



Mutualisation de données et de connaissances pour la Gestion Intégrée des Zones Côtières. Application au projet SYSCOLAG.

Julien Barde

► To cite this version:

Julien Barde. Mutualisation de données et de connaissances pour la Gestion Intégrée des Zones Côtières. Application au projet SYSCOLAG.. Mathématiques [math]. Université Montpellier II - Sciences et Techniques du Languedoc, 2005. Français. NNT: . tel-00112661

HAL Id: tel-00112661

<https://theses.hal.science/tel-00112661>

Submitted on 9 Nov 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ACADÉMIE DE MONTPELLIER

U N I V E R S I T É M O N T P E L L I E R I I

— S C I E N C E S E T T E C H N I Q U E D U L A N G U E D O C —

T H È S E

présentée à l'Université des Sciences et Techniques du Languedoc
pour obtenir le diplôme de doctorat

SPÉCIALITÉ : INFORMATIQUE
Formation Doctorale : Informatique
École Doctorale : Information, Structures, Systèmes

Mutualisation de données et de connaissances pour la Gestion Intégrée des Zones Côtières. Application au projet SYSCOLAG.

par

Julien BARDE

Soutenue le 9 Décembre 2005 devant le jury composé de :

M. Jean-Paul DONNAY , Professeur, Laboratoire Surfaces, Université de Liège, Belgique Rapporteur
Mme Anne DOUCET , Professeur, LIP6, Université Pierre et Marie Curie Paris VI Rapporteur
M. Marc ROBIN, Professeur, LTEG, IGARUN-Université de Nantes..... Rapporteur
M. Claude BOKSENBAUM , Professeur, LIRMM, Université Montpellier II Président
Mme Thérèse LIBOUREL , Maître de Conférences HDR, LIRMM, CNAM Directrice de Thèse
M. Pierre MAUREL , Ingénieur de Recherche Cemagref, UMR TETIS, Montpellier Co-directeur
M. Jean-Christophe DESCONNETS, Ingénieur de Recherche, IRD, Montpellier.....
M. David GENEST, Maître de Conférences, LERIA, Université d'Angers.....
M. Lionel LOUBERSAC, Directeur du Laboratoire Environnement - Ressources (LER/LR), IFREMER, Sète..
Mme Nabila MAZOUNI, Docteur en océanologie, chargée de mission au CEPRALMAR, Montpellier.....

Ce travail est dédié à la mémoire de Thomas Dricot et à notre passion partagée pour l’océan.

Remerciements

En premier lieu, je remercie la région Languedoc-Roussillon et le Cemagref qui ont financé la bourse de thèse dont j'ai bénéficié ainsi que l'université Montpellier II et le laboratoire du LIRMM qui ont assuré la direction scientifique de ce travail. En particulier, merci à mes encadrants, Thérèse Libourel et Pierre Maurel pour leur aide, leur disponibilité et leurs encouragements durant le déroulement de ma thèse et lors de la rédaction. J'exprime ensuite ma reconnaissance aux membres du jury qui ont accepté la lourde tâche qui a consisté à évaluer ce travail. Je remercie également tous les acteurs du programme de recherche Syscolag pour le plaisir que j'ai eu à travailler avec eux. En particulier, j'adresse une pensée à mes petits camarades thésards pour l'amitié que nous avons pris la peine de renforcer par les nombreuses réunions extra-professionnelles qui ont jalonné ces trois années et témoignent de l'ambiance agréable qui a régné durant ce programme. J'associe également Nabila qui aura, en tant qu'animatrice du projet, sûrement su apprécier les traits de mon caractère marqués par l'héritage génétique de mes aïeux périgordins, vers la vallée de la Vézère, là où Cro Magnon a semble-t-il, mieux qu'ailleurs, pu s'épanouir durant l'âge de Pierre. Remerciements également à l'équipe de ROSELT et à son représentant local pour le foie gras landais, sans Desconnets les liens entre départements du Sud-ouest aquitain n'auraient pas été renforcés de la sorte, qu'il en soit remercié.

Puisque je suis docteur, mes remerciements vont d'abord à Monique, docteur en psychiatrie, qui a accompagné de ma naissance jusqu'à ce vendredi 9 décembre, les moments les plus difficiles comme les plus heureux de ce sympathique début de vie. Il m'aura fallu attendre 28 ans pour pouvoir l'appeler chère collègue, puisse l'avenir m'apporter des collègues de cette qualité. Que ce titre de docteur me rappelle en premier lieu les liens qui nous unissent et les bienfaits de sa générosité.

Ensuite, puisque tout cela me rappelle au goût de l'eau salée, je rends hommage ici à ma famille consanguine ou assimilée, en bloc et sans distinction puisque les égards appellent les jalousies. Les éléments de cette tribu ont largement contribué à faire naître ma passion pour la nature, l'océan, et financer beaucoup de mes navigations pédestres ou côtières qu'ils en soient remerciés.

Ensuite, les amis, jeunes ou vieux. Les ancêtres d'abord : remerciements à Lulu et Rémi...quand Périgord rime avec Sainte Marie de Ré, ils ont fait de l'estran et de l'insularité un paradis caché pour périgord exilé. A Wil, mon copiale montagnard, compagnon de turne et tellement plus depuis chacune des années qui a suivie notre sortie de prison. La taule, ça forge des liens solides et c'est dans ses cellules que se préparent les meilleurs coups. De Wil en aiguilles, du midi ou des Pyrénées, à Jeny pour la préparation mentale qu'elle m'a apportée par expérience interposée, à Nico pour le charmant bout de route que nous avons parcouru ensemble et amicalement partagé. Aux collègues de la MTD, j'aurais pu tomber plus mal, vive les réunions clandestines et arrosées ! A mes potes de Rennes pour notre période fructueuse de réinsertion dans la société...moi aussi j'espère devenir Chef et être vouvoyé.

Enfin s'il fallait sortir le polish et la brosse à reluire pour lustrer les souliers de quelques repères éducatifs appréciés, je mettrai bien un petit coup sur les godillots de Gérard Douaire et son équipe de jedis ainsi que sur les palmes de Marc Taquet, bel exemple de passion partagée.

Enfin puisqu'il s'agit des dernières lignes de mon dernier cahier d'écolier, Agathe j'écris ton nom et pour le texte je te laisse deviner.

Sommaire

1	Introduction	15
1.1	Contexte de la thèse	15
1.2	Problématiques abordées dans la thèse	16
1.3	Objectifs de la thèse	19
1.4	Plan de la thèse	20
2	Contexte et problématique	23
2.1	Contexte général de la GIZC : concepts et caractéristiques	24
2.1.1	Origine de la GIZC : le Développement Durable, concept fondateur . . .	24
2.1.2	GIZC : un concept flou sur un espace flou	30
2.1.3	GIZC et information	35
2.2	Contexte local : le programme régional Syscolag	41
2.2.1	La zone côtière du Languedoc Roussillon	41
2.2.2	Le programme Syscolag	43
2.3	Systèmes d'information pour la GIZC : état de l'art	46
2.3.1	Directives générales pour la gestion de l'information du domaine	46
2.3.2	Systèmes de références pour la gestion de l'information en GIZC	49
3	Objectifs et propositions	61
3.1	Objectifs principaux de la solution proposée	62
3.1.1	Adaptation aux utilisateurs	62
3.1.2	Respect des infrastructures régionales existantes	63
3.1.3	Interopérabilité sémantique	64
3.1.4	Respect des standards / normes de référence	65
3.2	Architecture du système proposé	67
3.2.1	Niveau 1 : Accès à la connaissance	68
3.2.2	Niveau 2 : Fédération	69
3.2.3	Niveau 3 : Ressources	69
3.3	Les cas d'utilisations	70
3.3.1	Les rôles associés aux différents acteurs humains	70
3.3.2	Les contraintes d'accès	72
4	Partager les ressources de la GIZC	77
4.1	Intérêt des métadonnées pour le partage des ressources	78
4.1.1	Interopérabilité des bases de données environnementales	78
4.1.2	Définition du concept de métadonnée	79

4.1.3	Métadonnées et NTIC	80
4.1.4	Typologie d'éléments de métadonnées	81
4.1.5	Objectifs associés à l'utilisation des métadonnées	81
4.1.6	Les normes de métadonnées	82
4.1.7	Les recommandations européennes	85
4.2	Présentation de l'ISO 19115	87
4.2.1	Description des rubriques de l'ISO 19115	88
4.2.2	Description des éléments de métadonnées de l'ISO 19115	91
4.2.3	Implémentation de l'ISO 19115 : état de l'art	94
4.3	Implémentation de l'ISO 19115 pour la GIZC	101
4.3.1	Réalisation d'un profil Syscolag sur la base d'entretiens	101
4.3.2	Réalisation des gabarits associés au profil Syscolag	105
4.3.3	Modèle conceptuel proposé	106
4.3.4	Réalisation de la base de métadonnées	110
5	Partage de la connaissance en GIZC	115
5.1	Intérêt d'un vocabulaire contrôlé	116
5.1.1	Contexte et problématique du Web, de la GIZC et de l'ISO 19115	116
5.1.2	Glossaire	121
5.1.3	Thesaurus	123
5.1.4	Ontologie	131
5.2	Proposition d'un modèle d'ontologie spécifique à la GIZC	141
5.2.1	Démarche collaborative	142
5.2.2	Intérêt de l'approche écosystémique	147
5.2.3	Présentation du modèle d'ontologie	150
5.2.4	Spatialisation des concepts de l'ontologie	151
5.3	Typologies des classes du modèle d'ontologie	153
5.3.1	Méthodologie générale	153
5.3.2	Les Acteurs de la GIZC en LR	155
5.3.3	Les Ressources du littoral	157
5.3.4	Les Interactions de la GIZC en LR	162
5.3.5	Les outils de la GIZC en LR	164
5.4	Implémentation du modèle d'ontologie proposé	168
5.4.1	Utilisation "raisonnée" du modèle d'ontologie	168
5.4.2	Connexion entre métadonnées et connaissances	169
5.4.3	Réalisation : modèle physique de données	171
6	Le référentiel spatial	175
6.1	Intérêt d'une description spatiale contrôlée	176
6.1.1	Intérêt de l'ISO 19115 pour la description spatiale	177
6.1.2	Analyse et formalisation du lien entre concepts thématiques et spatiaux	181
6.2	Lien entre référentiel spatial et SMD	186
6.2.1	Modèle physique de données associé au référentiel spatial	187
6.2.2	Les requêtes spatiales désormais possibles sur la métadonnée	190
6.2.3	Intérêt de <i>Postgres</i> et <i>Postgis</i>	192
6.2.4	Interopérabilité du serveur de données spatiales mis en place	196
6.3	Alimenter le référentiel spatial par de l'IG pertinente	199

6.3.1	Méthodologie générale	199
6.3.2	Les sources de l'inventaire	202
6.3.3	Les entretiens auprès des partenaires du projet Syscolag	205
6.3.4	Intégration des analyses dans notre modèle d'ontologie	206
6.3.5	Caractéristiques de l'IG associée à la GIZC	208
7	Servir l'information et la connaissance	213
7.1	Choix techniques	215
7.1.1	Choix techniques pour le <i>SGBDR</i> et le <i>langage de script</i>	215
7.1.2	Choix techniques pour la gestion des référentiels sémantique et spatial	216
7.2	Interfaces graphiques de gestion du SMD	221
7.2.1	Interfaces graphiques de saisie des fiches de métadonnées	221
7.2.2	Interfaces graphiques de consultation du SMD	226
7.3	Interfaces graphiques de gestion de la connaissance	233
7.3.1	Ontologie	233
7.3.2	Atlas cartographique	236
8	Conclusion et Perspectives	243
8.1	Conclusion	243
8.2	Perspectives	246
8.2.1	Perspectives liées à l'évolution du service de métadonnées	246
8.2.2	Perspectives liées à l'utilisation du référentiel spatial	250
8.2.3	Perspectives liées à l'utilisation de l'ontologie	252

Liste des acronymes

AIZC	Aménagement Intégrée de la Zone Côtière
ASFA	Aquatic Sciences and Fisheries Abstracts
CEN	Comité Européen de Normalisation
CIADT	Comité Interministériel d'Aménagement et de Développement du Territoire
COI	Commission Océanographique Intergouvernementale
CORBA	Common Object Request Broker Architecture
CSDGM	Content Standard for Digital Geospatial Metadata
DC	Dublin Core
DD	Développement Durable
DTD	Document Type Definition (DTD)
ECO-IMAGINE	European Conferences and forum for Integrated coastal MANagement and Geo-INformation rEsearch
EEA	European Environment Agency
ETCCDS	European Topic Centre on Catalogue of Data Sources
FAO	Food and Agriculture Organization
FGDC	Federal Geographic Data Committee
FIGIS	Fisheries Global Information System
FOS	Fishery Ontology Service
GELOS	Global Environmental Locator System
GEMET	GEneral Multilingual Environmental Thesaurus
GIL	Gestion Intégrée du Littoral
GILIF	Gestion Intégrée du Littoral et des bassins Fluviaux
GIZC	Gestion Intégrée de la Zone Côtière
GMES	Global Monitoring for Environment and Security
GML	Geography Markup Language
GUI	Graphical User Interface (Interface Graphique Utilisateur en français)
IFREMER	Institut français de recherche pour l'exploitation de la mer
IG	Information Géographique
INSPIRE	INfrastructure for SPatial Information in Europe
IRD	Institut de recherche pour le développement

ISO	International Organization for Standardization
JDBC.....	Java Data Base Connectivity
KB.....	Knowledge Base
LR.....	Languedoc Roussillon
MADAME....	Methods for Access to data and Metadata in Europe
MCD.....	Modèle Conceptuel de Données
MIDA.....	Marine Irish Digital Atlas
MMI.....	Marine Metadata Interoperability
MOF.....	Meta Object facility
MOTIIVE....	Marine Overlays on Topography for Annex II Valuation and Exploitation
MPD.....	Modèle Physique de Données
NOAA	National Oceanic and Atmospheric Administration
NOKIS	Nord-Ostsee-Küsten-Information-System
NTIC	Nouvelles technologies de l'information et de la communication
OGC.....	Open GIS Consortium
OMG	Object Management Group
OWL.....	Web Ontology Language
PAP	Plan d'Actions Prioritaires
PHP	Hypertext Preprocessor
PLU	Plans Locaux d'Urbanisme
RDF.....	Resource Description Framework
RDFS.....	Resource Description Framework Schema
ROSELT.....	Réseau d'Observatoires de Surveillance Ecologique à Long Terme
SCOT.....	Schémas de COhérence Territoriale
SDI.....	Spatial Data Infrastructure
SFS.....	Simple Feature Specification
SGBD.....	Système de Gestion de Bases de Données
SGBDR.....	Système de Gestion de Bases de Données Relationnel
SMD.....	Service de métadonnées
SQL.....	Structured Query Language
SVG	Scalable Vector Graphics
SYSCOLAG..	SYStèmes COtiers et LAGUnaires
TC 211	ISO Technical Committee 211
UE.....	Union européenne
UNESCO....	United Nations Educational Scientific and Cultural Organization
URI.....	Uniform Resource Identifiers
W3C.....	The World Wide Web Consortium

WFS..... Web Feature Service
WMS..... Web Mapping Service
WWF..... World Wildlife Fund
XML..... Extensible Markup Language
XSL..... Extensible Stylesheet Language
XTM..... XML-based Topic Maps

Chapitre 1

Introduction

La machine ne fait que ce qu'on lui a demandé. Gérard Douaire.

1.1 Contexte de la thèse

Dans le contexte général du *Développement Durable* (*DD*) et dans celui plus spécifique de la *Gestion Intégrée de la Zone Côtière* (*GIZC*), a été énoncé un certain nombre de principes pour orienter la mise en œuvre concrète de ces nouveaux modes de développement et de gestion dont la portée opérationnelle est encore floue (en particulier la définition de la nature des *thématiques* et des *espaces* concernés par ces processus). Dans ce document, nous proposons de nous intéresser aux principes relatifs à la *gestion* de l'*information* et de la *connaissance* qui présentent une importance majeure pour amorcer et assister la mise en œuvre du *DD* et de la *GIZC*. Cependant, cette démarche se heurte globalement aux verrous suivants :

- au *cloisonnement institutionnel* d'*acteurs* variés,
- la nature très *hétérogène* de l'*information* à mobiliser dans ces processus et sa *dispersion*. Ceci est en lien direct avec la diversité des *acteurs*,
- les *NTIC* ont généré une explosion de la production et de la diffusion d'*information numérique* (ceci est particulièrement vrai pour l'*information géographique*) ce qui a révélé que certains des modes de gestion traditionnels de l'*information* et de la *connaissance* sont désormais obsolètes pour *décrire* ou *consulter* le *patrimoine* existant.
- le manque de *référentiels communs* pour assister le partage *informationnel* et *sémantique*.

Néanmoins, les *NTIC* apportent par le biais du Web un média de choix pour la diffusion de l'*information* et de la *connaissance* vers le *grand public*. En outre, dans le cas de la *GIZC* qui est, par essence, un *mode de gestion territorialisée*, la place de l'espace est centrale à l'image de l'*information géographique*, omniprésente dans les processus de *GIZC* (*télédétection*, *SIG*...). Le *support cartographique* peut donc aider au partage de l'*information* et la *connaissance*.

Cette thèse s'inscrit dans le cadre du programme de recherche pluridisciplinaire *Syscolag*¹, qui a pour objectif d'aider à optimiser la *gestion intégrée de la zone côtière* à l'échelle régionale du *Languedoc Roussillon*. A travers ce programme, nous disposons d'un échantillon d'*acteurs* et de *modes de gestion* des *ressources informationnelles* qui leur sont associées particulièrement représentatif des *verrous scientifiques* qui perturbent la *gestion* efficace de l'*information* et de la *connaissance* dans les processus de *DD* et de *GIZC*.

¹Ce programme est co-financé par la région *Languedoc Roussillon* et l'*IFREMER*.

En effet, la grande variété d'acteurs et de disciplines représentés pose des problèmes de collaboration d'ordre *physique* (le cloisonnement des acteurs étant renforcés par leur dispersion géographique) auxquels s'ajoutent des problèmes d'ordre *technique* (à cause des divers modes de *gestion* de l'*information*) qui sont liés notamment à des problèmes d'ordre *sémantique* (faute de *langage commun*) qui perturbent le *partage* de *connaissances* aussi bien entre les *acteurs* qu'entre les *systèmes*. Pour répondre aux problématiques soulevées, le programme *Syscolag* ne fournit ni n'impose de réponses pour les solutions à apporter. Cependant, il oriente implicitement la nature des choix potentiels en encourageant la prise en compte des points suivants :

- la prise en compte des *attentes des utilisateurs* est nécessaire pour la pertinence de la réponse qui sera apportée,
- les *NTIC* apportent une réponse pertinente aux *contraintes d'éloignement géographique*, en particulier grâce au Web. Les *NTIC* sont d'ailleurs globalement perçues comme un apport primordial pour supporter les processus de *GIZC* [101] et le programme *Syscolag* souhaite concrètement mettre à profit les réponses apportées au sein d'un *portail Web* accessible au *grand public*,
- le *portail* proposé hébergera une *base de connaissance* dont la nature dépendra de celles des solutions retenues à l'issue de ce programme (dont une partie doit provenir spécifiquement de ce sujet),
- les réponses informatiques proposées ne doivent pas contraindre les *acteurs* à un changement brutal de leur *mode de gestion* ou de *représentation* de l'*information* et de la *connaissance*. Au contraire, il s'agit d'accompagner, par des *systèmes adaptés*, la nécessaire élaboration de *référentiels* régionaux partagés pour décrire l'*information* et la *connaissance* existantes en vue de leur ***localisation*** et, à terme, si cela se justifie par les *contraintes d'accès* liées à chaque *ressource*, d'un éventuel *partage physique*.
- l'absence d'un ***langage commun*** pour la *GIZC* nuit au partage des *informations* et *connaissances* associées qui sont essentielles à cette démarche.
- la possibilité de trouver une partie de la réponse aux problèmes d'*échange terminologique* en appuyant le partage d'*information* et de *connaissance* par le *support cartographique*, qui facilite le travail collaboratif entre acteurs.

Nous proposons à présent de détailler, point par point, les problématiques informatiques qui peuvent aider à la résolution de ces verrous.

1.2 Problématiques abordées dans la thèse

Comme nous venons de le décrire, dans le domaine du *DD* et de la *GIZC*, les réponses apportées pour la *mutualisation* de l'*information* et de la *connaissance* doivent permettre la résolution des verrous scientifiques que sont, essentiellement, la maîtrise de la *description* d'*informations* et de *connaissances* de natures variées et *distribuées* chez les *acteurs* impliqués dans ces processus. Par ce biais, un premier niveau d'*interopérabilité des bases de données environnementales* mobilisées pour la *GIZC* sera alors possible et permettra une *localisation de l'information*. D'un point de vue informatique, un deuxième niveau d'*interopérabilité* concerne, en complément de la *localisation* de l'*information*, le *traitement physique* des *données* contenues dans les *SGBD* des partenaires. Compte tenu du contexte de la *GIZC*, ceci paraît plus délicat à atteindre et ne fait pas partie des objectifs accessibles par des méthodes génériques à court terme.

En outre, la *connaissance* accumulée à ces fins pourra être utilisée par les *systèmes* (pour une *recherche* et une *localisation* plus pertinentes de l'*information recherchée*) et restituée aux acteurs, les aidant ainsi à *élaborer* et *partager* un *langage commun*. A l'*interopérabilité syntaxique* des *SGBD*, il convient donc d'ajouter l'*interopérabilité sémantique* des systèmes et des acteurs.

Compte tenu de la nature des verrous évoqués et de l'intérêt de la réponse apportée par l'usage du Web, nous proposons de nous intéresser aux réponses apportées par ce dernier à des problématiques de recherches similaires. En effet, à l'échelle du Web, ces mêmes verrous sont renforcés car la complexité de la *gestion d'information et de connaissance hétérogènes* associées à un public "*World Wide*", ajoute une dimension supplémentaire aux problèmes d'*interopérabilités syntaxique et sémantique* entre les *systèmes* ou les *acteurs* par rapport à ceux d'une simple *communauté* voire d'une *communauté locale*. Les méthodes et réponses proposées se doivent d'être *génériques* et peuvent donc être déclinées pour appuyer la résolution de problématiques plus *spécifiques*, comme celles de la *communauté* mobilisée dans la *GIZC*. Après avoir démontré la nécessaire prise en compte du *facteur humain* dans notre contexte, nous proposons de regarder la nature des solutions *génériques* proposées pour le Web et spécifions celles sur lesquelles nous comptons nous appuyer pour proposer notre propre réponse.

Social Informatics Avant de s'intéresser aux problématiques essentiellement informatiques, il convient de relever que la *collaboration interdisciplinaire* est une condition nécessaire au succès des solutions que nous allons proposer. La réponse apportée trouvera, en effet, une part de sa pertinence dans la prise en compte de *facteurs humains* et donc *sociaux* associés à la *GIZC*. Il y a donc une rencontre entre les *sciences sociales* et *informatiques* qui raccroche la thématique générale de ce sujet au *champ interdisciplinaire* de *Social Informatics*². Par exemple, la part de *décloisonnement des institutions* nécessaires au partage de l'*information* et de la *connaissance* ne se résoudra pas, a priori, uniquement par la qualité technique des méthodes proposées (sans quoi le Web, seul, pourrait suffire) mais par l'agrément des *acteurs* à une solution spécifique qui prend en compte leurs attentes. Il faut donc s'adresser avec une écoute particulière à une communauté mûre et animée par une démarche volontaire. En ce sens, la prise en compte du *support cartographique* pour appuyer la mutualisation est un pas vers les besoins des utilisateurs de notre communauté.

Les problématiques de gestion de l'information et de la connaissance véhiculées par le Web La gestion optimale du Web, média *grand public*, passe nécessairement par la *gestion* de la *connaissance* portée par l'*information* que servent à visualiser les *navigateurs*. Or, actuellement, le traitement des documents se fait de manière passive alors qu'il serait souhaitable qu'ils soient conçus pour être interprétables par la machine à l'aide d'une *formalisation* qui soit appropriée à une description fine de l'*information* et de la *connaissance* qu'ils contiennent et rigoureusement respectée. En effet, ce point est central afin que les *moteurs de recherches* débroussaillent la jungle informationnelle du Web et répondent aux *requêtes* pleines d'espoirs de leurs *utilisateurs*. Les efforts de recherche et les enjeux économiques liés à la résolution de cette problématique sont colossaux et les algorithmes qui permettent la résolution

²Social Informatics est un champ interdisciplinaire. Au lieu de se concentrer uniquement sur la réalisation d'interfaces plus utilisables, cette discipline recherche comment rendre les systèmes entier plus utilisables dans leur contexte social.

pertinente de questions posées aux *moteurs de recherche*³ valent de l'or⁴ car ils se substituent à cet effort de *description du contenu des documents* que les producteurs de ressources du Web ont soigneusement pris la peine de ne pas effectuer (en particulier en utilisant le moins possible l'élément de *métadonnées* esseulé et encapsulé dans l'*en-tête* des fichiers *HTML*, tout en continuant à se demander comment bien être référencé par les *moteurs de recherche*). La *gestion* de l'*information* et de la *connaissance* du Web actuel en est donc encore à un stade *préhistorique* et selon Tim Berners Lee, son concepteur, le Web de demain sera *sémantique* [22] et permettra, en plus d'une simple visualisation des documents qu'il diffuse, de *gérer* l'*information* et la *connaissance* qu'ils contiennent implicitement par l'emploi de *langages de description normalisés* qui les explicitent et permettent leur interprétation et leur représentation aussi bien par la *machine* que par l'*homme*. Par ces approches, les problèmes de *palliatifs sémantiques* des *moteurs de recherche* devraient être en partie résolus.

D'une manière générale, il y a deux problématiques sous jacentes à la *mutualisation* de l'*information* et de la *connaissance* : la pertinence des *formalismes* ou *syntaxes* adoptés pour les décrire (*normalisés*, *standardisés* ou *libres*) que ce soit pour les *descripteurs* (pour la *mutualisation de l'information*) ou les *éléments* qui les valuent (pour la *mutualisation de la connaissance* associée). En effet, ces choix ne sont pas neutres si on souhaite également permettre au système proposé l'*interopérabilité syntaxique* et *sémantique* avec d'autres qui partagent des formalismes compatibles. Plus concrètement, nous allons voir que les *solutions techniques* mises en avant pour la résolution de problématiques similaires à celle du *Web sémantique* concernent les *métadonnées* pour *mutualiser l'information* et les *ontologies* pour maîtriser la *sémantique* des descriptions et la *connaissance* qui leur est associée.

Mutualiser l'information Par *mutualisation de l'information*, nous entendons implicitement *partage de la description des ressources informationnelles existantes*. Ceci suffit pour atteindre l'objectif préalable - essentiel pour la *GIZC* - de *localisation des ressources* avant de se soucier, à terme, de leur *partage physique*. A ces fins, la *description ressources existantes* doit s'appuyer, d'une part, sur des *métadonnées* qui sont des *descripteurs* pertinents permettant, entre autres, de *localiser l'information* et, d'autre part, sur la *séparation physique* entre *données* et *métadonnées* (contrairement au cas des fichiers *HTML*) qui facilite la localisation de l'information indépendamment de sa nature. La mise en place d'un *service de métadonnées* permet la gestion indépendante des ressources et des *métadonnées*. En effet, les *métadonnées* sont dédiées à la description de ressources informationnelles variées pour permettre de connaître quelles ressources existent, les contraintes d'accès qui leurs sont associées et à quels usages celles-ci peuvent servir. Elles permettent par ce biais l'*interopérabilité syntaxique* des bases de données (en rendant possible la *localisation des ressources* qui y sont *dispersées*). L'usage d'une *norme de métadonnées* permet, quant à lui, l'*interopérabilité* d'un *service de métadonnées* avec d'autres qui partagent ces mêmes *syntaxes normalisées de description*. Il existe des *normes spécialisées de métadonnées*, par exemple pour la gestion de l'*information géographique numérique* dont les caractéristiques techniques nécessitent des *descripteurs* qui lui sont spécifiques (notamment des éléments de *métadonnées* qui décrivent la *dimension spatiale de la ressource documentée*).

³Entre autres, pour localiser l'information pertinente dans les fichiers *HTML*, les *moteurs* utilisent les éléments *métadonnées*, *titre*, *producteur*... situés dans l'*en-tête* ainsi que des *recherches plein texte* sur le *contenu* et d'autres critères plus ou moins propres à chacun d'eux.

⁴L'argument se vérifie si on se réfère au cours de la bourse et à celui des actions *Google* à l'heure où ce document est rédigé.

Cependant une *norme de métadonnées* n'apporte pas systématiquement de *contrôle de la valuation* des *éléments de métadonnées* qu'elle propose. Elle balise (ou encadre), tout au plus, la nature des valuations ou des connaissances qu'elle héberge sans forcément la contrôler. Or, ceci est essentiel pour l'*interopérabilité sémantique* des *machines* voire des *acteurs* qui en prennent *connaissance*.

Mutualiser la connaissance Par *mutualisation de la connaissance*, nous entendons le *partage* de la *signification* des *termes* employés pour indexer l'*information* contenue dans une *ressource*. Pour cela, concrètement, le sens des *concepts* (ou *mots clés*) qui valuent les *éléments de métadonnées* doit être contrôlé par une mise en forme adaptée (*syntaxique* et *sémantique*) afin que les *agents* de recherche qui les traitent puissent interpréter sans ambiguïté l'*intention* du *producteur de la métadonnée* lors de l'*indexation*. L'*outil de contrôle terminologique* le plus fin, permettant aussi bien la *gestion* de *concepts* que de la *connaissance* qui leur est associée dans des *domaines* variés, est l'*ontologie* dont dépend la mise en œuvre du *Web sémantique* et plus largement de nombreux problèmes d'*interopérabilité sémantique*. En effet, elle permet la mise en place d'un *référentiel sémantique* nécessaire pour le partage d'un *langage commun* interprétables par les *machines* et les *acteurs*.

L'*ontologie* est donc le complément idéal d'un *service de métadonnées* [98, 73] puisqu'elle permet la *gestion des connaissances* portée par les *concepts* qui valuent les *éléments de métadonnées*. Le *contrôle sémantique* apporté optimise donc la *qualité d'indexation* et de la *localisation de l'information* et ce, d'autant plus, qu'elle est également utilisée pour alimenter les *moteurs de recherche* qui en assument la charge. L'*ontologie* participe donc également à la *mutualisation de l'information*.

Par le biais du Web, il est en outre possible, notamment par le biais des *Web services*, de mettre en pratique l'*interopérabilité* des *services de métadonnées normalisés* ou d'*ontologies* liées à différents *domaines* et de bénéficier en particulier de l'ensemble des spécifications pour la diffusion de l'*information géographique* (*Web Mapping Services*). La solution que nous proposons tient compte de ces possibilités.

1.3 Objectifs de la thèse

Cette thèse a pour objectifs d'apporter une ou des *méthodes génériques*, qui pourront être exportées dans des cas similaires, visant à améliorer la *mutualisation de l'information et de la connaissance* dans la *GIZC*. Elle s'appuiera sur un *démonstrateur* ou *prototype* qui permettra d'évaluer la pertinence de ces méthodes. En particulier, par l'utilisation de *sources*, *approches*, *outils*, *normes* ou *standards* de référence et par la capture des *besoins des utilisateurs*, il s'agit de pouvoir conjuguer l'*interopérabilité* du système mis en place avec un éventuel *réseau de systèmes équivalents* ainsi que son adaptation au contexte particulier d'implémentation et à ses utilisateurs. Ceci passe par la bonne appréciation de la notion de *GIZC* et de la nature de l'*information* et de la *connaissance* produites et par celle des spécificités régionales de la *zone côtière du Languedoc Roussillon*. Afin de faciliter le dialogue, l'analyse et la formalisation avec les *utilisateurs* nous utiliserons le formalisme *UML* et le *support cartographique*. Cette démarche doit garantir une réponse particulière au contexte d'implémentation.

Concernant la nature des solutions préconisées pour la mutualisation de l'*information* et de la *connaissance*, les problèmes techniques, quant à eux, ne dépendent pas de l'échelle géogra-

phique d'application. Au contraire même, puisque portée par le Web, l'application envisagée a tout intérêt à s'appuyer sur les méthodes génériques du Web pour pouvoir partager son contenu avec tout autre système qui partagent ces méthodes et objectifs conformément aux principes énoncés pour la *GIZC*.

Pour cela et compte tenu de la volonté d'utiliser le Web, nous proposons de nous appuyer sur les *spécifications* existantes de *Web Services* qui intéressent notre problématique pour faciliter la diffusion de nos solutions auprès d'un public large, conformément aux objectifs du *DD*.

Nous préconisons donc, compte tenu de notre contexte et des caractéristiques de la *GIZC*, la mise en place d'un *service de métadonnées* qui va aider à *mutualiser l'information* sur les ressources existantes pour la *GIZC* et d'une *ontologie*, basée sur le contenu d'un *référentiel sémantique*, qui va permettre de *mutualiser la connaissance* associée aux *concepts* de ce *domaine* (en particulier pour les *concepts thématiques* et *spatiaux*). En outre, nous proposons l'élaboration d'un modèle d'*ontologie* spécifique à la *GIZC* pour le *contrôle* de la *valuation* de certains *éléments de métadonnées* ainsi que pour la consultation en ligne d'une *base de référence terminologique* sur les *concepts* de la *GIZC* et leurs *propriétés sémantiques*. Afin d'exploiter au mieux le *support cartographique* sur notre *portail Web*, nous proposons la mise en place d'un *atlas cartographique* qui s'appuiera sur les *représentations géométriques géoréférencées* des *concepts spatiaux* ou *toponymes* que nous gérons dans le *référentiel sémantique* (à l'aide d'un *SBGDR à cartouche spatiale*). En outre, pour l'ensemble de ces *composants*, il s'agit de proposer des applications simples à *administrer* et à *utiliser* puisque certaines fonctionnalités de l'outil visé seront à la disposition du *grand public*.

1.4 Plan de la thèse

Ce manuel est structuré en huit chapitres. Nous avons voulu, pour faciliter la lecture, faire en sorte que chaque chapitre puisse constituer un tout intelligible. Chaque chapitre constitue une avancée à l'édifice proposé mais peut être parcouru de manière indépendante, au prix de quelques rappels obligatoires pour resituer le contexte général.

Dans le **chapitre 2**, nous présentons le *contexte* général de ce travail et précisons les origines et les définitions des termes qui définissent le cadre conceptuel de notre *problématique*. Dans un premier temps, nous revenons sur l'importance de la *gestion de l'information* et de la *connaissance* dans les principes fondamentaux du *Développement Durable*. Nous montrons ensuite que la *Gestion Intégrée de la Zone Côtière (GIZC)* est une *application territorialisée* du *DD* sur la *zone côtière*, qui, par ses particularités, exacerbe les *verrous humains* et *informatiques* liés à la problématique de mise en œuvre d'une *gestion* efficace de l'*information* et de la *connaissance*. En particulier, la diversité des *ressources informationnelles* et l'importance de l'*information géographique* (liée à l'importance de l'espace, capitale dans la *GIZC* en tant que mode de gestion territorialisé) ainsi que la diversité et le cloisonnement des *acteurs* mobilisés génèrent des problèmes de *partage physique* de l'*information* et en particulier des *problèmes sémantiques*. Nous illustrons cette problématique générale en précisant les objectifs spécifiques du programme de recherche *Syscolag* pour la *gestion* de l'*information* et de la *connaissance* associées à la *GIZC* dans le contexte régional du *Languedoc Roussillon*. Nous terminons par un *état de l'art* sur les *systèmes* qui apportent des réponses aux verrous que nous venons d'énumérer.

Dans le **chapitre 3**, nous présentons les *objectifs* et *propositions* de cette thèse. Nous détaillons les principaux *composants* du système que nous proposons pour résoudre les verrous

de notre problématique. Nous justifions la nécessité de respecter les modes de *gestion* de l'*information* en place chez les acteurs mobilisés et la pertinence d'un système de médiation basé sur l'utilisation des *métadonnées* pour gérer l'*information* sur les *ressources existantes* pour la *GIZC* en *Languedoc Roussillon* et permettre leur *localisation*. En complément, nous proposons la mise en place d'un *référentiel sémantique* qui répond à la problématique de *gestion* de la *connaissance* du *domaine*, en particulier aux problèmes *terminologiques* (pour la gestion du *service de métadonnées* ou le partage de *connaissances* entre *acteurs*). En outre, nous proposons une gestion perfectionnée des *descriptions spatiales* pour permettre de nous appuyer sur des *supports cartographiques*. Nous terminons en présentant les *cas d'utilisation UML* associés au système que nous préconisons.

Dans le **chapitre 4**, nous détaillons la pertinence des *métadonnées* pour résoudre les problèmes d'*interopérabilité des bases de données environnementales* liées à la *GIZC* en *LR*. Nous présentons ensuite la *norme de métadonnées* choisie (*ISO 19115*), ses propriétés et quelques *outils d'édition* qui l'implémentent. Nous justifions les adaptations (*profil* et *gabarits*) que nous avons effectuées pour l'adapter au contexte général de la *GIZC* et du programme *Syscolag* en particulier (en tenant compte des remarques de nos partenaires). Pour cela, nous proposons de distinguer différentes catégories d'*éléments de métadonnées* avec des *modes de gestion* adaptés à leur nature. Ceci est possible par une *approche mixte* que nous recommandons pour l'élaboration du *modèle conceptuel* de *base de métadonnées* dans un *SGBDR*. En effet, nous gérons de manière *générique* (multi-norme) certains *éléments de métadonnées* et d'autres de manière *spécifique* dans des tables dédiées qui permettent le *contrôle des valuations* qui leurs sont associées et donc leur *automatisation* ainsi que la *gestion* de leurs *redondances*.

Dans le **chapitre 5**, nous expliquons l'importance des *éléments de métadonnées* clés pour la *gestion* de l'*information* et de la *connaissance* liées à la *GIZC*. En particulier ceux qui renseignent sur la *thématique* et la *dimension spatiale des ressources décrites* sont essentiels car ils peuvent, si on ne contrôle pas leur utilisation, générer des problèmes d'*interopérabilité sémantique* préjudiciables à la gestion du *service de métadonnées* par une interprétation erronée des *connaissances* du domaine associées aux fiches de métadonnées lors de leur *saisie* ou de leur *consultation*. En outre, sans maîtrise de la *signification* de ces *concepts*, l'élaboration d'un *langage commun* et le *partage de connaissance* entre les *acteurs* seraient *illusoirs*. Après avoir expliqué les différentes options qui permettent de *contrôler* la *description* et la *restitution* de la *terminologie* et de la *connaissance* d'un *domaine*, nous proposons la mise en place d'une *ontologie du domaine de la GIZC* afin d'optimiser le *contrôle* du *sens* des *valuations* des *éléments de métadonnées* impliqués ainsi que le *partage* des *connaissances* qui leur sont associées. En nous basant sur l'*approche éco-systémique* préconisée en *GIZC*, nous proposons un *modèle générique d'ontologie* qui exploite le contenu du *référentiel sémantique* (*inventaire* et *structuration des concepts*) en rendant compte de la nature des *interactions* entre *environnement* et *société* conformément aux exigences du *DD* et de la *GIZC*. Nous implémentons ce modèle pour le cas particulier de la *GIZC* en *LR* et proposons une *méthodologie* qui permet d'alimenter le *référentiel sémantique* et de l'exploiter sous la forme d'un *outil de contrôle de vocabulaire* ainsi que d'une *base de connaissance*.

Dans le **chapitre 6**, nous nous intéressons spécifiquement à la *gestion* de l'*information* et de la *connaissance spatiales* nécessaires pour respecter l'importance de l'espace dans la *GIZC*. Pour cela nous nous appuyons sur la pertinence des *descripteurs spatiaux* proposés par les spécifications de l'*ISO 19115* basées sur celles de l'*OGC*. Nous explicitons ensuite le *lien* implicite existant entre *concepts thématiques* et *spatiaux* et levons de fait une ambiguïté de l'*ISO 19115*. Ainsi, nous enrichissons le *modèle* de notre *référentiel sémantique* présenté dans

le chapitre 5 par un sous ensemble dédié à la *gestion des descripteurs spatiaux* (*concepts* et *relations* associées) qui constitue un *référentiel spatial* pour la *GIZC*. Nous proposons l'implémentation des *modèles physique et conceptuel de données génériques* qui en découlent au sein d'un *SGBDR* doté d'une *cartouche spatiale*. Ceci permet de faciliter l'exécution des *requêtes spatiales* et en particulier la *description* et la *consultation* des *fiches de métadonnées*. Finalement, nous précisons le contenu de notre *référentiel spatial* en discutant de la nature *thématique* de l'*information géographique* à y intégrer pour proposer aux utilisateurs du système un inventaire pertinent d'*objets géographiques de référence pour la GIZC*.

Dans le **chapitre 7**, nous présentons les *interfaces graphiques utilisateurs* de notre *prototype* qui permettent la réalisation concrète des *rôles utilisateurs* décrits dans le chapitre 3. Nous présentons les *choix techniques* de l'*environnement de développement libre* sur lequel s'appuie notre système : *langages* et *applications* libres de droit qui répondent à leur manière à l'éthique nécessaire à la mise en œuvre des projets de *Développement Durable*. Puis nous illustrons les différents *cas d'utilisation* du système par des extraits d'*interfaces graphiques* qui ont été développées sur la base de ces techniques. En particulier nous décrivons les modalités de *saisie* et de *consultation* des *fiches du service de métadonnée* et nous détaillons les cas de *description* des *éléments de métadonnées* essentiels pour notre contexte, contrôlés par le *référentiel sémantique* : les *concepts thématiques* et *spatiaux* grâce à la mise en place d'*interfaces conviviales* (notamment *cartographique*) dédiées à ces valuations et incluses dans les *formulaires de recherche* ou de *saisie*. Nous illustrons également les *interfaces* qui utilisent les valeurs stockées dans l'*annuaire électronique* ou la *base de référence bibliographique*. Alors que les interfaces précédentes répondent directement à la problématique de *gestion* de l'*information* sur les *ressources informationnelles* existantes pour la *GIZC*, nous proposons en complément, la consultation séparée du *référentiel sémantique*, à travers une *base terminologique*, un *réseau sémantique* et un *atlas cartographique* de la zone côtière régionale, qui répondent plus directement à la problématique de *consultation* et de *partage* des connaissances.

Dans le **chapitre 8**, nous concluons sur la qualité des résultats apportés par les méthodes proposées pour la résolution des verrous scientifiques de notre problématique. En outre, nous proposons, en perspectives, quelques réflexions sur les suites à donner aux méthodes et aux développements proposés concernant le *service de métadonnées* et les *référentiels sémantique et spatial*. En particulier, sont discutées les perspectives d'*interopérabilité* entre *service de métadonnées* ou *serveurs de données spatiales*, équivalents à ceux que nous préconisons, distribués en réseau. Enfin, nous terminons par les perspectives de *formalisation* plus stricte envisageable pour la réalisation d'une *ontologie formelle* et la modélisation plus fine de *processus environnementaux* plus complexes, à l'aide du travail de *formalisation élémentaire* réalisé dans cette thèse.

Chapitre 2

Contexte et problématique

"On croit que l'homme est libre... On ne voit pas la corde qui le rattache au puits, qui le rattache, comme un cordon ombilical, au ventre de la terre". Antoine de Saint-Exupéry (Terre des hommes).

Dans cette partie nous souhaitons présenter au lecteur quelles sont les caractéristiques essentielles du domaine complexe dans lequel nous intervenons. Nous insistons sur les enjeux, difficultés et contraintes associés à la mise en œuvre du *Développement Durable* (DD) et de la *Gestion Intégrée de la Zone Côtière* (GIZC) en particulier ceux qui concernent l'amélioration de la gestion de l'information et de la connaissance auxquels ce travail s'intéresse.

Nous allons donc revenir sur la définition des concepts qui se rapportent à notre domaine d'application et notamment sur la filiation entre le concept de développement durable et le concept de GIZC, le premier étant à l'origine du second. Ainsi, en présentant certains objectifs généraux du DD, nous permettrons de mieux appréhender le concept de GIZC. Nous nous intéresserons spécifiquement au partage de l'information et de la connaissance qui est un des enjeux majeurs dans la mise en œuvre du DD sans détailler toutefois l'ensemble des principes du DD. Puis nous discuterons du concept de GIZC en considérant les spécificités de ce domaine qui nous intéresseront plus directement pour la mise en place d'un système adapté au partage des ressources et de la connaissance.

Nous reviendrons ensuite sur les caractéristiques du littoral du *Languedoc Roussillon* et du programme de recherche *Syscolag*, dans lequel s'insère ce travail, ainsi que ses attentes en terme de gestion de l'information à l'échelle de la zone côtière régionale.

Puis nous concluons par un état de l'art sur les projets et systèmes existants qui proposent une gestion de l'information et de la connaissance sur la GIZC (ou des domaines connexes) et les outils associés ainsi que leurs limites (en particulier en ce qui concerne l'utilisation de référentiels thématiques et géographiques et autres standards pour l'interopérabilité entre systèmes, en particulier les métadonnées). Nous y présentons les réponses informatiques apportées par d'autres programmes de recherche, et fournissons des éléments de comparaison nécessaires pour apprécier, dans le chapitre suivant, le contenu de notre proposition pour le programme Syscolag, au regard des spécifications générales de la GIZC et des caractéristiques propres à notre région d'application.

2.1 Contexte général de la GIZC : concepts et caractéristiques

A travers une présentation du concept de DD sur lequel est fondée, en grande partie, la définition de la GIZC, nous allons introduire le concept de GIZC et ses enjeux essentiels sous l'angle particulier de la gestion de l'information et de la connaissance.

2.1.1 Origine de la GIZC : le Développement Durable, concept fondateur

Nous proposons dans cette partie de décrire les origines et le cadre conceptuel du DD. Sans rentrer dans le détail, notre objectif ici est d'illustrer les considérations (ou inquiétudes) scientifiques qui ont donné naissance à ce concept. Puis, nous retranscrivons les grandes étapes de sa formalisation politique, ses enjeux et les difficultés inhérentes à la mise en œuvre des principes de ce nouveau mode de développement, pour mieux situer, par la suite, le cadre de notre intervention : l'aide à la gestion de l'information et de la connaissance.

Retour aux racines

Les idées qui ont engendré le concept de *DD* sont bien plus anciennes que le concept lui-même. Le bouleversement des modes de développement humain par les avancées technologiques majeures au cours des deux derniers siècles ont conduit de nombreux scientifiques à s'interroger sur la capacité des *écosystèmes* à supporter des *pressions anthropiques* en constante évolution.

Actuellement, parmi l'ensemble des *principes* reconnus nécessaires à la mise en place d'un *DD* à l'issue de la convention de Rio, le *principe 20* reconnaît à la femme une *place centrale dans la gestion de l'environnement et le développement*. Dans certains pays, cela est, paraît-il, acquis. En hommage à cette évidence encore contestée, voici, pour introduire le sujet du *développement durable*, une citation issue d'un des écrits laissés par *Anita Conti*¹ [35], qui (alors qu'elle n'avait, à cette époque, pas encore le droit de vote en France) exprime en substance les enjeux actuels du *DD* : "[...] *Nous sommes aujourd'hui les gérants, fugacement passagers, de terres, d'airs et d'eaux qui devront nourrir les foules de l'avenir. En conséquence, il faut léguer à l'avenir un domaine correctement entretenu. Cela nous interdit de polluer l'air, qui est notre premier aliment puisque nous respirons les précieux éléments qu'il contient. Cela nous interdit également d'altérer la valeur de l'humus dans lequel poussent nos végétaux nourriciers, et l'on sait bien que cet humus si vulnérable ne représente qu'une infime pellicule sur la croûte du globe terrestre. Il y a aussi les eaux de sources, de puits, et le volume énorme des eaux océaniques d'où monte la vapeur qui sera la bienfaisante pluie.*

Cette énorme volume océanique contient des animaux variés et l'on a cru longtemps que les espèces animales en évolution dans la mer étaient si nombreuses et abondantes qu'elles étaient inépuisables.

Forts de cette croyance illusoire, les gros bateaux modernes ont pêché de leur mieux, avec de puissants filets, et progressivement ils ont dépeuplé les zones d'exploitation proches de leurs ports d'armements. Pour trouver d'autres zones intactes il fallut aller plus loin, et de plus gros bateaux furent construits...On voit la suite : comme la Terre est ronde on ne peut pas aller

¹Anita Conti (1899-1997) était écrivain, photographe, cinéaste et biologiste. En 1935, elle fut une pionnière en tant que première femme océanographe au service de l'Office Scientifique et Technique des Pêches Maritimes (OSTPM). De ses missions, à Terre-Neuve jusqu'en Afrique (pour l'évaluation des ressources halieutiques, les techniques de pêche, la température des eaux, la salinité, ...), elle a ramené des écrits et des photos qui retranscrivent l'histoire de la pêche et du développement de l'exploitation des ressources marines au long du 20 ième siècle [35, 34, 36].

toujours plus loin, car il y a nécessairement un instant où l'on revient à son point de départ. Cette évidence paraît comique : en vérité elle est grave, et pose des problèmes techniques innombrables, ainsi que de subtiles questions de droit maritime international. A qui appartiennent les poissons qui évoluent avec le déplacement des eaux qui les contiennent ? Appartiennent-ils au sol sous-marin du pays côtier ? Ou au contraire à l'eau en mouvement ? A-t-on le droit d'installer en mer libre n'importe quel système de pêche ? Et ce pillage inconscient va-t-il résoudre le problème de malnutrition des populations déshéritées ? [...]".

A en croire les chalutages actuels pour exploiter les ressources halieutiques des fonds abyssaux, beaucoup de réponses à ces questions ont été soigneusement ignorées. En attendant la prise en compte de ces considérations, l'environnement évolue d'un pas assuré vers un avenir plus qu'incertain.



FIG. 2.1 – Redéfinir le rapport entre l'homme et les ressources naturelles (Photo : Anita Conti)

Cependant, dans l'action politique, la germination d'une réponse formelle à ces inquiétudes a été plus lente. Aujourd'hui, ces mises en garde scientifiques, d'ordre écologique, sont reconnues pour leur importance, d'ordre économique, et ont engendré, avec un ensemble d'autres considérations relatives au développement socio-économique, la naissance du concept de développement durable et, par la suite, celle du concept de GIZC.

En écho à la citation précédente, il est intéressant de lire à présent les dernières lignes de la conclusion du premier guide méthodologique pour la GIZC de l'UNESCO [170, 94] paru en 1997 : "*On assiste déjà à un changement de comportement de la part du public qui revendique le droit de savoir ce que contient l'air qu'il respire, l'eau qu'il boit ou le sol sur lequel il vit. A mesure que s'affinera sa compréhension de l'environnement et des principales **interactions** qui déterminent la qualité de la vie, il demandera des comptes de plus en plus détaillés*".

Certes les enjeux du DD ne se limitent pas au contenu de ces citations mais il s'agit bien en substance, comme cela est suggéré, de changer les modes d'exploitation actuels, de redéfinir la manière dont l'homme interagit avec l'environnement naturel, et plus particulièrement pour notre part avec la zone côtière (voir Figure 2.1).

Remarques :

- L'ensemble des points développés dans cet épilogue est fort à propos pour introduire le sujet qui nous intéresse et nous verrons que cette synthèse visionnaire peut encore éclairer bien des parties que nous présenterons par la suite, dans ce document, et nous ne manquerons pas d'y faire référence à nouveau.
- L'idée n'est certes pas neuve, c'est, en fait, une bonne intention que Saint Exupéry a également formalisée en disant : "*Nous n'héritons pas de la Terre de nos ancêtres, nous l'empruntons à nos enfants*" et d'autres encore avant lui à n'en pas douter.

Après avoir introduit ces quelques idées fondatrices, nous allons présenter, dans la partie suivante, du point de vue historique, quelques étapes essentielles dans l'évolution de la formalisation de ce concept et dans la mise en place de réponses politiques à ces idées.

Historique et cadre conceptuel du DD

Dans son acception actuelle, le *développement durable* est un mode de développement économique en rupture avec les précédents qui cherche à équilibrer le rapport entre progrès économique et social et préservation de l'environnement. Ce dernier étant considéré comme un patrimoine à transmettre aux générations futures [179, 52].

C'est à travers la création du *Club de Rome*² que le concept de *croissance zéro*³ est apparu en 1970. De 1970 à 1987, cette approche a fait son chemin.

En 1972, un rapport du Club de Rome a pour la première fois abordé ces problèmes d'une *manière systématique* et consistante donnant lieu aux célèbres ouvrages "*Les limites de la croissance*" [124], et, plus tard, "*Au delà des limites*" [123]. La même année, lors de la *Conférence de Stockholm*, les Nations Unies élaborent un modèle de développement respectueux de l'environnement à travers une gestion efficace des ressources naturelles.

L'expression *développement durable* a finalement été proposée pour la première fois en 1980 dans la *stratégie mondiale de la conservation* publiée par l'Union Internationale pour la Conservation de la Nature (UICN), devenue depuis l'Union mondiale pour la nature, le Fonds mondial pour la nature (WWF) et le Programme des Nations Unies pour l'Environnement (PNUE).

C'est en 1987 que le terme *durable* ou *viable* a pris son sens actuel à travers une réponse politique formulée par la commission mondiale sur l'environnement et le développement dans le *rapport Brundtland*⁴[25, 33] (*Notre avenir à tous*) où apparaissent les notions de *synergie* entre développement économique et social, protection de l'environnement et conservation des ressources naturelles. Il y est défini comme un "*développement qui répond aux besoins du présent sans compromettre les capacités des générations futures à répondre aux leurs*". Certains préfèrent parler de *développement soutenable*, c'est-à-dire ce que notre environnement peut supporter sur le long terme, ce qui serait aussi une traduction plus littérale du terme anglophone *sustainable development*.

²Le Club de Rome est une association internationale, ONG fondée le 8 avril 1968 et organisée autour d'un groupe d'experts internationaux (scientifiques, économistes, hommes d'affaires ou d'états des cinq continents) qui souhaitent que la recherche s'empare du problème de l'évolution du monde pris dans sa globalité pour tenter de cerner les limites de la croissance ; <http://www.clubofrome.org/>.

³En 1970, le Club de Rome lançait l'idée de la *croissance zéro*, en réaction à une croissance économique et démographique exponentielle et à la surexploitation des ressources.

⁴Gro Harlem Brundtland (née le 20 avril 1939) : femme politique norvégienne. Chef du parti travailliste, ministre de l'environnement de 1974 à 1979, première femme premier ministre en Norvège en 1981, puis de nouveau de 1986 jusqu'en 1989, et de 1990 jusqu'en 1996. Elle est à l'origine du concept de développement durable.

En 1991, la nouvelle stratégie de conservation de la nature publiée par l'UICN, le WWF et le PNUE définit le développement durable comme : *"le fait d'améliorer les conditions d'existence des communautés humaines, tout en restant dans les limites de la capacité de charge des écosystèmes"*.

Une synthèse possible de ces définitions pourrait être : *"le DD a pour objectif de rester dans les limites soutenables des interactions entre l'homme et les ressources naturelles pour définir, en fonction de chaque espace anthropisé, un équilibre de développement optimal pour l'ensemble des éléments du système"*⁵.

Voici donc le cadre général théorique du développement durable, nous allons à présent nous intéresser, plus concrètement, à l'opérationnalisation de ce concept.

Les enjeux du DD et l'accès à l'information

Pour opérationnaliser ce nouveau mode de développement, la synthèse des idées débattues lors de conférences ou dans les ouvrages dédiés à ce domaine a été déclinée autour de grands principes qui doivent structurer les plans d'action du DD. Après avoir brièvement présenté ces principes, nous verrons comment, malgré ces efforts de formalisation, la méthodologie générale d'application du DD reste encore floue.

DD point par point (Principes généraux) L'ensemble des réflexions menées depuis 1970 est synthétisé dans une série de principes nécessaires à la bonne marche d'un programme de DD.

Les conclusions de la *conférence des Nations Unies sur l'Environnement et le Développement de Rio de Janeiro en Juin 1992* se déclinent en 27 principes parmi lesquels on retrouvera les idées véhiculées par des discours scientifiques similaires à ceux d'Anita Conti, mais aussi, plus largement, à titre d'exemples et sans être exhaustif, des directives relatives à la *nécessaire égalité entre hommes et femmes, Développement centré sur l'être humain, Souveraineté et responsabilité, Équité inter- et intragénérationnelle, Intégration de l'environnement aux autres dimensions du développement, Élimination de la pauvreté, Besoins particuliers des pays en développement, Responsabilités communes mais différenciées, Consommation et production soutenables et politiques démographiques appropriées, ...*⁶.

Le *plan d'action global Action 21* est l'une des pièces maîtresses issues de la Conférence de Rio. Adopté par 179 nations, ce *plan* traite de 38 sujets (aux thématiques aussi variées que celles présentées à Rio) et constitue un document de référence pouvant inspirer les pays dans la mise en œuvre du développement durable.

Recommandations spécifiques à la gestion de l'information et du savoir Pour ce qui nous concerne, nous allons mettre de côté l'essentiel de ces principes pour nous intéresser prioritairement aux réflexions et principes consacrés à l'accès au savoir et à l'information. Parmi les principes de Rio, certains mettent un accent particulier sur ceux auxquels notre travail se réfère : l'accès au savoir par l'optimisation de la gestion de l'information et de la connaissance. Il est en effet stipulé que les mesures favorisant l'éducation et l'accès à l'information doivent être encouragées de manière à stimuler l'innovation ainsi qu'à améliorer la sensibilisation et la participation effective du public à la mise en œuvre du DD [146]. On peut,

⁵Certains comme [127] reprochent aux définitions de référence du DD une vision trop anthropocentrique.

⁶http://www.geneve.ch/agenda21/pme/annexe_a.asp.

pour comprendre la démarche, évoquer l'adage qui dit que l'information est le pouvoir, en ajoutant que ne pas la diffuser c'est le conserver. Offrir l'information au grand public c'est donc lui transférer une part de pouvoir, l'impliquer dans le développement et la prise de décision en soutenant la participation citoyenne au débat politique (ce qui est cohérent avec les objectifs du concept de gouvernance⁷).

Trois principes énoncés à Rio concernent le sujet qui nous intéresse [139, 6] :

1. **Principe 9 : Coopération en matière de renforcement des capacités** Les Etats devraient coopérer ou intensifier le renforcement des capacités endogènes en matière de développement durable *en améliorant la compréhension scientifique par des échanges de connaissances scientifiques et techniques et en facilitant la mise au point, l'adaptation, la diffusion et le transfert de techniques*, y compris de techniques nouvelles et novatrices.
2. **Principe 10 : Accès à l'information et participation aux décisions** : La meilleure façon de traiter les questions d'environnement est d'assurer la *participation de tous les citoyens concernés*, au niveau qui convient. *Au niveau national, chaque individu doit avoir dûment accès aux informations relatives à l'environnement que détiennent les autorités publiques*, y compris aux informations relatives aux substances et activités dangereuses dans leurs collectivités, et avoir la possibilité de participer aux processus de prise de décision. *Les Etats doivent faciliter et encourager la sensibilisation et la participation du public en mettant les informations à la disposition de celui-ci*. Un accès effectif à des actions judiciaires et administratives, notamment des réparations et des recours, doit être assuré.
3. **Principe 19 : Communication d'information transfrontalière** : Les États doivent prévenir suffisamment à l'avance les États susceptibles d'être affectés et leur *communiquer toutes informations pertinentes sur les activités* qui peuvent avoir des effets transfrontières sérieusement nocifs sur l'environnement et mener des consultations avec ces États rapidement et de bonne foi.

Des conclusions similaires à celles de la conférence de Rio sont également reprises dans *Action 21* et la *Convention d'Aarhus*⁸ en 1998 (notamment les **articles 4 et 5**), ainsi que dans la déclaration concernant le volet de l'information pour la prise de décision du *Sommet Mondial sur le Développement durable de Johannesburg* en 2002, qui insiste sur l'importance de la gestion de l'information dans les démarches de DD : "*dans le cadre du développement durable, chacun est un utilisateur et un fournisseur d'informations, au sens large. Il faut entendre par là, des données, des renseignements, des expériences présentées de façon appropriée et des connaissances. Le besoin d'information se fait sentir à tous les niveaux, du niveau national et international chez les principaux décideurs au niveau local et à celui de l'individu..."*

Finalement, pour celui à qui revient la charge de proposer une gestion pertinente de l'information dans un programme de DD, il s'agit donc de faciliter, par une "**présentation appro-**

⁷La substitution de "gouvernance" à des notions telles que "gouvernement" ou "gestion" est révélatrice d'une évolution significative dans l'exercice du pouvoir. La gouvernance se distingue de la notion traditionnelle de "gouvernement", qui reste associée dans la langue française à une forme organisée, rationnelle et cohérente du pouvoir. La gouvernance, par ailleurs, renvoie à un processus de coordination qui permet à l'exercice des pouvoirs politiques, économiques et administratifs de s'effectuer à tous les niveaux de la structure du système national, régional et local par différents acteurs disposant à des degrés divers de pouvoirs de décision. Elle se traduit donc concrètement par une participation accrue de la société civile organisée à l'élaboration des décisions et à leur mise en œuvre.

⁸Convention sur l'accès à l'information, la participation du public au processus décisionnel et l'accès à la justice en matière d'environnement.

prise", le partage et l'intégration des connaissances et des ressources existantes dans diverses disciplines scientifiques afin de permettre, à terme, de traiter au mieux et tous ensemble les questions complexes soulevées par le DD.

Cependant, derrière ces recommandations générales et théoriques auxquelles tous généralement adhèrent, l'application concrète de l'ensemble de ces principes se fait au cas par cas selon l'espace géographique d'application qui détermine les éléments physiques et humains et leurs interactions dans le système.

DD au cas par cas

Depuis l'apparition du concept de DD, de nombreuses autres définitions ont été proposées pour opérationnaliser les recommandations générales que nous venons d'évoquer. Cependant, aucune ne peut satisfaire simultanément les besoins légitimes de tous les individus, des peuples, des entreprises ou des organisations qui cherchent à le mettre en œuvre dans une grande variété de territoires. Toutefois, quelle que soit la définition utilisée pour ce concept et la façon de le mettre en application, l'*intégration* en un tout opérationnel des dimensions sociales, économiques et environnementales du développement reste toujours le fondement pour l'atteinte d'un développement qui est durable [52].

Ceci revient à dire que si l'application du DD ne peut être supportée par une méthodologie universelle (mais se décline au cas par cas, par une réévaluation de l'ensemble des mesures nécessaires à ce mode de développement, en fonction de l'espace d'application et des potentialités offertes par les ressources naturelles présentes, les sociétés humaines impliquées et leurs modes de développement socio-économiques), il demeure néanmoins un trame commune à chaque cas d'application, qui consiste à intégrer les principaux éléments humains et écologiques en interactions et à considérer nécessairement pour apporter une réponse, en terme de développement, qui respecte le fonctionnement naturel du système. Or, si l'intégration de ces différentes dimensions est essentielle, nous verrons qu'elle nécessite un décloisonnement entre organismes et acteurs (gestionnaires, politiques, scientifiques...).

En pratique, le facteur de variation principal, d'un cas de DD à l'autre, est l'espace. En effet, chaque espace est caractérisé par les éléments de géographie physique et humaine qui caractérisent les milieux. Ainsi, selon la géographie de l'espace concerné, les possibilités d'application du DD varient selon que l'on se situe, par exemple, dans les pays riches ou dans les pays pauvres et selon la quantité et la qualité des ressources naturelles présentes. A ce propos, nous rappelons que l'application du DD à l'échelle globale pose des problèmes éthiques de droits d'interdiction des pays riches aux pays pauvres de certains modes de développement (par le contrôle des émissions de CO₂ par exemple). En effet, le DD a un coût que tous les pays ne peuvent pas forcément assumer de la même manière.

Par contre, l'application du DD à l'échelle nationale, pour un pays riche, n'est pas soumise à controverse dans la mesure où un tel Etat peut, à l'intérieur de ses frontières, décider librement d'un mode de développement durable respectueux des recommandations internationales. Charité bien ordonnée commençant par soi-même, c'est à travers les propres efforts des pays riches qu'on pourra mesurer leur bonne foi et volonté à mettre en pratique les principes qu'ils défendent.

Nous allons désormais orienter notre présentation sur le cas particulier de DD auquel nous allons nous intéresser spécifiquement pour notre travail de gestion de l'information : la *Gestion Intégrée de la Zone Côtière*. En effet, nous allons voir que la GIZC est une sorte de DD appliquée à l'espace littoral et qu'ainsi, ce concept hérite donc des propriétés et méthodes

attribuées au DD en général, y-compris d'un certain flou opérationnel au cas par cas. Du point de vue de l'*accès au savoir* en particulier, nous verrons que la GIZC porte une attention particulière à la mise en place de systèmes d'information et identifie des tâches principales pour l'amélioration des systèmes en place.

2.1.2 GIZC : un concept flou sur un espace flou

A la suite de ce que nous venons de présenter en partie 2.1.1, la GIZC peut être appréhendée comme une application des principes du développement durable à l'environnement côtier. En des termes plus informatiques, la gestion intégrée de la zone côtière est *une sorte de DD*, ou une *instanciation du DD spatialisée sur la zone côtière*.

Nous allons donc nous consacrer à décrire ce qui fait de la zone côtière un espace si particulier d'application du DD, à l'échelle planétaire, qu'il est justifié de lui dédier un concept particulier. Puis nous établirons une synthèse générale des définitions et enjeux de ce concept (en particulier l'importance de l'*espace* et de l'*information* pour le succès d'un processus de GIZC). Enfin, un inventaire des recommandations spécifiques à la gestion de l'information, issues de la littérature du DD et de la GIZC, nous permettra de situer au mieux notre approche de gestion de l'information parmi l'ensemble des objectifs nécessaires à la réalisation d'un programme de GIZC.

Caractéristiques générales de la zone côtière

Des espaces d'application possibles pour le développement durable (voir partie 2.1.1), l'espace littoral est un de ceux qui rassemble le plus des problématiques énoncées dans les principes de Rio. Par ses richesses écologiques et humaines, il illustre, voire exacerbe, certains enjeux du DD.

Un espace naturellement riche et convoité La zone côtière est un espace particulier de transition entre la terre et la mer (que certains qualifient de continuum terre-mer, voir figure 2.2), écologiquement et humainement riche, aux limites géographiques floues [50].

Zone mythique (frontière de fantasmes, de peurs ou de rêves), ligne d'ouverture vers un "au-delà" longtemps mystérieux à travers les mers et océans qui représentent 2/3 tiers de la surface planétaire, la zone côtière possède, en outre, les écosystèmes les plus productifs de la planète (récifs coralliens) particulièrement propice au développement humain d'un point de vue économique avec de nombreux échanges commerciaux mais aussi culturels. Le patrimoine maritime et terrestre y est important, et les possibilités de loisirs uniques. Exploitées comme telles depuis l'antiquité, la pression démographique a augmenté à tel point qu'actuellement 2/3 de la population mondiale est concentrée à moins de 50 km du trait de côte et cet espace est toujours plus convoité depuis un demi-siècle.

Zone de productions et d'échanges économiques intenses à forte qualité de vie, les zones côtières soumises à une urbanisation inquiétante, sont à présent des espaces convoités par des acteurs aux intérêts souvent divergents. En revanche, à la suite de ces modes de développement humain, l'état écologique du littoral est devenu alarmant et les capacités d'accueil sont désormais restreintes.

L'indicateur le plus évident de la nécessité d'un changement de gestion pour cet espace, face à cette pression démographique croissante, a été l'apparition de *conflits d'usages* [51] entre acteurs et activités. Ils illustrent les convoitises actuelles associées à l'attribution des ressources

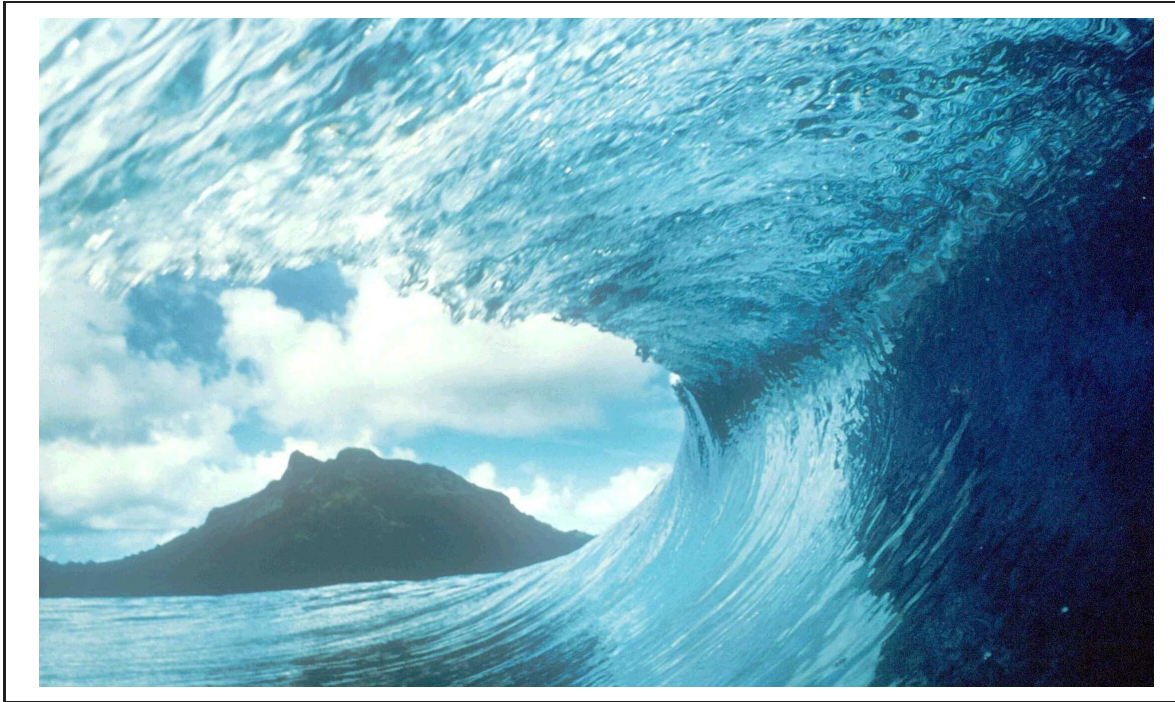


FIG. 2.2 – Zone côtière : continuum terre-mer

littorales. En particulier, la raréfaction des espaces disponibles pour la pratique d'activités humaines est telle que les conflits d'usages associés à la *consommation d'espace* augmentent considérablement avec le temps.

Ainsi, au nom du principe de respect de la capacité de charge des écosystèmes que nous avons évoqués dans la partie 2.1.1, cet espace est prioritaire pour mettre en œuvre les principes de DD. On comprend mieux alors, au vue des enjeux et des spécificités liés à cet espace, qu'un nouveau concept lui soit dédié.

Voyons à présent, plus en détail, quelles sont les caractéristiques des acteurs de la zone côtière à mobiliser dans les projets de GIZC.

Un domaine humainement, thématiquement et spatialement hétérogène Humainement, la zone côtière est la rencontre des communautés terrestres et maritimes, traditionnellement séparées. La réunion de ces populations distinctes fait cohabiter des modes d'exploitation ou de productions économiques variés ainsi que différentes activités scientifiques ou culturelles. La gestion *intégrée* de la zone côtière est, pour ces raisons, un cas complexe d'application du développement durable. En effet, cette sectorisation des activités a, plus qu'ailleurs, généré des différences de vocabulaires, de coutumes ou de méthodes propres à deux mondes qui jadis se côtoyaient peu et doivent désormais apprendre à faire route ensemble puisqu'à présent intégrés et considérés comme éléments en interaction au sein d'un même système.

Or, la réussite du DD (et donc la GIZC) repose en partie sur le *décloisonnement* des approches et des institutions [51]. La confrontation des acteurs et des thématiques associées sont des facteurs d'intégration au service d'une meilleure gouvernance. On comprend alors que la prise en compte de l'importante variété d'acteurs (et donc d'organismes) impliqués sur la zone

côtière ainsi que les thématiques rassemblées dans les différents programmes est capitale pour la réussite d'un processus de GIZC (d'où une grande diversité de méthodes/outils/ressources..). En effet, la connaissance des caractéristiques et usages associés à chaque catégorie d'acteurs (et organismes associés) est indispensable à la bonne intégration des composantes essentielles à la gestion d'une zone côtière donnée.

Pour cela, le guide méthodologique 37 de la Commission Océanographique Internationale⁹ (COI) [170] propose de distinguer différentes sortes d'acteurs en fonction de leur domaine d'appartenance :

- le **domaine environnemental** avec les "acteurs scientifiques intervenant ou contribuant à la connaissance des milieux et des hommes". Ils appartiennent essentiellement à des *organismes de recherche* (*universités, instituts, observatoires, stations spécialisées dans le domaine marin côtier*), à des *bureaux d'études*, des *associations non gouvernementales de défense ou de protection de la nature...*,
- le **domaine socio-économique** avec des utilisateurs du milieu côtier qui sont soit des *agents économiques* (pêcheurs ou conchyliculteurs par exemple) soit des *usagers* fréquentant l'espace littoral à titre permanent (habitation) ou temporaires (loisirs).
- le **domaine administratif** qui réunit *gestionnaires, décideurs et planificateurs*. Ils appartiennent aux *services de l'état*, aux *collectivités territoriales* ou assurent la tutelle des diverses activités exercées sur le littoral.

Cette diversité d'acteurs a été illustrée ultérieurement en 2002 dans le guide méthodologique 42 de la COI [87] (voir Figure 2.3).

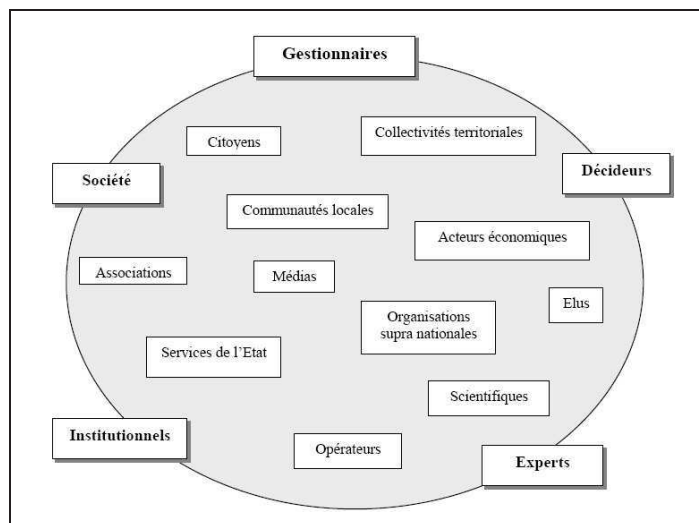


FIG. 2.3 – "La ronde des acteurs" dans la GIZC d'après [87]

On peut au passage relever dans cette Figure que l'interprétation de la mise en forme de l'information est ambiguë (pour ne pas dire inappropriée, en écho aux recommandations du DD). Ceci illustre l'indissociable nécessité de partager des référentiels en même temps que les formalismes communs qui les présentent aux acteurs dans le contexte pluridisciplinaire de la GIZC.

⁹<http://ioc.unesco.org/iocweb/index.php>.

En particulier, du point de vue de l'échange d'information pour la GIZC, cette démarche d'identification des acteurs est fondamentale [170] : "***Tous les systèmes producteurs de données, détenus par les diverses organisations locales, publiques ou privées, à caractère scientifique ou administratif, etc., doivent être identifiés pour mettre en place les passerelles nécessaires à l'échange de données***". Nous proposerons en nous appuyant sur les sources de référence du domaine une typologie d'acteurs dans la partie 5.3.2.

Mais cette diversité d'acteurs est à l'origine en GIZC, plus encore que dans d'autres cas d'application du DD, de difficultés quant à la mise en œuvre de ce mode de gestion, notamment de lourds problèmes terminologiques pour le partage des informations et des connaissances apportées par les acteurs.

Théorie et pratique de la GIZC

Nous allons présenter ici des définitions plus ou moins opérationnelles de la GIZC. En effet, la GIZC hérite d'un certain flou du DD, d'autant plus, qu'historiquement, on a voulu, à tort, appliquer à ce milieu complexe et méconnu les mêmes modes de gestion qu'au milieu terrestre¹⁰.

Définitions et enjeux de la GIZC D'après [33], la GIZC aide à préserver la nature et ses ressources pour les générations futures, elle aide à préserver et promouvoir l'équité sociale et aide à protéger l'usage et les droits traditionnels et l'accès équitable aux ressources côtières. La GIZC aide donc littéralement à appliquer, pour la zone côtière, les principes généraux du DD et de la déclaration de Rio en particulier. Nous ne reviendrons pas ici sur l'évolution historique de la politique environnementale qui a conduit au développement durable, qui a été présentée dans la partie précédente.

Nous retiendrons qu'on peut attribuer à la commission scientifique de la conférence générale de l'UNESCO réunie en 1993 la parenté du concept de gestion intégrée de la zone côtière qui prend corps à travers le guide méthodologique 37 de la COI [170].

Une définition normative et canonique de la GIZC pourrait être celle-ci : c'est un "*processus dynamique qui réunit gouvernements et sociétés, sciences et décideurs, intérêts publics et privés en vue de la protection et du développement des systèmes et ressources côtières ; ce processus vise à optimiser les choix à long terme privilégiant les ressources et leur usage raisonné et raisonnable*" [30] repris par [87, 170].

Cela signifie implicitement qu'une meilleure gestion passe par un décloisonnement des institutions et une participation de l'ensemble des acteurs. En outre, avec de tels objectifs [87] note que, plus encore que dans l'énoncé général de la problématique du DD, il y a, pour le processus de GIZC, nécessité de définir une ***méthodologie commune*** qui sera utilisée pour ***décrire la complexité des interactions entre le système des ressources et ses utilisateurs potentiels***. Il y a nécessité également de planifier et de contrôler ce processus d'une façon ***systématique et durable***.

Néanmoins, en pratique, comme pour la mise en œuvre du DD, on peut remarquer que la GIZC hérite, malgré l'unanimité autour de sa définition et de ses enjeux, d'un flou opérationnel d'autant plus important que les problèmes à résoudre sont amplifiés par les particularismes importants de l'espace côtier.

¹⁰ //asso.objectifs-sciences.com.

Le flou de la GIZC Du DD, la GIZC hérite des difficultés relatives à la définition de sa portée opérationnelle [18]. [27] parle de "concept flou ", de même, le Conseil Economique et Social du Conseil Régional de Bretagne [28] parle de la part de certains acteurs d'une "perception plus symbolique, philosophique, que pratique " de ce concept. A ceci s'ajoute une diversité d'acronymes et de définitions qui nuit globalement à la définition d'un cadre de référence stabilisé pour la gestion du littoral (Gestion Intégrée de la Zone Côtière - GIZC ; Gestion Intégrée du Littoral - GIL ; Gestion Intégrée du Littoral et des bassins Fluviaux - GILIF ; Aménagement Intégrée de la Zone Côtière - AIZC ; ...), et ce, tant d'un point de vue spatial que thématique. Retenons qu'il y a convergence des analyses sur le fait que ce concept s'appuie sur l'intégration exhaustive, généralement par une *approche systémique* (voir partie 5.2.2), des thématiques impliquées ainsi que sur la définition d'*unités géographiques cohérentes*.

Cela se traduit concrètement par des difficultés d'appropriation de la part du public, manifestées par exemple, lors de rencontres, par des prises à parti des scientifiques ou gestionnaires pour demander la mise en œuvre de mesures concrètes, et plus largement par une demande publique d'action urgente face à une théorie traînante.

Il est néanmoins possible, dans un premier temps, de contribuer à rendre opérationnelle cette démarche en luttant contre les déficits d'accès à l'information et à la connaissance qui sont unanimement reconnus dans la GIZC.

De plus, en pratique, la GIZC s'appuie malgré tout sur des outils de politiques publiques et des unités spatiales de gestion qui lui sont propres.

La GIZC ou la mise en œuvre d'une politique de gestion territorialisée

D'une manière générale, la place de l'espace est capitale dans la GIZC. Le concept fondamental de *zonage* traduit la nécessaire adaptation des unités de gestion aux outils de politiques publiques dont disposent les gestionnaires. A ce sujet, le guide UNESCO [170] consacre son deuxième chapitre à la définition d'*unités cohérentes de gestion*.

Dans le cadre français et dans la logique actuelle du développement durable, il s'agit de réaler les principes d'aménagement, de protection et de mise en valeur du littoral ¹¹ à l'échelle d'un territoire pertinent d'intégration des problématiques et des objectifs [18]. La gestion intégrée des territoires, mise en œuvre par des *dispositifs* spécifiques, est avant tout une politique dépassant la simple gestion intersectorielle avec des orientations environnementales [152, 153]. Des actions de programmation et de projet doivent être poursuivies et sont à présent reconnues au titre de projets de territoires privilégiant l'intercommunalité en termes de *contractualisation des politiques publiques*. Outre les échelles locales auxquelles ces projets font nécessairement appel, les questionnements autour de la définition de territoires de références ainsi que de la prise en compte de plusieurs niveaux d'analyse et de leurs recoupements sont ici centraux (avec notamment la définition de zones de concernement des populations [151]). La finalité correspond ici à la détermination d'*unités de gestion* en fonction des problématiques présentes et des réponses possibles sur l'espace considéré.

Les *instruments de gestion et d'organisation de l'espace*, à savoir les *dispositifs institutionnels* que sont les documents de planification et les démarches contractuelles (respectivement les *Schémas de COhérence Territoriale (SCOT)*, les *Plans Locaux d'Urbanisme (PLU)*, les *Pays*, etc.), constituent les bases de la définition de projets intégrés établis de manière concertée ; toutefois, si certains sont intégrateurs (les SCOT par exemple), ils n'ont pas été pensés pour

¹¹Ces principes ont été promus par la loi Littoral n.86-2 du 3 Janvier 1986.

l'espace littoral. La contractualisation des politiques publiques devrait être instrumentalisée par des *Pays Maritimes* et la création d'*Agglomérations Maritimes* en tant que *projets de territoire côtiers*. Le *Comité Interministériel d'Aménagement et de Développement du Territoire (CIADT)* en 2001 a ainsi reconnu les "*Pays Maritimes et Côtiers* " dont les orientations des chartes les régissant se distinguent des Pays terrestres.

Or, comme nous l'avons évoqués, la GIZC ainsi appréhendée doit être matérialisée par une gestion négociée avec une approche participative (institutionnalisation du passage d'une démocratie représentative à une démocratie participative) offrant notamment au travers de phases de concertation et de médiation des éléments concrets et pérennes de régulation de conflits d'usages existants.

Dès lors, une question essentielle apparaît : quels sont les besoins d'information dans un processus de GIZC et en quoi sont-ils essentiels pour la territorialisation d'une politique de gestion ? En premier lieu, il semble que la mobilisation et la coopération de certains acteurs pour l'action collective ne nécessitent pas forcément d'appuis techniques informatisés. En revanche, certaines catégories d'acteurs jouant un rôle de clé de voûte dans une démarche de GIZC peuvent être les premiers à tirer partis de telles avancées technologiques ; les scientifiques contribuant notamment à des diagnostics d'états et de suivi, et participant pour une large part à la réalisation de bilans sociaux-environnementaux sont les premiers à développer un réseau local d'information dans un processus GIZC. En outre, soulever l'introduction du concept de territoire pour articuler différentes échelles d'analyse nécessite l'apport d'objets géographiques pour spatialiser les données existantes et produites. Voyons à présent quel est le rôle affecté à la gestion de l'information.

2.1.3 GIZC et information

Là encore, ce qui vaut pour le DD vaut pour la GIZC (nous renvoyons donc au cadre général de gestion de l'information dans ce domaine développé en partie 2.1.1). Cependant, dans le cas de la mise en œuvre de la GIZC, la gestion de l'information prend une importance particulière si on en croit les guides méthodologiques. En effet, en avant-propos du guide méthodologique de la *COI* [170] on trouve :

*"[...] Ce guide avant tout pratique, a pour ambition d'aider à **construire un système d'information** cohérent pour l'aide à la décision. Outil indispensable à un mode de gestion intégrée de la zone côtière, ce système doit pouvoir fournir aux décideurs et aux aménageurs des éléments objectifs de choix, les conflits naissants souvent de l'absence de données et d'indicateurs pertinents. [...]"*

Avant de détailler la place de l'information dans la GIZC, nous allons, au préalable, définir la terminologie des principaux concepts du domaine de la gestion d'information.

Terminologie

D'un point de vue terminologique, afin d'éviter les confusions, nous nous appuyons sur les définitions de Doody utilisées dans une réflexion concernant l'information nécessaire pour la GIZC [47].

La **donnée** se distingue de l'**information** :

- La **donnée** est la matière première à partir de laquelle l'**information** est produite.
- L'**information** est une collection de **données** pertinente à un endroit et à un temps donnés.

Autrement dit l'**information** est de la **donnée** dans un *contexte* : elle a une *signification*, une *pertinence* et un *objectif*.

D'où la proposition suivante :

$$\begin{aligned} \text{Données} + \text{Contexte} &= \text{Information} \\ \text{Information} + \text{Analyse} &= \text{Compréhension / connaissance} \\ \text{Compréhension / connaissance} + \text{Gestion} &= \text{Possibilité d'action durable} \end{aligned}$$

Pour éviter toute ambiguïté entre *donnée* et *information*, nous utiliserons également le terme générique *ressource* ou *ressources informationnelles* qui, pour nous, représente l'*ensemble des données et informations*. Selon les organismes, les individus, on entendra par *ressources* aussi bien un rapport papier, une carte, une image, une vidéo, un jeu ou une série de données, une base de données, un modèle, une valeur mesurée par un capteur, un résultat d'interview ... [16, 15]. Nous verrons par la suite qu'il est possible d'élargir cette définition et de considérer qu'un *acteur* (et les activités qu'il pratique), un *projet* ou du *matériel d'observation* par exemple, sont également des types de *ressources* qui méritent de faire l'objet d'une information et d'être portés à connaissance de la communauté (voir partie 4.1.3).

Voyons à présent quelles sont les caractéristiques des *ressources informationnelles* de la GIZC.

Caractéristiques des *ressources informationnelles* utilisées dans les processus de GIZC

Les ressources nécessaires à la GIZC sont dispersées et produites dans des contextes fortement variables. Elles sont souvent rares, chères avec des contraintes d'accès particulières selon les organismes producteurs : il n'y a pas de cas général. De plus, elles sont de natures très hétérogènes du fait des différentes catégories d'acteurs ou d'organismes qui interviennent à des échelles spatiales et avec des méthodes différentes (voir partie 2.3). Sans chercher à être exhaustif, nous nous contenterons d'évoquer les différentes ressources qui peuvent être utilisées pour la gestion d'espaces aussi bien à l'échelle internationale, que nationale ou locale (*systèmes d'information, SGBD, télédétection / Système d'information géographique (SIG), modèles, indicateurs, ressources issues d'outils d'observation, relevés, outils de politiques publiques, de gestion ou de protection, outils législatifs, d'aménagement, ressources papiers plus traditionnelles : articles, livres, rapports...*) [33]. Ces *ressources informationnelles* sont associées à différentes catégories d'*outils* qui permettent leur production ou leur gestion.

Le lien entre *ressources* et *outils* est illustré sur le portail Web du *Programme d'Actions Prioritaires*¹² (PAP) [147] qui évoque la "*boîte à outils*" nécessaire pour un processus de GIZC et les catégories générales d'*outils* indispensables dont nous proposons une version simplifiée en Figure 2.4. Nous reviendrons en détail sur les *typologies d'outils* nécessaires à la GIZC en partie 5.3.5.

Parmi ces "*outils*" certains ont une importance capitale dans la GIZC. Les *indicateurs*, mais aussi les *SIG* ainsi que la *télédétection* font l'objet respectivement des chapitres 4 du guide spécial [171] et 5 du guide UNESCO [170]. L'importance accordée aux *SI* et aux *bases de données* en général et *SIG* en particulier (ainsi qu'à la *télédétection* pour la production

¹²http://www.pap-medclearinghouse.org/fra/Toolbox_start.asp.

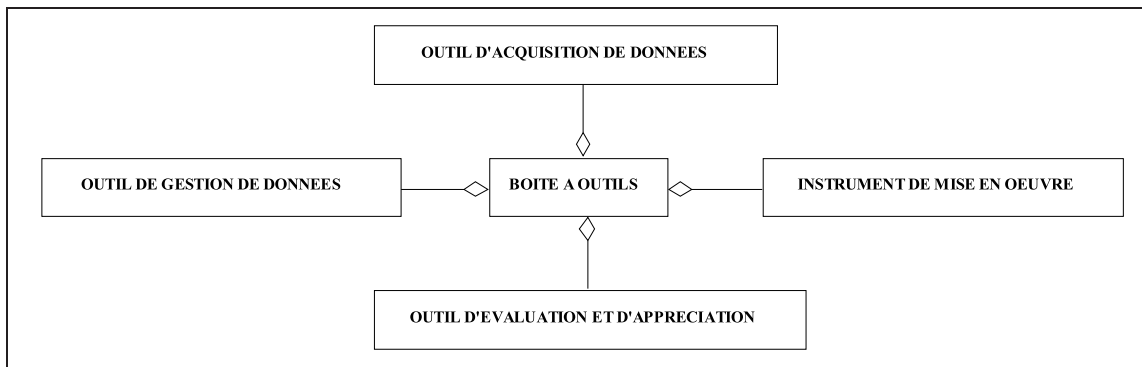


FIG. 2.4 – "Boîte à outils" pour la GIZC d'après [147]

d'information géographique en amont) répond à la place centrale de l'espace dans la GIZC (comme cela est détaillé en partie 2.1.2).

Ainsi par leur *contexte de production*, leur *nature* et leur dispersion physique, les *ressources informationnelles* de la GIZC constituent un labyrinthe d'information dans lequel on ne peut se retrouver sans expérience approfondie du domaine ou sans volonté, de la part des acteurs impliqués, de partager les caractéristiques des *ressources informationnelles* existantes.

A cette diversité de ressources produites s'ajoute une difficulté supplémentaire pour le partage des connaissances qu'elles contiennent : une forte ***hétérogénéité des formalismes de représentations de la connaissance***. Que ce soit les formalismes de représentations graphiques ou plus basiquement la sémantique des termes employés pour décrire de manière différente des processus identiques, une même ressource ou une même connaissance peut être représentée, selon les thématiques, voire selon les acteurs, sous des formes très diverses. Nous verrons par la suite que nous proposons l'emploi d'un formalisme commun pour gérer ces ambiguïtés.

Comment alors gérer ces ressources hétérogènes (sur le *fond* et sur la *forme*) et réparties pour améliorer la prise de décision ?

Nous retiendrons qu'un des enjeux majeurs du domaine pour permettre de répondre aux objectifs d'accès au savoir du DD sera donc la gestion de la diversité des ressources produites en vue de les mobiliser pour une prise de décision qui intègre l'information pertinente d'une part et, d'autre part, permette des modes de représentations partagés des connaissances qu'elles contiennent. Ces modes de représentations partagés constituent ce que nous appellerons dans la suite de notre documents des *référentiels sémantiques*¹³ (*thématiques, spatiaux, temporels...*) ou des *formalismes de représentations*.

Gestion adaptée aux ressources informationnelles pour la GIZC

Les informations et les connaissances nécessaires à la mise en œuvre de chaque cas de GIZC sont nécessaires tout au long de ces processus. Cependant, la proposition d'une solution de gestion adaptée à la GIZC implique la prise en compte de verrous spécifiques à ce domaine.

¹³En informatique, un référentiel est un ensemble structuré d'informations, utilisé pour l'exécution d'un logiciel, et constituant un cadre commun à plusieurs applications [168]. Nous élargissons cette définition du référentiel en ouvrant l'utilisation du référentiel à des acteurs humains.

Place de la gestion de l'information et de la connaissance dans la GIZC La prise en compte des *connaissances* et *ressources informationnelles* pertinentes (d'un point de vue thématique et spatial) est donc un préalable nécessaire pour *amorcer* et *soutenir* le déroulement d'un processus de GIZC. Ceci permet de répondre à la nécessité de proposer des solutions appropriées aux problématiques soulevées par chaque cas de gestion, comme le montre la partie 2.1.2. La Figure 2.5 montre la place de l'accès à l'information dans les étapes principales de la GIZC.

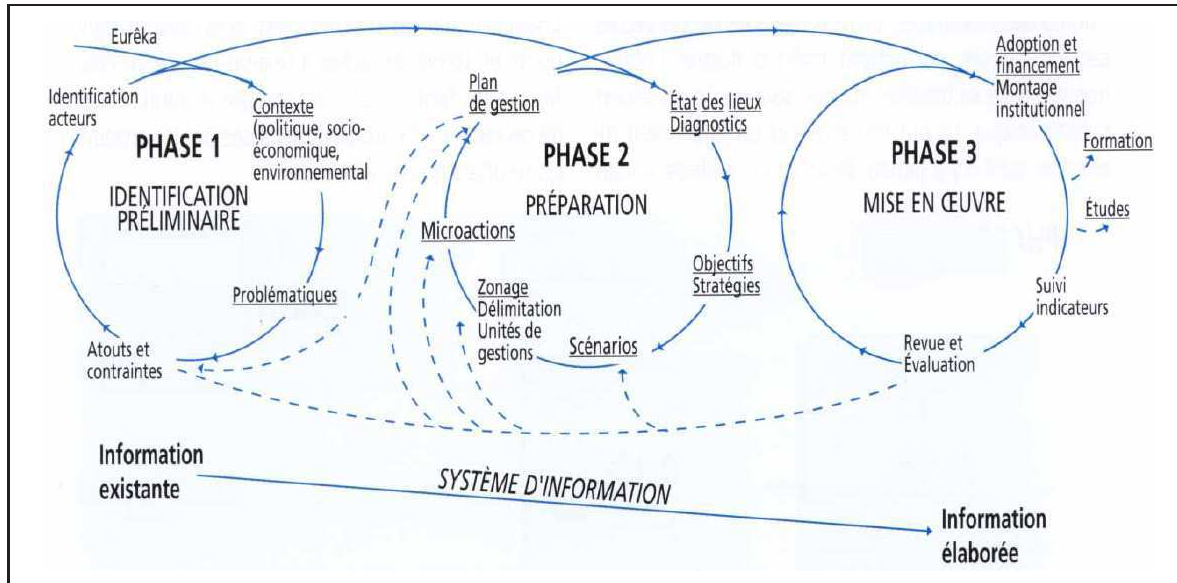


FIG. 2.5 – Place du SI dans le processus de GIZC d'après [87]

Cependant, il est en pratique difficile de *mobiliser l'information et la connaissance pertinentes* pour chaque cas de gestion. Il existe en effet, des verrous propres à ce domaine qui bloquent cette intégration.

L'ensemble des caractéristiques du *DD* et de la *GIZC* est donc à l'origine de verrous propres à ces domaines, à présent bien identifiés et pour lesquels certaines approches ou outils spécifiques apportent des solutions pertinentes.

Les verrous Si l'enjeu de l'*aide à la décision* est bien de fournir aux décideurs les éléments nécessaires à une prise de décision raisonnée, cela passe par un partage et un accès efficace aux ressources existantes. Or cette tâche complexe dans les domaines plus traditionnels est amplifiée par la diversité des *acteurs* et des *types de ressources* utilisées dans le domaine de la *GIZC* (voir parties 2.1.2 et partie 2.1.3).

Alors qu'on reconnaît fondamentalement la nécessité d'une approche pluridisciplinaire pour la *GIZC* et, plus largement dans le domaine du *DD*, on constate que la mise en place de méthodologies et d'outils informatiques pour intégrer les ressources et connaissances réparties dans différents organismes constitue un vaste sujet de recherche. Les principales difficultés recensées [15], tant du point de vue sémantique que technique, sont directement liés aux caractéristiques du domaine :

- la mise en pratique de l'*interdisciplinarité*, notamment entre sciences environnementales

et sciences humaines,

- le manque de *référentiels communs* : partage de la signification des *concepts thématiques* et *spatiaux* en particulier *via* des *vocabulaires contrôlés*, problème d'ordre *sémantique*,
- l'hétérogénéité des supports, des modes de gestion des ressources informationnelles, des formats d'échange, problèmes d'ordre *syntactique* auxquels on peut ajouter ceux liés aux *contraintes d'accès* liés au *coût* et à la *rareté* de certaines *ressources*.

Nous pouvons dès lors concevoir que si les résultats d'un processus *GIZC* sont dépendants de la bonne *intégration de l'information et de la connaissances* nécessaires à la résolution de chaque problématique, alors, les méthodologies de travail associées à un processus *GIZC* doivent se doter puis s'appuyer sur des techniques communes de recueil d'information ou de connaissance qui soutiennent les objectifs d'intégration visés. Parmi celles possibles, certaines sont prioritaires. En effet, la mobilisation physique de l'information nécessite la connaissance de l'existant. Pour le recueil des connaissances, une approche est particulièrement recommandée pour sa pertinence en *GIZC* : l'*approche écosystémique* sur laquelle nous nous appuyons pour proposer une mode de gestion de l'information et de la connaissance adapté au domaine de la *GIZC* (voir partie 5.2.2).

A présent nous proposons de justifier quelles sont les priorités actuelles en termes de gestion de ressources informationnelles de la *GIZC*.

Les priorités dans la gestion des *ressources informationnelles* de la *GIZC*

Dans le cas de la *GIZC*, l'objectif prioritaire est le partage de l'information sur les *ressources informationnelles* existantes plutôt que le partage physique des ressources elles-mêmes. En effet, face à la variété des *ressources informationnelles* existantes, dispersées chez des acteurs utilisant des systèmes de gestion de données hétérogènes, la description pour l'évaluation quantitative et qualitative des ressources disponibles de ce stock (permettant la gestion de la localisation, des droits de diffusion...) sont un préalable indispensable à leur partage physique¹⁴.

Ceci fait, il est alors possible, de restituer le contexte de production d'une donnée pour permettre, si elle convient, son partage physique en vue d'une utilisation raisonnée et pertinente dans un processus de *GIZC*. La restitution du contexte est donc essentielle pour une intégration pertinente des données essentielles à la mise en œuvre d'un processus de *GIZC*¹⁵.

Actuellement l'ensemble de ces objectifs de gestion peut être intégrés dans un nouveau type d'outil pour la *GIZC* nommé *Observatoire*.

Notion d'observatoire L'ensemble des recommandations que nous venons de présenter s'intègrent dans les objectifs d'un type particulier de *système d'information*, nommé *observatoire*.

Dans le guide méthodologique [170], la fonction d'observatoire est définie de la manière suivante : "*L'observatoire environnemental contribue à une meilleure compréhension des situations et des tendances intéressant l'ensemble des relations entre l'environnement et le développement. Il doit apporter aux décideurs les éléments d'information objective permettant*

¹⁴C'est une question de bon sens, il en va de même pour la gestion durable d'une ressource naturelle (halieutique, forestière...).

¹⁵Nous verrons par la suite que les métadonnées apportent une réponse pertinente en permettant de décrire le contexte en complément de la donnée brute, et contribuent donc à une meilleure gestion de l'information.

d'orienter leurs actions, de fournir les indications (indicateurs) nécessaires pour dimensionner et déterminer des actions de protection. Il est le lieu privilégié d'intégration pour une **compréhension plus globale des processus** : développement et validation des modèles, obtention des paramètres clefs d'évolutions sur le long terme, évaluation de l'impact des activités anthropiques sur le milieu physique. Il doit permettre d'appréhender les aléas naturels au regard des activités humaines, de mettre au point et de valider des outils et des méthodes de prévention, d'alerte et de remèdes aux catastrophes naturelles. Enfin, il doit **faciliter les liens entre les acquis de la recherche (l'amont) et les besoins en information (l'aval), par le transfert des informations vers l'ensemble des utilisateurs**".

Pour Doody [47], l'observatoire intervient pour la gestion de l'information au niveau local ou régional. "L'approche observatoire procure. . . Il forme une plaque tournante qui amène de la cohérence dans l'espace concerné, aide à identifier les sources d'information, **donne la possibilité de décrire et rechercher l'information**, et contribue à valider l'information dans les niveaux supérieurs. Ce modèle pré-suppose que les systèmes d'information et de données locaux et régionaux procurent les unités élémentaires pour la mise en place d'un système d'information sur l'état de la côte. **Ce système doit intégrer tous les secteurs concernés par la GIZC**".

Synthèse de la section 2.1

- La GIZC est une application territorialisée des principes de mise en place du *Développement Durable*. Parmi ces principes, la gestion de l'information et de la connaissance est essentielle pour amorcer et assister les processus de mise en œuvre de ces nouveaux modes de gestion.
- Cette gestion de l'information et de la connaissance pour la GIZC s'appuie sur les autres principes du DD et doit se réaliser dans le souci de porter à connaissance de tous les publics les informations et les connaissances existantes.
- Afin de faciliter cette tâche, la mutualisation d'information et de connaissances doit se réaliser sous une forme / représentation compréhensible par tous (avec des représentations communes / partagées, tant sur le fond que sur la forme).
- Ainsi la participation de tous devient possible (processus participatif permis par le décloisonnement) et conduit à une amélioration des modalités de prise de décision.
- L'hétérogénéité des acteurs et des ressources informationnelles complexifie la gestion de l'information et de la connaissance, les difficultés d'échanges terminologiques inhérentes aux travaux interdisciplinaires sont exacerbés dans le contexte de la GIZC par l'absence de référentiels sémantiques (référentiels thématique et spatial).
- En outre, les données sont dispersées et peu documentées (en particulier la restitution du contexte de production des données, essentiel pour comprendre l'information). La description des ressources informationnelles est donc un préalable indispensable à leur partage physique, afin de pouvoir estimer leur pertinence et leur localisation et de pouvoir ainsi les utiliser dans un processus de GIZC.
- Pour pallier ces contraintes et apporter une réponse adaptée au mode de gestion territorialisée qu'est la GIZC, le support cartographique procure un mode d'échange privilégié.

Nous allons à présent nous intéresser à un cas pratique d'implémentation de la gestion de l'information et des connaissances pour la GIZC à travers le cas de la zone côtière du *Languedoc Roussillon* et du programme de recherche *Syscolag*.

2.2 Contexte local : le programme régional Syscolag

Comme nous l'avons présenté en partie 2.1.1, chaque cas de GIZC est a priori unique et caractérisé, selon l'espace concerné, par le réseau d'acteurs, les thématiques et les ressources impliquées. De plus, chaque programme de GIZC est susceptible d'apporter une réponse spécifique à une même problématique.

Nous allons donc présenter, dans cette section la problématique générale de GIZC en *Languedoc Roussillon* : les caractéristiques et enjeux liés à cet espace ainsi que la réponse apportée par le programme de recherche *Syscolag* dans lequel s'inscrit ce travail de thèse. Là encore, nous insisterons sur les attentes spécifiques à ce programme qui concerne l'élaboration d'outils informatiques au service de la démarche de GIZC envisagée (en particulier du point de vue de la gestion de l'information et de la connaissance des partenaires impliqués).

2.2.1 La zone côtière du Languedoc Roussillon

Le littoral du Languedoc Roussillon est un espace privilégié d'application de la GIZC. En effet, cet espace riche écologiquement et humainement, qui a autrefois été aménagé à l'encontre des principes de GIZC, cherche aujourd'hui à harmoniser l'aménagement de sa zone côtière alors que la pression démographique y est importante¹⁶.

Histoire de la zone côtière régionale

Certes la GIZC est un concept flou (voir partie 2.1.1) mais certains de ses ennemis sont bien identifiés, parmi lesquels on trouve, au premier rang, l'urbanisation et le tourisme de masse (destructeurs des usages traditionnels) qui font du Languedoc Roussillon un espace d'application privilégié du DD.

En effet, certaines portions de cet espace ont été historiquement sacrifiées, dans les années soixantes, au nom de l'intérêt national, sur l'autel du développement économique lié au tourisme, lorsque de Gaulle et la *mission Racine*¹⁷, leur ont données l'ambition de retenir le flux touristique massif qui se dirigeait, alors, plutôt en Espagne qu'en France. Le Languedoc Roussillon a alors été massivement aménagé, pour retenir la manne financière fuyante, de la *Camargue* jusqu'aux dernières limites régionales en terres catalanes françaises.

Depuis cette période, le Languedoc Roussillon est plus souvent cité en exemple de GIZC pour la qualité douteuse de certains de ces aménagements plutôt que pour la qualité de son environnement littoral actuel.

¹⁶L'accroissement démographique du Languedoc Roussillon est deux fois et demi supérieur à la moyenne nationale.

¹⁷Dans les années 1960, la Mission Racine (mission interministérielle d'aménagement touristique présidée par Monsieur Racine) a engagé l'aménagement de plusieurs stations balnéaires sur le littoral languedocien dont la vocation est devenue le tourisme. Ainsi, Port-Camargue et La Grande Motte sont des stations créées par la Mission Racine. Aujourd'hui, une nouvelle mission interministérielle a été créée pour faire le bilan de la première et re-qualifier le littoral languedocien [49].

Et pourtant...cette zone côtière est particulièrement riche, écologiquement et humainement, et la pression démographique est telle que la région Languedoc Roussillon est aujourd'hui au cœur d'une réflexion de GIZC.

Etat actuel

Voici la présentation qui est faite dans la convention cadre du programme *Syscolag* au sujet des caractéristiques actuelles de la zone côtière du Languedoc Roussillon :

"Le littoral du Languedoc-Roussillon constitue un patrimoine exceptionnel au niveau écologique (avec, en particulier, 40 000 hectares de zones humides) et socio-culturel. C'est un territoire particulièrement fragile et vulnérable, fruit d'un équilibre subtil entre les apports de la terre et ceux de la mer. Le développement des infrastructures de transport, l'assainissement et l'irrigation des plaines littorales, l'urbanisation des bassins versants ont entraîné, ces trente dernières années, des mutations radicales. Le littoral n'est pas seulement le lieu d'une fréquentation touristique massive durant la saison estivale. L'expansion démographique du Languedoc-Roussillon se concentre aussi dans les agglomérations littorales. Son réseau de villes connaît la plus forte croissance démographique de France mais aujourd'hui, la situation est pour le moins contrastée. En effet, alors que 55 % de la population régionale (contre 40 % en 1990) se concentre aujourd'hui sur les 16 % du territoire régional que constitue la frange littorale, 60 % du littoral reste vierge de toute urbanisation. Par ailleurs des demandes de plus en plus nombreuses émergent, portant sur l'utilisation de ces espaces naturels et de nouvelles activités se développent sur le littoral, cadre idéal pour les loisirs récréatifs ou sportifs, la découverte de la nature et des richesses paysagères, ornithologiques, culturelles..."

L'accroissement de la population permanente et temporaire et la multiplication des modes



FIG. 2.6 – Une zone côtière particulière : le littoral du Languedoc Roussillon

d'utilisation du littoral génèrent de nombreux conflits. L'ignorance du besoin des autres, l'absence de sensibilisation des populations nouvelles ou saisonnières à la vulnérabilité des milieux littoraux, rendent particulièrement difficile la résolution de ces conflits. Face à une dégradation permanente des milieux littoraux et aux nombreux conflits d'usages, le fonctionnement trop sectorialisé des politiques publiques n'a pas permis, à ce jour, de promouvoir une politique à la mesure des enjeux du littoral".

Face à ce constat, les responsables régionaux avec l'appui de l'Etat ont décidé d'engager des initiatives¹⁸ de façon à favoriser la mise en œuvre de plans de gestion intégrée des ressources littorales. Le programme Syscolag propose une approche originale.

2.2.2 Le programme Syscolag

Le programme **SYSCOLAG** (*SYStèmes COTiers et LAGunaires*¹⁹), est un programme de GIZC, co-financé par la *région Languedoc Roussillon*, l'*Ifremer* et autres acteurs scientifiques régionaux (Universités de Montpellier I, II et III, Brgm, Ird et Cemagref) dont la mise en œuvre a débutée en 2002 à travers le financement de sujets de recherches interdisciplinaires. Par son action, ce programme souhaite contribuer à l'intégration des différents points de vue thématiques régionaux et au décloisonnement des institutions impliquées pour le bénéfice des processus de GIZC à venir.

A ces fins, ce programme a privilégié le financement de moyens humains (plutôt que matériel) de manière à favoriser l'échange, la concertation et la mise en place d'outils collaboratifs par les partenaires impliqués.

Objectifs et enjeux généraux du programme Syscolag

Syscolag est un programme de recherche appliquée qui cherche à assister la mise en œuvre opérationnelle de la *GIZC* par la construction d'outils qui solutionnent ou appuient l'aide à la décision.

Les moyens humains sont importants et regroupent actuellement une soixantaine de personnes directement mobilisées dans le programme *Syscolag*. L'interdisciplinarité doit se réaliser autour de projets de recherche transversaux qui impliquent une vingtaine de laboratoires de recherche (réparties en différentes localisations dans la région), plus d'une cinquantaine de scientifiques réunis en particulier autour de onze sujets de doctorat. Les disciplines représentées sont les suivantes : *Sciences politiques, Sciences économiques, Géologie, Ecologie marine et biostatistique, Informatique, Halieutique, Géographie, Sédimentologie*. Syscolag a, par ce biais, chercher à s'appuyer sur un panel d'acteurs et de thématiques représentatif de la variété d'*acteurs* potentiellement mobilisés dans la GIZC (voir partie 2.1.2). En conséquence, les systèmes de gestion de l'information et des connaissances manipulées par ces différents acteurs et les terminologies disciplinaires sont donc variés et les problèmes interdisciplinaires liés à l'interopérabilité entre les modes de gestion des *ressources informationnelles* (à commencer par la mutualisation des descriptions des ressources informationnelles) et les terminologies des différents *acteurs* font face aux mêmes verrous scientifiques que ceux évoqués en partie 2.1 [16, 15].

Les objectifs principaux énoncés par le programme Syscolag visent à :

- à court terme, au cours de la réalisation de ces travaux de recherche :

¹⁸Par exemple la *Mission littoral, le programme AGIL...*

¹⁹<http://www.syscolag.org>.

- mobiliser et *mutualiser prioritairement les ressources et connaissances disponibles* pour optimiser l'intégration des paramètres pertinents pour chaque problématique,
- favoriser une *approche interdisciplinaire et intégrée*.
- à moyen terme, à l'issue des travaux de recherches :
 - proposer des *indicateurs pertinents de suivi des milieux et des usages*,
 - construire des outils scientifiques d'aide à la décision,
 - ceci devrait permettre, à moyen terme, de constituer un *observatoire régional du littoral du Languedoc-Roussillon*, à l'attention du grand public.

Pour atteindre ces objectifs, le programme est composé de 2 volets complémentaires et indissociables :

- **volet opérationnel** : définition d'outils d'aide à la gestion, en particulier pour l'élaboration d'un accès commun à l'information et à la connaissance,
- **volet recherche** : élaboration de nouveaux concepts, travail de recherche propre à chaque thèse.

Le bon fonctionnement du programme repose sur une animation importante et sur les échanges d'expériences qui sont réalisés au cours de réunions régulières.

Selon les termes de *Syscolag*, la réalisation du programme doit aboutir au développement d'une *Base de Connaissance* commune par l'ensemble des partenaires du programme, sous la responsabilité de l'Ifremer et qui sera intégrée, à terme dans l'*Observatoire*. Le sens attribué au terme *Base de Connaissance* est encore flou et la nature de cet outil dépendra en partie des conclusions apportées par ce travail de thèse.

Projet Informatique : gestion de l'information et de la connaissance

Le développement d'une *Base de Connaissances Commune*, répondant aux spécificités et aux contraintes du programme *Syscolag* comprend trois étapes principales (extrait de la Convention Cadre du programme [164]) :

- étude des niveaux d'organisation du système et des fonctionnalités nécessaires,
- spécifications détaillées de cette *Base de Connaissances Commune* et de ses outils,
- élaboration d'un prototype.

Ce prototype de *Base de Connaissances* régionale partagée sera conçu afin d'intégrer les *ressources informationnelles Syscolag* et hors *Syscolag*. Il intéressera l'ensemble du *continuum terre-mer* (bassin versant, lagunes, mer ouverte, à l'échelle du Golfe du Lion) et permettra de rendre accessible et disponible l'ensemble de ces *ressources et connaissances* à l'ensemble des partenaires du programme *Syscolag*.

Démarche générale

Compte tenu des objectifs du programme et des moyens disponibles, la stratégie de ce programme de recherche est divisée en trois phases :

1. **La première phase**, qui correspond à la fédération des moyens humains énumérés et des savoirs, est en cours. Cette mise en commun de moyens a été réalisée afin de mettre en œuvre une approche qui soit réellement pluridisciplinaire pour constituer une aide à l'action, en particulier par l'élaboration d'indicateurs de suivi. Ce travail de mise en réseau et de mutualisation s'appuiera sur la mise en place d'un *Extranet Syscolag* accessible aux membres du programme.

2. **La deuxième phase** du programme qui consiste à *organiser la concertation des différents types d'acteurs* (scientifiques, gestionnaires, professionnels, associations, ...) a pour objectif la *définition d'un langage commun*. Elle sera réalisée notamment, sous la forme de Forum de discussion ou d'ateliers d'échange d'expérience. Ce qui est recherché ici n'est pas uniquement une consultation mais une véritable *participation de tous les acteurs* afin que l'*information soit restituée de façon adaptée* à l'acquisition de nouvelles connaissances et soit en adéquation avec la demande des acteurs locaux.
3. Enfin, l'objectif de **la troisième phase** du programme consiste à finaliser la construction d'une maquette préfigurative de l'*observatoire du littoral*. Cet observatoire consiste en un outil informatique associé à des forums d'experts, qui sera accessible à tous via *Internet*. Il offrira à chacun des acteurs un accès aux connaissances disponibles (données, bibliographie, cartes, synthèses, indicateurs, simulations, diagnostics, ...).

Position de ce travail de thèse

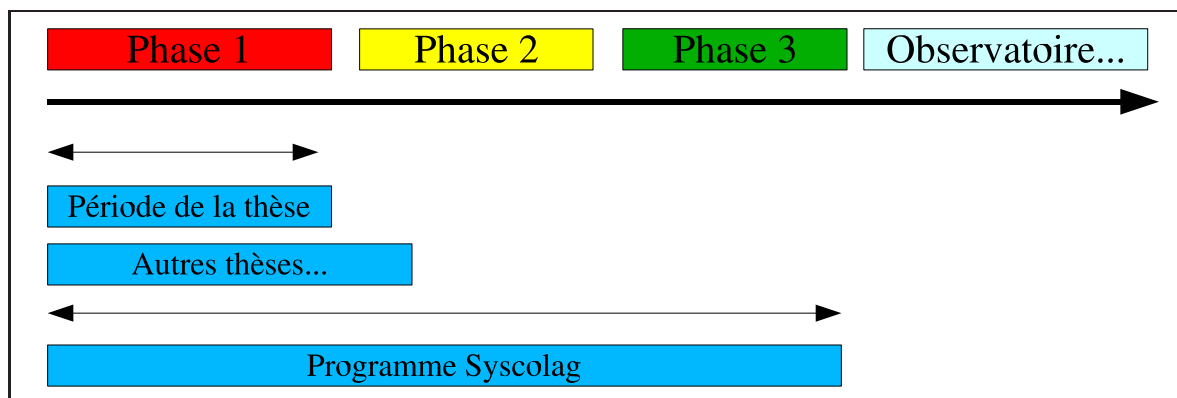


FIG. 2.7 – Position de la thèse dans le programme Syscolag

Le travail présenté dans ce document prend donc place dans la phase *un* du programme Syscolag et vise donc à assiter la définition d'outils ou de méthodes qui permettront la mise en place dans les phases *deux* et *trois* de prototypes de *mutualisation de l'information et de la connaissance* qui seront à terme intégrés dans un *observatoire régional du littoral*.

Synthèse de la section 2.2

- La zone côtière du *Languedoc Roussillon* est un espace privilégié d'application d'un mode de gestion intégrée compte tenu de l'histoire de son aménagement parfois opposé à l'intérêt de ses richesses écologiques.

- Pour soutenir ce nouveau mode de gestion, le programme *Syscolag* propose une approche qui tient compte des recommandations générales de la *GIZC* qu’il adapte aux caractéristiques régionales.
- Une attention particulière est portée à la mise en place d’outils de gestion de l’information et de la connaissance adaptés aux acteurs présents et aux ressources informationnelles présentes.
- Sans imposer la nature des solutions à apporter, le programme *Syscolag* suggère la prise en compte de certaines caractéristiques du domaine d’application.
- En particulier, il s’agit de prendre en compte l’intérêt du Web pour la diffusion de l’information entre les différents acteurs, la nécessaire adaptation de la solution proposée aux habitudes des partenaires : respecter les systèmes de gestion de l’information et de la connaissance en place et les sémantiques disciplinaires variées.
- Parmi les outils pressentis, on trouve en particulier la mise en place d’un *portal Web*, d’un *Observatoire*, d’une *Base de Connaissance* (dans un sens et sous une forme qui restent à définir) auxquels se travail doit apporter des éléments de réflexion.
- La thèse en informatique présentée dans ce document a pour objectif initial de proposer des solutions de mutualisation de l’information et de la connaissance qui aideront à définir la nature de la solution retenue à plus long terme.

Ces objectifs correspondent aux problématiques émergentes analysées par les projets internationaux en charge de guider les programmes de *GIZC* désireux d’améliorer leur gestion de l’information et de la connaissance. Nous passerons en revue la nature des problèmes et réponses inventoriés par ces projets et présenterons quelques systèmes d’informations opérationnels aux objectifs comparables à ceux du programme *Syscolag*

2.3 Systèmes d’information pour la *GIZC* : état de l’art

Il s’agit ici de voir quelles sont les réponses spécifiques, apportées par d’autres programmes de recherche, aux questions soulevées dans les parties 2.1 et 2.1.3.

En effet, compte tenu des principes énoncés pour la mise en œuvre d’une gestion performante de l’information dans un programme de DD et plus particulièrement de *GIZC* et des objectifs du programme *Syscolag* (voir partie 2.2.2), nous proposons de présenter, dans un premier temps, une synthèse des recommandations et directives internationales concernant la gestion de l’information pour la *GIZC*. Dans un deuxième temps, nous rapporterons quelques implémentations de référence qui existent de par le monde et visent à opérationnaliser ces recommandations. Cela permettra, par la suite, d’évaluer la pertinence des solutions proposées pour résoudre les problématiques du programme *Syscolag*.

2.3.1 Directives générales pour la gestion de l’information liée aux zones marines et côtières

De nombreux projets de recherche pluridisciplinaires pour la *GIZC* ont cherché à mettre en pratique les recommandations générales du DD à travers des *systèmes d’information* per-

formants pour la localisation et l'accès aux ressources et connaissances disponibles pour leurs propres cas d'études. Leurs objectifs sont proches des outils de type *observatoire* présentés en partie 2.1.3. Il ressort de ces expériences un ensemble de recommandations opérationnelles pour assister les équipes scientifiques en charge de la gestion de données qui veulent un avis et des méthodes pour gérer correctement leurs données. Les projets qui synthétisent ces propositions n'oublient pas cependant de souligner les lacunes des systèmes actuellement en place.

Oceans 21 est un programme coopératif de travail entre la *Commission Océanographique Intergouvernementale*²⁰ (COI) et l'*International Geographic Union*²¹ (IGU) [95]. Créé en 1998 dans le cadre de l'*année internationale de l'Océan*, ce programme est ciblé sur la GIZC. Son objectif principal est le développement de méthodes, de concepts et de techniques au service de la GIZC. Les recommandations de ce programme ont été adressées au *sommet de Johannesburg* en 2002 et ont obtenu un consensus unanime. Parmi les orientations d'*Oceans 21* pour la période 2005-2008, on note une proposition d'optimisation de l'utilisation de l'information spatiale et des *SIG* pour la gestion du littoral en procurant des outils d'analyse, de modélisation et de visualisation ainsi que des techniques de mise en place d'*infrastructures de données spatiales* pour les zones marines et côtières au niveau régional, national et international. En effet, d'après *Oceans 21*, de par le monde, l'information géospatiale est perçue comme le lien qui aide à intégrer l'information provenant de différentes sources et disciplines au sein de systèmes d'analyse et de gestion unifiés nécessaires pour mettre en œuvre une e-gouvernance moderne et efficace et aidant à l'amélioration de la qualité de vie des citoyens.

La mise en œuvre opérationnelle de ces recommandations se fera à travers des projets liés que sont *MOTIVE*, *ECO-IMAGINE*²² et la série de conférences *COASTGIS*²³. A ce titre, le thème 1 de *Oceans 21* est l'"*accumulation de connaissances sur la zone côtière pour le développement durable*". Les concepts et méthodes proposés sont discutés en relation avec le groupe de travail 1 de *ECO-Imagine* "*Construction de connaissance et d'information géographique côtières*".

MOTIVE²⁴ est un projet financé par l'*Union Européenne* de février 2005 à février 2007 qui évalue les aspects opérationnels de l'harmonisation de données terre-mer et leur intégration trans-disciplinaires et frontalières en Europe. *MOTIVE* exploite le travail effectué dans le cadre de **MarineXML**²⁵[126] concernant le développement d'ontologies et les techniques d'interopérabilité, en relation avec les travaux de l'*Open Gis Consortium*²⁶ (OGC) et a pour objectif de produire des outils directement applicables au public cible d'*Oceans 21*. Au niveau européen, la directive **INSPIRE**²⁷ reconnaît l'importance particulière des besoins en information pour la communauté côtière / marine. De la même manière le **GMES**²⁸ (*Global Monitoring for Environment and Security*) a 6 thèmes principaux dont un nommé : "*Applications marines et océaniques*".

En Amérique du Nord, l'agence américaine **NOAA**²⁹ met en avant le besoin d'améliorer la

²⁰<http://ioc.unesco.org/>.

²¹<http://www.igu-net.org/>.

²²<http://www.gisig.it/eco-imagine/>.

²³<http://www.coastgis.org/>.

²⁴Marine Overlays on Topography for Annex II Valuation and Exploitation.

²⁵<http://www.marinexml.net/>.

²⁶<http://www.opengeospatial.org/>.

²⁷<http://inspire.jrc.it/>.

²⁸<http://www.gmes.info/>.

²⁹National Oceanic and Atmospheric Administration, <http://www.noaa.gov/>.

coordination et l'intégration des sources de données terre-mer et les Etats Unis et le Canada sont, au niveau international, les pays les plus en avance dans ce domaine actuellement.

Dans la lignée de ces initiatives, le projet américain *Marine Metadata Interoperability* [82] *project* propose, regroupés au niveau d'un site Internet³⁰, un ensemble de méthodologies et de documentations visant à rendre les données marines et côtières plus facilement localisables, diffusables, réutilisables, et compatibles avec d'autres jeux de données du domaine. Le nom de ce projet porte déjà une grande partie de la nature des réponses proposées. La solution retenue dans ce but consiste, en effet, à guider les utilisateurs vers une gestion plus pertinente de leurs données à l'aide de *métadonnées*.

En simplifiant la complexité d'implémentation des métadonnées en recommandations simples, le *MMI* souhaite permettre aux scientifiques et aux gestionnaires de données de tous niveaux d'appliquer de bonnes pratiques de métadonnées dès le départ de leur projet de gestion de l'information.

Pour cela, le *MMI* cerne un ensemble de lacunes dans la gestion des métadonnées généralement disponibles dans les domaines marin et côtier (voir figure 2.8). Ces analyses rejoignent celles présentées en partie 2.1.3. En effet, les difficultés énumérées correspondent bien aux verrous évoqués précédemment :

- *métadonnées incomplètes* : à cause d'un manque de standards de métadonnées pour ces domaines spécifiques,
- *ambiguïté sémantique* : à cause des différentes perceptions du système,
- *métadonnées distribuées et disconnectées* : à cause du nombre de collections de données réparties dans différents organismes,
- *incompatibilité syntaxique* : à cause de la variété de logiciels et de formats.

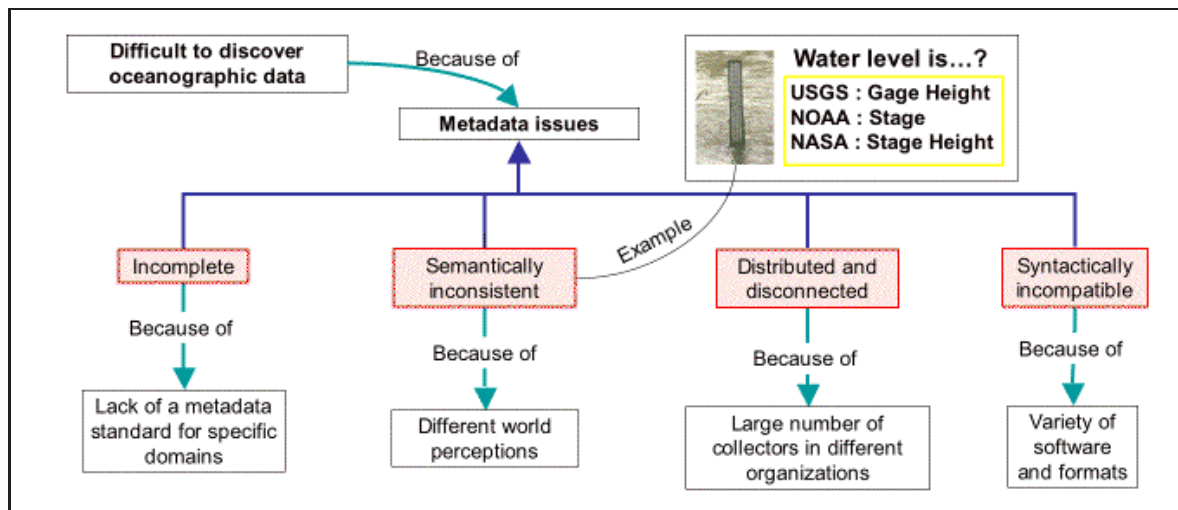


FIG. 2.8 – Les verrous analysés par le MMI [82]

En réponse à ces lacunes, le *MMI* propose donc des recommandations sur les *vocabulaires contrôlés* (*glossaires, thésaurus* ou *ontologies*) et sur des outils existants pour la mise en œuvre de service de métadonnées (*SMD*) et d'ontologies (standards et logiciels).

On peut donc synthétiser le contenu de ces propositions en disant que l'ensemble des

³⁰<http://marinemetadata.org/>.

projets et directives présentés encouragent explicitement l'utilisation des *métadonnées* et la création de *référentiels thématiques et spatiaux partagés* (compte tenu de l'importance des données spatiales), en particulier la mise en place d'*ontologies*. Et ceci doit préférentiellement être réalisé dans le respect des standards existants.

Le Web étant reconnu comme le média de prédilection pour la diffusion d'information pour la GIZC [101], l'ensemble des projets existants sont généralement accessibles en ligne (aux contraintes d'identification prêt) sous forme de *portails Web*. Il est donc possible de faire un état de l'art représentatif de l'existant sur le Web.

Nous allons présenter une partie de ces initiatives. Nous nous intéresserons uniquement à des systèmes évolués permettant la saisie, la description, le traitement dynamique et surtout la recherche d'information distribuée avec le souci de résoudre les verrous que nous venons de rappeler.

2.3.2 Systèmes de références pour la gestion de l'information en GIZC

Dans cette partie, nous déclinons différents systèmes produits par des projets internationaux, de type *observatoire* ou gestion de ressources environnementales, en présentant le contexte, les objectifs et les solutions retenues pour mettre en œuvre les recommandations énumérées dans la partie 2.3.1 (éléments principaux de l'architecture et choix techniques effectués, en particulier pour la diffusion d'information et les référentiels terminologiques et spatiaux) et améliorer les lacunes actuelles dans le domaine de la GIZC.

FIGIS *Contexte et objectifs* : Dans le cadre de son rôle de coordination des informations environnementales à l'échelle internationale, les *Nations Unies* et la *FAO* en particulier, entreprennent actuellement d'importants efforts pour la mise en place de divers systèmes d'information à l'échelle mondiale, en particulier en ce qui concerne les *pêcheries*, l'*aquaculture*... Il s'agit d'intégrer, entre autres, en un même système d'information, des données statistiques distribuées dans différents pays sous des formats (syntaxe) et modes de descriptions (sémantique) hétérogènes afin de produire dynamiquement des statistiques de pêche à l'échelle mondiale.

Solutions proposées : Pour offrir une *interopérabilité* optimale la *FAO* fournit un effort important sur l'*utilisation de métadonnées*, en particulier avec *Geonetwork* (voir partie 4.2.3), et de *vocabulaires contrôlés*. Ceci devant faciliter la diffusion, par le biais de sites Web générés dynamiquement (utilisant les langages *XML* et *XSL*), des données statistiques relatives à l'exploitation des pêcheries.

Pour atteindre ces objectifs, la mise en place de *thésaurus*, voire d'*ontologies*, dédiés à l'aquaculture ou aux pêcheries est capitale afin, en particulier, de gérer les problèmes de synonymies et de multi-linguisme. Ces démarches ont commencé avec différents thésaurus : *ASFA*, *FIGIS*, *OneFish* ou encore *Agrovoc*, et ces efforts tendent aujourd'hui à être unifiés au sein d'un même service par la mise en place d'une *ontologie* qui chapeaute ces différents outils de contrôle du vocabulaire et permet des correspondances entre eux afin de les rendre interopérables : l'ontologie *FOS* [75, 74]. Ces solutions techniques utilisent le langage *UML* (diagrammes d'activités) pour le partage des représentations et le langage *OWL* comme standard pour le *Web sémantique*, en utilisant notamment le logiciel *Protégé 2000*. Par ce biais, ces systèmes souhaitent, entre autres, permettre l'*expansion de requêtes*, les *requêtes en langage naturel*, la *consultation de métadonnées*.

Afin d'améliorer la convivialité du service et la représentation de l'information, on note également l'utilisation d'*interfaces cartographiques*.

La figure 2.9 illustre un exemple d'interface graphique utilisateur qui restitue l'information contenue dans les bases de données par espèces ou activités en utilisant les supports textuels, cartographiques voire des traitements statistiques, et diffusée par le biais du Web [65].

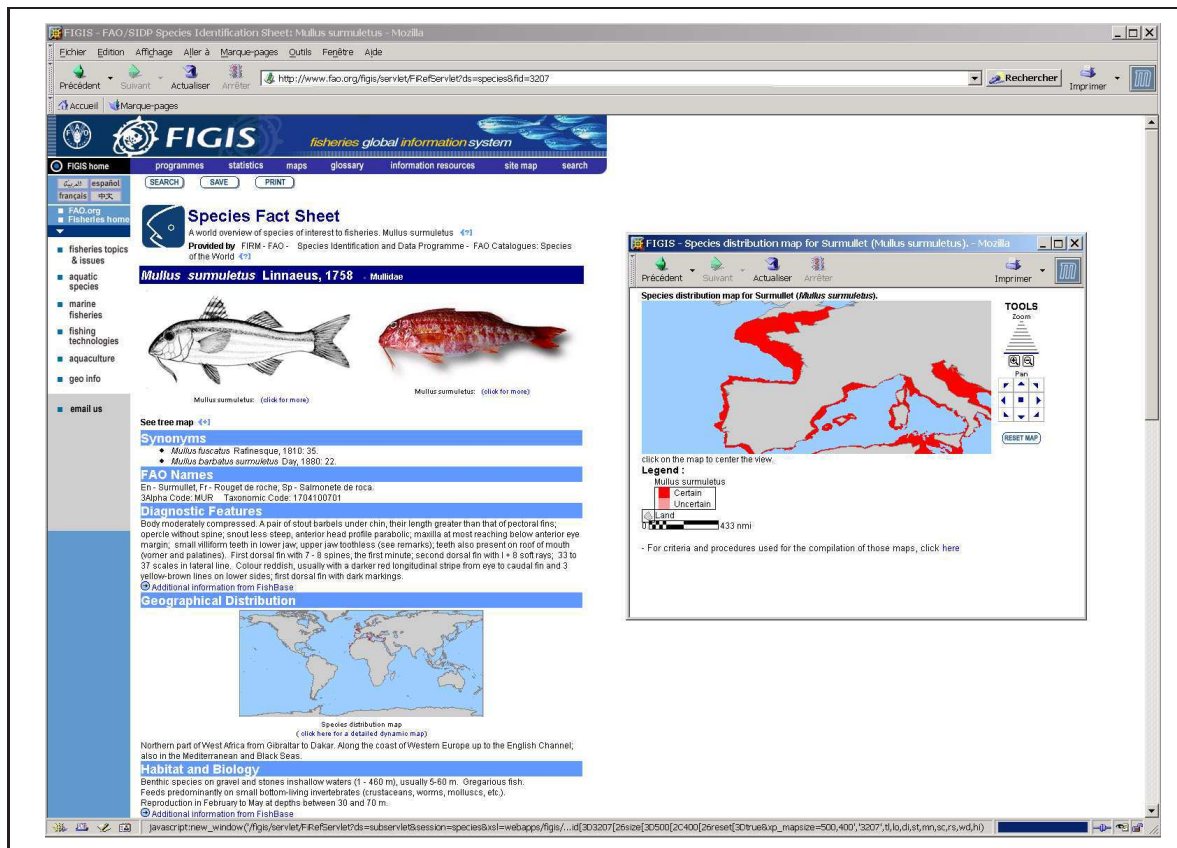


FIG. 2.9 – Le SI FIGIS

THETIS, ARION et POSEIDON *Contexte et objectifs* : Sur le constat du développement d'une majorité d'applications scientifiques isolées les unes des autres et sur celui du développement d'Internet et des *systèmes distribués*, les projets *THETIS*, *POSEIDON* et *ARION* tendent à *procurer un point d'accès unique par le Web à une collection de serveurs en réseau contenant des données, programmes et outils pour des applications de gestion de la zone côtière*.

POSEIDON est un programme de recherche de *modélisation océanique dans un environnement informatique distribué* mené par le MIT (*Massachusetts Institute of Technology*) dans le cadre d'un groupe de recherche pluridisciplinaire associant océanographie physique et biologique avec l'acoustique océanique.

L'objectif de ce projet est l'amélioration de la prévision en temps réel et les applications concernent la gestion des pêcheries, le contrôle des pollutions et diverses opérations maritimes.

Les partenaires impliqués proviennent du laboratoire d'acoustique du *MIT* et du laboratoire d'océanographie physique et biologique d'Harvard DEAS.

Il s'agit d'améliorer l'accès aux données et informations concernant l'évolution océanique et de sa dynamique qui impliquent des interactions multi-échelles [144]. En conséquence, l'efficacité et la rapidité de prévisions pluridisciplinaires ne pourront être atteintes que si les systèmes de prédiction et d'observation se comportent et s'adaptent de la même manière. Il convient alors de mettre en place les bases d'un système distribué pour un travail de prévision interdisciplinaire en temps réel basé sur des modèles et des échantillons adaptés. Le but scientifique est le couplage de l'océanographie physique et biologique avec l'acoustique sous-marine. Le but technique est la construction d'un système dynamique basé sur des infrastructures évoluées, distribuées avec des recherches et visualisations d'information performantes [165].

De la même manière *ARION* [26] est un système de librairie scientifique numérique qui vise à améliorer l'accès aux ressources scientifiques telles que les données ou les programmes. Ce projet a débuté en 2001 pour une durée de trois ans. Il s'agit d'un consortium ouvert qui fédère des organismes de recherches, des producteurs de données. Le leader du projet est *the Institute of Computer Science-Foundation for Research and Technology* (Grèce), la *National Technical University of Athens* (Grèce), le Consiglio Nazionale delle Ricerche CNR-IMA (Italie), la Commission de la communauté européenne, le Joint Research Centre (Italie), l'Université de Crète (Grèce), le SMEs HR Wallingford Ltd (Royaume-Uni), la Oceanographic Company of Norway ASA (Norvège) et la Enterprise LSE Limited (Royaume-Uni).

Par le biais d'Internet *ARION* souhaite permettre [90] :

- l'utilisation de ressources scientifiques distantes stockées à différents endroits ;
- l'élaboration de recherches perfectionnées de ressources scientifiques ;
- une publication facilitée des ressources scientifiques.

THETIS s'inscrit dans la continuité des deux programmes précédents. Il vise à la mise en place d'un système fédéré pour l'intégration de *données ou bases de données patrimoniales*³¹, de *SIG*, de *programmes de simulation* et d'*outils de visualisation* via un portail *Web*.

Solutions proposées :

Dans ces trois projets, l'utilisation des *métadonnées* et *ontologies* est mise en avant pour optimiser l'*interopérabilité syntaxique et sémantique* des systèmes (voir Figure 2.10), le tout à la disposition des utilisateurs par le biais de *portails Web*.

Plus spécifiquement, les solutions proposées par *ARION* se sont appuyés sur les standards existants (*FGDC 94* pour les métadonnées et *OGC* pour l'information géographique) ainsi que sur la réutilisation d'ontologies existantes. D'un point de vue langages et logiciels, *RDF*³² et *RQL* sont utilisés, respectivement, pour la description et l'interrogation de métadonnées et d'ontologies (pour l'interrogation de *RDF*) associé à *Protégé 2000* pour sa simplicité d'utilisation, ainsi qu'au *SGBDR PostgreSQL* [162]. De plus, *ARION* offre également des possibilités avancées de visualisation de l'information géographique.

Concernant *POSEIDON*, le système est conçu comme un réseau de données, de ressources et de connaissances distribuées communiquant entre elles par le protocole *CORBA*³³ [141]. Les interfaces graphiques utilisateurs sont interprétées par un navigateur standard qui télécharge

³¹Données qui sont issues de langages, plates-formes et techniques dépassés ou encore qui ont été héritées d'une autre entreprise.

³²Resource Description Framework (*RDF*) est un modèle de graphes pour décrire les (méta-)données et permettre un certain traitement automatique des métadonnées. Une des syntaxes (sérialisation) de ce langage est *RDF/XML* [90]. Il s'agit d'un dialecte *XML* développée par le consortium W3C.

³³Common Object Request Broker Architecture.

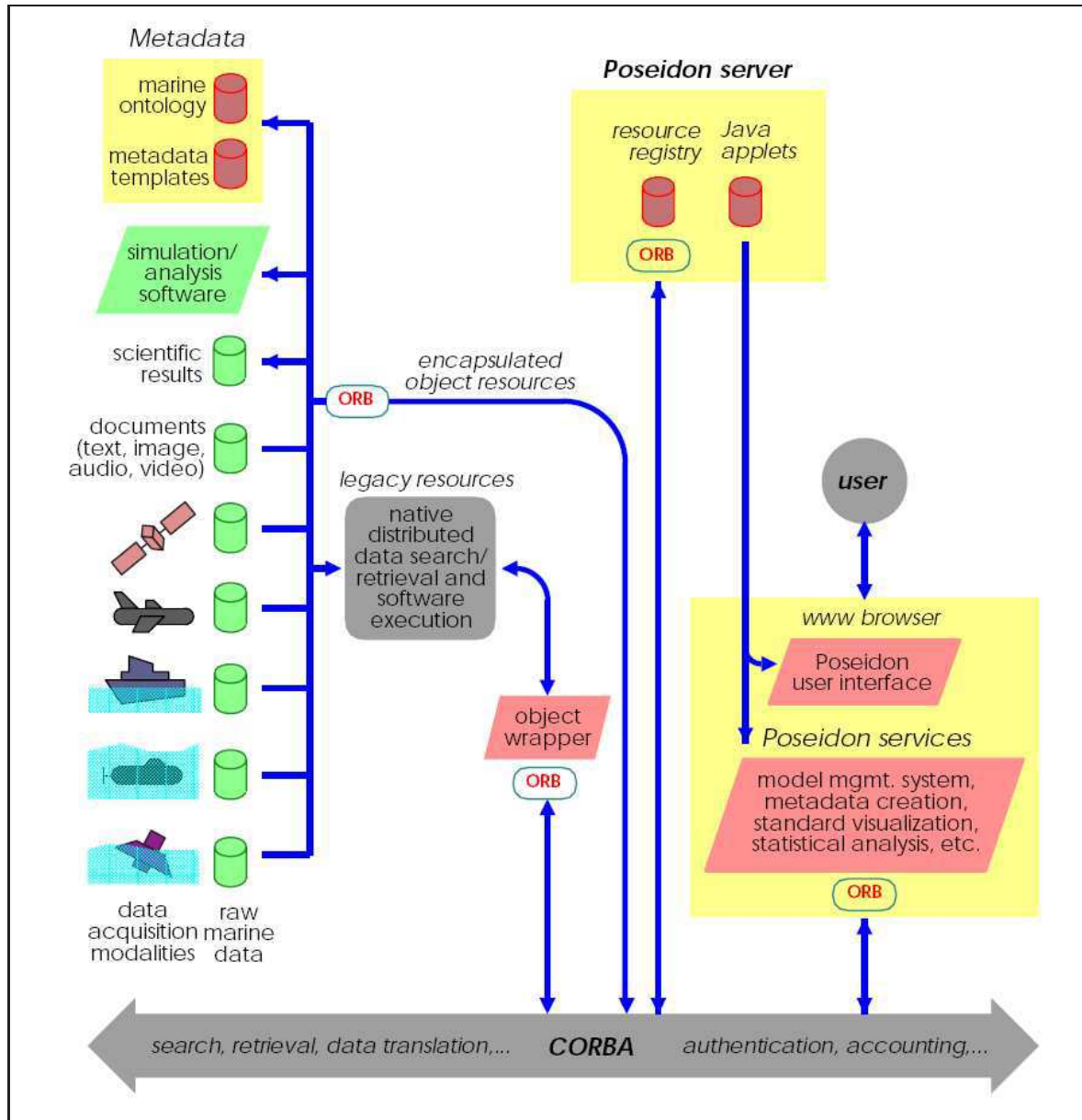


FIG. 2.10 – L'architecture du SI POSEIDON [141]

des applets *Java* sur un serveur *POSEIDON* et permettent en particulier la visualisation d'information spatiale par l'utilisation d'une *interface cartographique* [177].

Afin d'améliorer les recherches d'information *POSEIDON* a développé un *service de métadonnées* (sur la base des normes *Dublin Core* et *FGDC*) associé au développement d'*ontologies partagées*.

De la même manière, le projet *THETIS* apporte des solutions pour la *découverte*, *l'intégration* et la *visualisation d'information*, dans un réseau de SGBD distribués, par la mise en place d'un *service de métadonnées* basé sur des fichiers d'échange *XML*. La restitution étant faite à l'utilisateur par un *portail Web* constitué d'interfaces graphiques exploitant la dimen-

sion spatiale des ressources par la visualisation d'information géographique ou la formulation de requêtes spatiales à l'aide d'une *interface cartographique*. De plus, *THETIS* s'intéresse à l'utilisation d'*ontologies* pour la description de concepts issus de domaines spécifiques ainsi qu'à des méthodes de création ou de gestion d'ontologies. A ces fins, ce programme reprend les spécifications de *POSEIDON* et utilise l'environnement de développement d'ontologies *Ontolingua* [12].

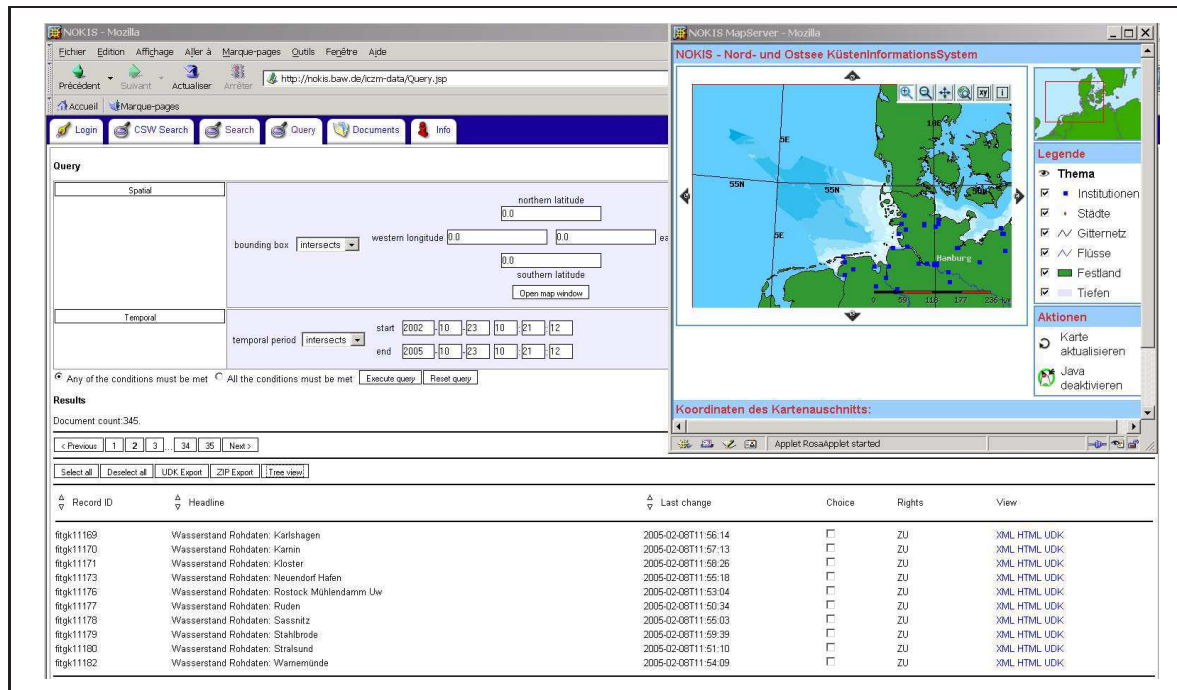


FIG. 2.11 – Le SI NOKIS

NOKIS Contexte et objectifs : Le projet *NOKIS* a pour objectifs principaux de faciliter et d'intensifier la communication ainsi que le partage d'information et de données entre institutions impliquées dans la recherche sur la zone côtière pour les régions allemandes de la mer Baltique et de la mer du Nord [115]. Un portail Web a été mis en place pour se lier à la communauté de recherche internationale. En outre, ce projet revendique un environnement de développement libre pour résoudre les problèmes de paiement de licences qui pourraient nuire à l'utilisation de l'outil et favoriser l'adhésion de nouveaux partenaires [113]. Le moteur de recherche a pour objectif de proposer des recherches en *texte libre* ou par *mots clés* voire sur *emprise spatiale*.

Solutions proposées : La solution proposée repose sur des fichiers de description XML et des interfaces graphiques utilisateurs pour la description et la recherche de métadonnées. La norme de métadonnées choisie est l'ISO 19115 dont le modèle de métadonnées est fidèlement interprété par l'outil Java. L'infrastructure est constituée de bases de métadonnées réparties qui échangent de manière automatisée avec un serveur centralisé (les fichiers d'échanges sont en XML, conforme à l'ISO 19119) [104]. Le profil élaboré tient néanmoins compte du FGDC Shoreline Profile. L'éditeur de métadonnées repose sur une technologie Open Source : XML pour le contenu, mise en forme par l'utilisation d'XSL. Apache est utilisé comme serveur

Web avec *Tomcat* et *MySQL* comme *SGBD* pour l'archivage des métadonnées et données dans un premier temps, puis *Postgres* dans un deuxième temps [111].

Deux outils complémentaires sont utilisés pour la consultation de métadonnées : *Dig* comme *moteur de recherche plein texte* et une *interface cartographique* utilisant *Mapserver* (conforme aux recommandations de l'*OGC*) pour permettre les requêtes spatiales [86], voir illustration en Figure 2.11. En outre ce projet a pour projet l'élaboration d'un *thésaurus* pour la zone côtière, afin d'augmenter la qualité des recherches proposées [112]. Le *glossaire international d'hydrologie de l'Unesco* et *GEMET* sont les *vocabulaires contrôlés* actuellement utilisés mais les auteurs précisent leurs limites pour le domaine côtier et la nécessité d'un outil dédié au domaine pour l'instant inexistant [135, 114].

Ce système permet, par exemple, par des recherches aiguillées par les métadonnées, la visualisation en ligne de données bathymétriques.

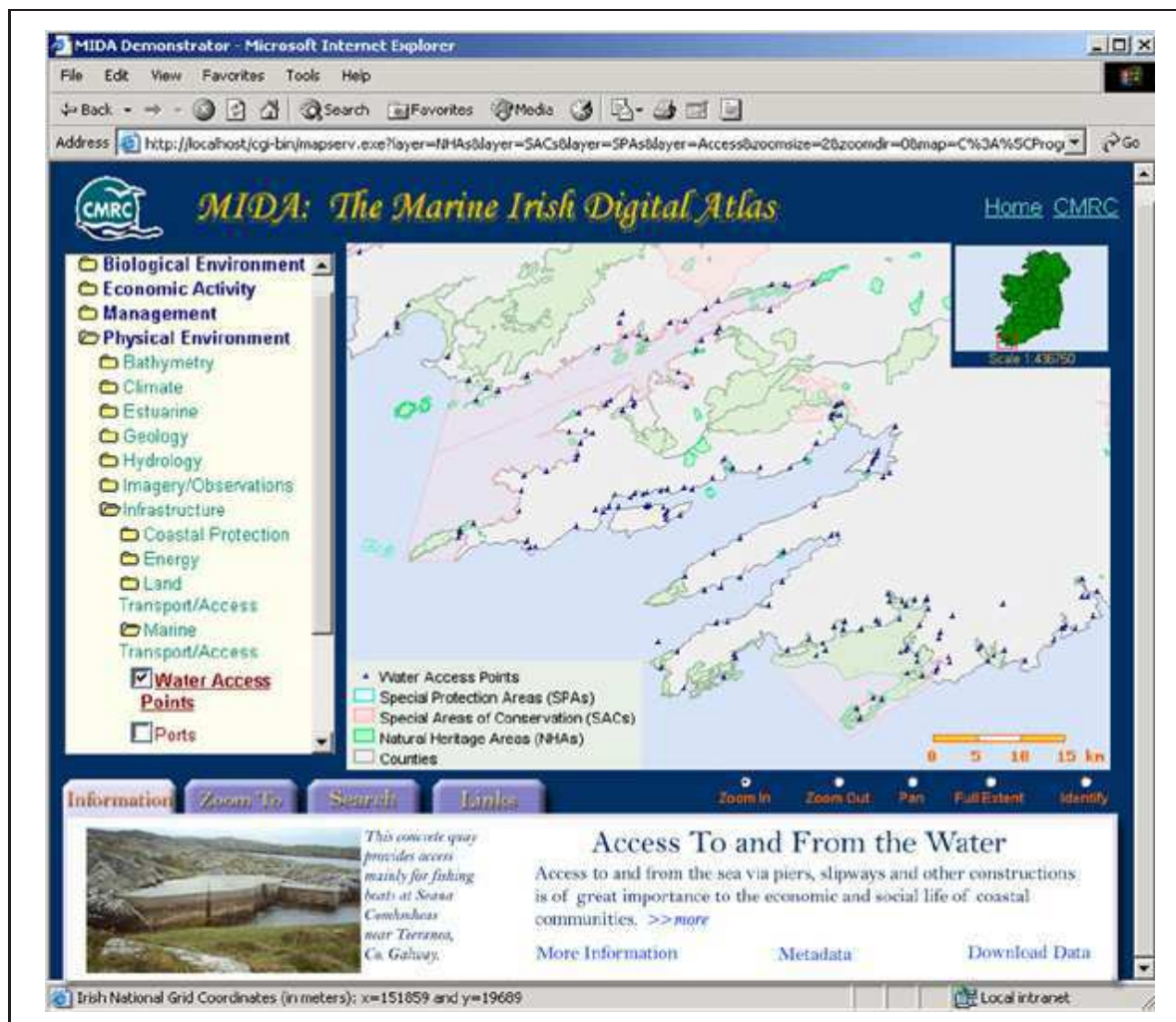


FIG. 2.12 – Le SI du MIDA

MIDA *Contexte et objectifs* : Le *MIDA* est un *portail Web* sur l'information marine et côtière en Irlande. Face à l'absence de source unique pour la recherche et la consultation / visualisation d'information sur ces domaines, le *Marine Irish Digital Atlas*³⁴ propose de changer l'accès aux données marines et côtières en fournissant, en une source unique, les informations nécessaires à l'évaluation quantitative et qualitative des données disponibles [85]. L'objectif principal du projet est de développer un portail Web qui sera une base de référence pour les données spatiales en même temps qu'un outil d'information destiné à un large public (scientifiques, administration, éducation et grand public en général).

Solutions proposées : Le *MIDA* utilise les technologies *Open Source* du Web pour mettre à disposition en ligne les informations sur les données du domaine et en particulier permettre la visualisation d'information géographique. La norme de métadonnées retenues est l'*ISO 19115* implémentée en *XML*. La description et la recherche spatiale de métadonnées ainsi que la visualisation d'information géographique sont permises par l'utilisation de *MapServer* [53]. Le cœur du système repose sur l'interface cartographique *Open Source MapServer* qui permet à l'utilisateur de sélectionner, visualiser et interroger de multiple layers d'information spatiale structurés par une classification thématique (voir Figure 2.12).

Chaque jeu de données présent dans l'atlas est documenté par des métadonnées basées sur l'implémentation des recommandations de l'*ISO 19115*. Les métadonnées de découverte sont gérées au sein d'une base de données *XML* [137].

De plus, le *MIDA* souhaite désormais utiliser des vocabulaires contrôlés (*ontologies*) pour la qualité du service de métadonnées proposé[138].

COASTBASE *Contexte et objectifs* : Le projet *COASTBASE*³⁵ (cinquième PCRD) a été initié en janvier 2000 par un consortium pluridisciplinaire de onze organisations de différents pays et thématiques avec les objectifs suivants [102] :

- améliorer la gestion de la zone côtière à travers l'Europe en améliorant l'accès à l'information pertinente,
- mettre en place une coopération européenne sur l'échange de données et d'information,
- promouvoir la standardisation des données,
- stimuler l'utilisation d'indicateurs sur l'environnement côtier et marin ainsi que pour la GIZC au niveau européen.

COASTBASE est un *portail Web* permettant la recherche et l'accès à de l'information distribuée sur la zone côtière ou marine en Europe [174].

La validation du prototype présenté est effectuée sur la base de trois scénarios (par exemple, l'"eutrophisation en Mer du Nord").

Solutions proposées : Il s'agit de procurer des outils de manipulation des données s'appuyant, en particulier, sur un service de métadonnées et d'une interface cartographique (raster et vecteur)[132].

Les métadonnées sur les sources clés d'information pour la GIZC sont stockées dans le serveur central de *COASTBASE*. Le standard de métadonnées retenu est le *GELOS* conforme au *Dublin Core*.

L'accès est ouvert à tous pour les fonctions principales et aux abonnés pour certaines fonctions comme la saisie de métadonnées. La recherche d'information est articulée autour de paramètres clés : localisation (*où ?*), datation (*quand ?*), contenu thématique (*quoi ?*), per-

³⁴<http://mida.ucc.ie/>.

³⁵<http://herakles.fzi.de:8181/CatalogServer/index.jsp>.

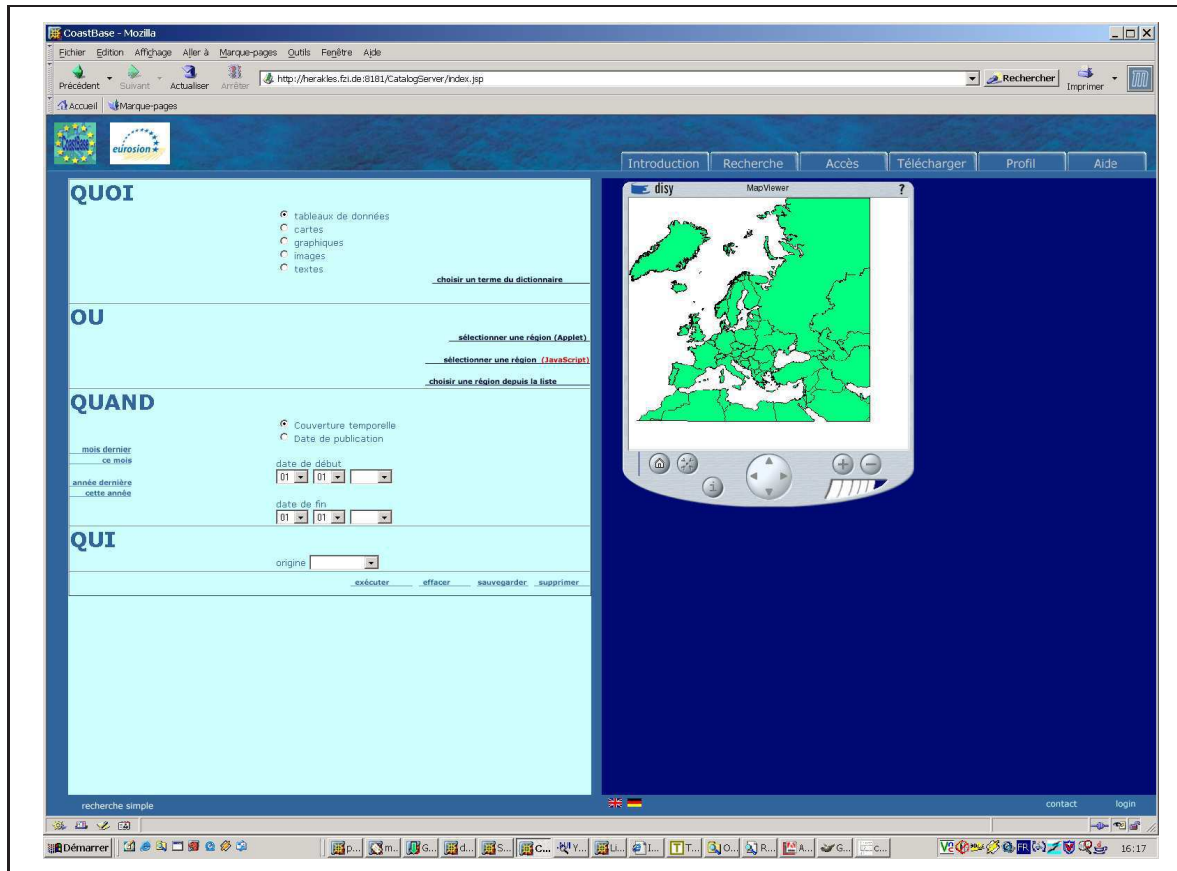


FIG. 2.13 – Le SI COASTBASE

sonnes impliquées (*qui ?*) (voir Figure 2.13).

Du point de vue du contrôle de la terminologie *COASTBASE* propose l'utilisation d'un *référentiel sémantique* et propose le *thésaurus* de l'*Union Européenne GEMET*³⁶ [103], bien que non spécifique à la GIZC. Une *interface cartographique*, qui utilise la librairie *Luciad Map*, permet la recherche spatiale de métadonnées et permet la visualisation d'information géographique (en mode raster et vecteur). Les langages identifiés pour le développement sont *XML* et *XSL-T*. Dans cette perspective, une collaboration est envisagée avec le projet *NO-KIS* pour la mise en place d'un catalogue *XML* sur le serveur *COASTBASE*. Le protocole de communication utilisé par cette architecture est *SOAP*³⁷.

³⁶GEneral Multilingual Environmental Thesaurus, <http://www.eionet.eu.int/gemet>.

³⁷Simple Object Access Protocol.

Synthèse de la section 2.3

- Les directives et initiatives internationales de référence en charge d'orienter la *gestion de l'information et de la connaissance* dans les domaines marins ou côtiers se multiplient devant la nécessité et l'ampleur de cette tâche. Elles mettent en avant des solutions à cette problématique qui s'articulent autour de l'utilisation des *métadonnées*, des *ontologies* et de l'*information géographique*.
- Le Web apporte des possibilités de diffusion précieuses dans ce contexte. L'information géographique occupe une place centrale et doit être prise en compte aussi bien pour assister la gestion de l'information que celle des connaissances.
- Les *métadonnées* apportent une réponse concrète pour la *mutualisation d'information*. Cependant, si elles sont utiles pour documenter la donnée, elles sont souvent *incomplètes, distribuées et disconnectées, incompatibles syntaxiquement et sémantiquement*.
- Le problème sémantique est incontournable dans ces thématiques pluridisciplinaires.
- En appui à ces orientations théoriques, des programmes de recherche ont déjà tenté d'apporter des solutions à ces verrous scientifiques.
- Globalement, les réponses apportées s'appuient sur des portails Web qui hébergent des *services de métadonnées, des outils de contrôle et de partage de la terminologie* employée ainsi que sur des *interfaces cartographiques*. On constate effectivement une hétérogénéité des syntaxes de métadonnées employées et l'absence de vocabulaire de référence dans le domaine.
- En outre les domaines d'application ne correspondent pas directement à celui de la GIZC et l'échelle d'application est internationale ou nationale et plus rarement régionale.

Conclusion du Chapitre 2

La problématique du travail présenté dans ce document trouve son origine dans les principes généraux du *Développement Durable* dont l'objectif est de permettre un équilibre harmonieux entre *écosystèmes* et *développement humain*. En effet, celui-ci recommande l'amélioration des modes de *gestion de l'information et de la connaissance*. Cette gestion doit désormais permettre une *diffusion de l'information vers un public large* ainsi que le *décloisonnement des acteurs*, par une *mutualisation de l'information et de la connaissance* à l'aide de *formalismes de descriptions partagés*.

La *Gestion intégrée de la Zone Côtière (GIZC)* est un *mode de gestion territorialisé* qu'on peut percevoir comme une application du *Développement Durable* à l'*espace côtier*. Un mode de gestion spécifique est dédié au littoral car ce dernier concentre les problématiques majeures de mise en œuvre du *Développement Durable*. En effet, la zone côtière, en raison notamment de ses richesses écologiques, regroupe aujourd'hui près de deux tiers de la population mondiale. Pour apporter des réponses à la *problématique de mutualisation de données et de connaissances pour la GIZC*, nous proposons donc de nous appuyer sur les recommandations héritées du *Développement Durable*. Cependant, nous devons tenir compte, dans le domaine de la GIZC, de la spécificité suivante : la problématique de *gestion de l'information et de la connaissance* est complexifiée par la vision actuelle de l'*espace côtier* selon un *continuum terre-mer*. En effet,

cette nouvelle représentation de l'espace implique la collaboration d'*acteurs* variés qui ont évolués jusqu'à présent dans des milieux professionnels et socio-culturels clairement séparés, en particulier par les modes de gestion traditionnels qui distinguaient milieux terrestres et marins. Cette caractéristique a engendré une variété de *ressources informationnelles* et de *terminologies disciplinaires* considérable et constitue un verrou scientifique que la résolution de notre problématique doit débloquent.

Le *programme de recherche pluridisciplinaire Syscolag* s'appuie sur un échantillon d'*acteurs* et une zone côtière représentatifs des problématiques actuelles de la GIZC. D'un point de vue informatique, ce sujet de thèse doit participer à l'orientation des solutions qui seront retenues, à court terme, pour *améliorer la mutualisation de l'information et de la connaissance* à l'échelle de la zone côtière du *Languedoc Roussillon*.

Au delà de notre contexte d'application, dans des domaines associés à la GIZC, des programmes de recherches étudient la résolution de cette problématique. Ces derniers s'accordent sur la nécessaire résolution des problèmes d'*interopérabilité syntaxique et sémantique* pour l'amélioration de la gestion de l'information et de la connaissance associée aux zones marines et côtières. Les solutions techniques proposées s'appuient sur l'utilisation de **standards** pour les *métadonnées* et les *référentiels sémantiques* ainsi que sur l'utilisation d'*interfaces cartographiques* pour optimiser l'utilisation de données spatialisées (là encore dans le respect des standards en vigueur, généralement **Open Source**, avec en particulier l'*OGC*).

Ces outils sont en construction, leur émergence est récente et liée aux problématiques plus générales du *Web sémantique et cartographique*. L'absence remarquée de référentiels terminologiques spécifiques à notre domaine [115, 177] font de ces projets des contributions de références pour la future résolution des problèmes d'interopérabilité sémantique du domaine. Ils aideront à la mise en place d'*ontologies* plus générales à l'avenir.

A l'inverse, la problématique de référence syntaxique pour l'utilisation de métadonnées est, quant à elle, mieux cernée. Les standards de références sont clairement ciblés par les directives internationales d'aide à la gestion de l'information dans ce domaine.

L'accès au savoir dans la GIZC se fait quasi exclusivement par le biais de **portails Web** dont l'intérêt pour le domaine est unanimement reconnu (idéal pour élargir l'accès au grand public), et l'information est *présentée de manière appropriée* en utilisant des référentiels syntaxiques (standards de métadonnées pour l'interopérabilité de *SGBD* distribués et hétérogènes) et sémantiques (en particuliers thématiques et spatiaux).

Synthèse du Chapitre 2

- La GIZC est une *application territorialisée* des principes du *Développement Durable*. La problématique de *gestion de l'information et de la connaissance*, étudiée dans ce document, est essentielle pour *amorcer* et *assister* sa mise en œuvre.
- Le souci de diffuser vers tous les publics les informations et les connaissances existantes sous une *forme appropriée* (ie compréhensible par tous, tant sur le fond que sur la forme) permet la participation de tous et assiste le *décloisonnement* des *acteurs* ainsi que la *prise de décision*.
- L'*hétérogénéité* et la dispersion des *acteurs* et des *ressources informationnelles* ainsi que les difficultés d'échanges terminologiques inhérentes aux travaux interdisciplinaires sont exacerbés dans le contexte de la GIZC et complexifient cette tâche. A ceci s'ajoute l'absence de *référentiels sémantiques* (référentiels thématique et spatial).
- En outre, les données sont peu documentées alors que la *description des ressources informationnelles* est un *préalable indispensable* à leur *partage physique*, afin de pouvoir estimer leur pertinence et leur localisation avant leur utilisation pour la GIZC.
- Pour pallier ces contraintes et apporter une réponse adaptée au mode de gestion territorialisée qu'est la GIZC, le *support cartographique* procure un mode d'échange privilégié.
- Ce travail s'insère dans le programme de recherche pluridisciplinaire *Syscolag* qui étudie une approche de GIZC adaptée aux caractéristiques du *Languedoc Roussillon*.
- Parmi les outils pressentis par *Syscolag*, on trouve en particulier la mise en place d'un *portail Web*, d'un *Observatoire*, d'une *Base de Connaissance* (dans un sens et sous une forme qui restent à définir) auxquels se travail doit apporter des éléments de réflexion.
- Les directives et initiatives internationales de référence dans les domaines marins ou côtiers se multiplient devant la nécessité et l'ampleur de cette problématique. Elles mettent en avant des solutions qui s'articulent autour de l'utilisation sur le Web de *métadonnées*, d'*ontologies* et d'*information géographique*.
- Les *métadonnées* apportent une réponse à la *mutualisation d'information*. En effet, elles permettent la *description des ressources informationnelles existantes*. Cependant, elles sont souvent incomplètes, distribuées et disconnectées, incompatibles syntaxiquement et sémantiquement.
- Le problème sémantique entre les *acteurs* ou les *machines* est incontournable dans ces thématiques pluridisciplinaires.
- En appui à ces orientations théoriques, des programmes de recherche ont déjà tenté d'apporter des solutions à ces verrous scientifiques et confirment la pertinence de ces solutions. On constate effectivement une hétérogénéité des syntaxes de métadonnées employées et l'absence de vocabulaire de référence dans le domaine.

Chapitre 3

Objectifs et propositions

Le programme de recherche *SYSCOLAG* souhaite opérationnaliser la démarche de *Gestion Intégrée de la Zone Côtière en Languedoc Roussillon* (voir partie 2.2.2). Pour ce qui concerne la mise en place d'outils informatiques, la proposition reprend les objectifs généraux d'amélioration du *partage de l'information et de la connaissance associés à la GIZC*, en considérant la nécessaire adaptation de l'outil proposé aux contraintes locales.

En nous appuyant sur les caractéristiques et l'état de l'art du domaine présentés dans le chapitre 2, nous détaillons, dans ce chapitre, la nature des différentes solutions retenues pour le système que nous préconisons.

Notre proposition repose sur un certain nombre de *concepts* généraux liés au domaine de la *mutualisation d'information et de connaissances* : en particulier nous proposons l'utilisation de *métadonnées* et d'une *ontologie* pour répondre aux problèmes d'*interopérabilités syntaxique et sémantique*. Cependant, la solution que nous proposons est adaptée spécifiquement au domaine de la GIZC et au cahier des charges du programme *Syscolag*. La réponse proposée se présentera sous la forme d'un *portail Web* et tiendra compte des *besoins des acteurs* ainsi que des caractéristiques de leurs actuels *systèmes de gestion de données*. A ces fins, nous proposons la mise en œuvre d'un *service de métadonnées* (pour gérer les descriptions des *ressources informationnelles* de nos partenaires) et celle complémentaire de *référentiels thématique et spatial* (pour assurer la qualité de la *localisation des ressources existantes* en même temps que le partage de *connaissances* entre les *acteurs*), adaptés à ce domaine. Le *support cartographique* aura une place centrale dans la solution proposée et devra assister la *médiation* entre les *acteurs* et les *machines*.

En outre, le souci de qualité et de pérennité du système mis en place, nous a conduit à proposer une *architecture* générique qui implémente les *normes et standards* en vigueur pour permettre l'*interopérabilité* de la solution proposée au delà de notre contexte d'implémentation. Autant que possible, nous souhaitons proposer des applications dans un *environnement de développement libre* afin d'exclure les problèmes de licences et permettre l'accessibilité d'un tel outil au large public concerné par la GIZC. Nous proposons d'utiliser le langage *UML* pour traduire les attentes des *acteurs* et guider le développement de l'outil. Nous présentons finalement les cas généraux d'utilisation du système proposé et les différents utilisateurs identifiés.

3.1 Objectifs principaux de la solution proposée

En réponse aux attentes générales propres à notre domaine d'application formulées dans les principes du DD et de la GIZC et en tenant compte de celles exprimées spécifiquement dans la convention cadre du programme *Syscolag* [164], nous énumérons dans cette section l'ensemble des propositions techniques retenues sur lesquelles nous allons nous appuyer pour apporter des solutions opérationnelles à la problématique de gestion de l'information et de la connaissance pour la GIZC du Languedoc Roussillon (conformément à ce qui a été présenté dans le chapitre 2).

3.1.1 Adaptation aux utilisateurs

Nous souhaitons proposer un système ouvert à tous qui tient compte, d'une part, des attentes des utilisateurs, et, d'autre part, du réseau d'acteurs impliqués dans la GIZC en Languedoc Roussillon.

Analyse des attentes des utilisateurs

Dans les recommandations générales du DD et de la GIZC¹, comme dans les objectifs spécifiques au programme *Syscolag*, il est jugé nécessaire de valider les propositions par une présentation régulière des choix possibles aux futurs utilisateurs du système d'information de *Syscolag*.

En particulier, ceci est recommandé dans [170] : "[...] *Il faut que le système puisse fournir au décideur l'information qu'il considère nécessaire et suffisante pour trouver la solution à un problème qui lui est posé. Cette information doit lui être présentée sous une forme facilement compréhensible et utilisable*². [...] *Il est donc nécessaire de concevoir le système avec les utilisateurs auxquels il est destiné*".

Une partie du travail de mise en place du système d'information est donc réservée à la ***prise en compte des attentes des utilisateurs***. Cela a été entrepris sous la forme d'entretiens, de réunions, par le biais d'exercices puis d'évaluation collective de la pertinence des solutions proposées par rapport aux attentes formulées. Ainsi un aller retour permanent entre *analyse, formalisation et présentation* des résultats aux futurs utilisateurs a permis d'adapter l'outil au public visé.

L'enjeu est d'éviter un décalage dans la conception du système entre les fonctionnalités proposées, supposées pertinentes par les concepteurs, et les attentes des utilisateurs. En outre, cette démarche permet une appropriation globale de l'outil en cours de développement avant son utilisation opérationnelle.

Prise en compte de l'ensemble des organismes et acteurs impliqués

Nous souhaitons également répondre concrètement aux recommandations de mise en réseau des acteurs impliqués dans les processus de GIZC et favoriser le décloisonnement des institutions pour une meilleure intégration des thématiques, connaissances et ressources nécessaires aux démarches de GIZC en Languedoc Roussillon.

¹En particulier dans le principe 10 de la convention de Rio qui insiste sur la nécessaire participation de tous les citoyens concernés.

²Cela rejoint les conclusions du sommet de Johannesburg où il est jugé nécessaire de présenter l'information sous une forme appropriée, voir partie 2.1.1

Nous proposons donc la mise en place d'un *annuaire régional des acteurs et organismes impliqués dans la GIZC* qui permettra de contacter les acteurs du domaine (scientifiques, gestionnaires ou acteurs économiques) mais aussi de restituer leurs activités et leurs espaces d'intervention. En outre, ce composant pourra servir de référence pour la description des acteurs associés aux *ressources informationnelles* que nous cherchons à localiser. Ainsi nous souhaitons contribuer à la coordination des projets existants ou à venir et éviter d'éventuelles redondances dans les actions engagées.

Représentation des connaissances avec un formalisme partagé

Il existe une telle disparité de formalismes pour la représentation d'analyses et de connaissances du domaine de la GIZC que le *choix d'un formalisme commun*, quel qu'il soit, nous semble être une absolue nécessité. En effet, l'étude de la littérature du domaine montre, selon les disciplines, voire selon les acteurs, une forte hétérogénéité des formalismes de représentations, et ce, parfois, pour illustrer d'une même connaissance [122]. Le lecteur doit ainsi d'une ressource à l'autre comprendre et emprunter les modes de représentations de chacun ce qui nuit globalement aux partages des connaissances et informations à partager. Pour séparer le *fond* de la *forme*, nous proposons donc de formaliser les analyses que nous présentons dans ce travail en utilisant un *mode de représentation unique et partagé par tous*. Ceci était impératif pour optimiser le dialogue entre acteurs en facilitant la compréhension à travers des représentations uniformisées, partagées et interprétables côté système. A ces fins nous proposons donc l'emploi du langage de modélisation *UML* (voir détails de ce standard en partie 3.1.4). Nous répondons ainsi aux recommandations générales sur la nécessité d'une *présentation de l'information sous une forme appropriée* (voir partie 2.1.1) ou, selon la recommandation que nous venons d'évoquer, "*sous une forme facilement compréhensible et utilisable*".

3.1.2 Respect des infrastructures régionales existantes

Pour localiser (voire partager) les ressources (voir caractéristiques en partie 2.1.3) au sein de la communauté SYSCOLAG, plusieurs types d'infrastructures réseau étaient envisageables, avec comme critère de choix, le pouvoir sur la gestion et le contrôle de l'information [129], depuis une solution *centralisée* jusqu'à une solution totalement *distribuée*, en passant par des infrastructures intermédiaires (fédérée ou appuyée par un *médiateur*). Compte tenu de la diversité des partenaires et des ressources produites (voir partie 2.1.3), tant au niveau du contenu que du mode de gestion, il paraissait délicat de proposer un système d'information centralisé. En effet, comme le note [163], l'information disponible est généralement distribuée et personne n'est prêt à partager sa propre information sans garanties [...] cela était le cas dans notre programme de recherche. De plus, une solution centralisée supposerait de définir un modèle de données unifié pour relier toutes les sources d'informations.

Nous avons donc opté pour une *infrastructure de médiation*, que nous définissons comme un système permettant à des utilisateurs divers d'accéder à des sources d'information distribuées, hétérogènes en percevant celles-ci comme si elles étaient issues d'une source unique homogène [116], et tout *en conservant intacts les SGBD des partenaires*.

Cette infrastructure repose sur la *mutualisation des métadonnées* pour décrire et qualifier les ressources partagées (rôle du producteur) et permettre une recherche pertinente (rôle de l'utilisateur).

Compte tenu des problèmes habituellement reprochés aux services de métadonnées par les utilisateurs, nous souhaitons optimiser la consultation et la valuation de certains éléments de métadonnées par l'utilisation d'outils conviviaux. Par exemple, l'*annuaire* des acteurs et des organismes permettra d'automatiser la valuation de certains éléments de métadonnées. En outre, pour la qualité du service de métadonnées, l'élaboration d'un référentiel sémantique (thématique et géographique) constitue un point clé (voir partie 2.3.1) et exige, en complément, un effort collectif de l'ensemble de la communauté *Syscolag*. Le niveau ultime de perfectionnement de ces référentiels correspond en fait à une *ontologie thématique et spatiale du domaine de la GIZC*.

3.1.3 Interopérabilité sémantique

Les problématiques de GIZC s'articulent synthétiquement autour de *concepts thématiques et spatiaux* [143, 87] dont il convient de partager le sens pour la bonne conduite d'une démarche de GIZC, comme cela est exposé en partie 2.1.2. En effet, le contrôle de ces termes clés assure un gain en *qualité d'indexation et de recherche de métadonnées* ainsi qu'un meilleure *partage des connaissances*.

Ainsi, nous proposons la mise en place d'un *référentiel sémantique unique* qui permet l'intégration de l'ensemble des concepts nécessaires à la description des problématiques de GIZC en *Languedoc Roussillon* en utilisant les possibilités de représentation cartographique pour le cas particulier des concepts spatiaux.

Ontologie de la GIZC

Nous souhaitons nous appuyer sur les compétences thématiques des partenaires scientifiques du programme *Syscolag* pour contribuer à la mise en place d'une *ontologie du domaine de la GIZC en Languedoc Roussillon*. A travers l'analyse de l'existant et l'expression de l'expertise des acteurs du programme, nous proposons la mise en place d'un modèle qui permettra d'intégrer ces différents points de vue disciplinaires et de partager les concepts essentiels à la GIZC régionale. Par ce biais, nous obtiendrons la terminologie nécessaire à la mise en place d'un *vocabulaire contrôlé* en complément du *service de métadonnées*. Il s'agit, en pratique, d'une *ontologie thématique et spatiale* qui permettra l'intégration, à terme, des *connaissances* que les partenaires du programme souhaitent partager. En particulier, nous proposons de nous appuyer sur l'*approche écosystémique* plébiscitée par les guides pour formaliser nos propositions et unifier les différentes perceptions du système (voir figure 2.8).

Conscients de la difficulté de cette tâche, par l'expérience précédente de programmes de recherche aux objectifs similaires (voir partie 2.3.2), et convaincus de l'apport des concepts spatiaux dans les processus de GIZC, nous proposons de nous appuyer plus particulièrement sur leurs représentations cartographiques pour faciliter cette démarche.

Le concept spatial comme médiateur

La dimension spatiale, par son omniprésence (implicite ou explicite) dans le concept de GIZC, a fait que *nous avons opté, dans le programme, pour une médiation entre acteurs qui privilégient l'utilisation des concepts spatiaux*. Ceux-ci sont liés aux choix d'espaces appropriés à l'application des mesures de gestion et ont la capacité de résoudre certaines des difficultés constatées lors d'échange entre terminologies disciplinaires (interopérabilité sémantique) [95, 15].

En effet, leurs représentations graphiques constituent un complément à l'utilisation parfois complexe d'outils de contrôle du vocabulaire thématique (ontologies en particulier) et facilitent la définition de référentiels thématique et spatial. Nous proposons donc la mise en place d'une *interface cartographique* pour exploiter les *propriétés des concepts spatiaux* de l'ontologie mise en place. Cette dernière permettra non seulement la visualisation séparée de l'information géographique partagée (constituant une base de référence spatiale pour la GIZC) mais répondra également à l'attente, de la part des utilisateurs, d'interfaces conviviales et simples d'utilisation pour *alléger la saisie ou la consultation de métadonnées*.

3.1.4 Respect des standards / normes de référence

Un des enjeux de notre système est de proposer un outil adaptable aux contraintes du domaine et à son espace d'application qui respecte néanmoins les standards de référence à l'échelle internationale. Ceci permettra d'assurer la *pérennité et la qualité du service proposé* à l'échelle régionale d'une part, et son *interopérabilité* avec d'autres services au plan national et à l'étranger d'autre part.

Il s'agit alors d'arriver à conjuguer un *effort générique de normalisation* avec l'adaptation aux attentes spécifiques des utilisateurs de ce domaine.

En outre, il s'agit de proposer des solutions qui puissent permettre l'implémentation dans un *environnement de développement libre* tant sur le plan des spécifications ou langages utilisés que des logiciels.

Ces deux contraintes ont été respectées, autant que possible, pour l'ensemble des solutions techniques retenues.

Métadonnées et standards pour la gestion de l'information

La mise en évidence de la nécessité des *services de métadonnées* en réponse à la diversité des ressources du domaine et des *SGBD* associés a été faite précédemment.

La question du *choix du standard* est importante pour l'*interopérabilité syntaxique* des systèmes en dehors de l'application régionale avec d'autres systèmes nationaux et à l'étranger voir partie 2.3.1). Compte tenu de l'état de l'art, nous proposons le choix de la norme de métadonnées *ISO 19115*. Le détail des caractéristiques de cette norme (syntaxe et éléments de métadonnées associés) et de son adaptation au contexte du programme *Syscolag* fait l'objet du chapitre 4.

Standards pour la gestion des connaissances

De la même manière que le choix des éléments de métadonnées doit permettre l'interopérabilité avec d'autres systèmes existants, le choix des mots clés de description thématiques et spatiaux doit permettre un partage de la sémantique associée à ces concepts par l'emploi de *vocabulaires contrôlés de référence*.

Globalement les *vocabulaires contrôlés de référence* disponibles sont de trois sortes : *glossaire*, *thésaurus* ou *ontologie*. Nous avons vu que, compte tenu des objectifs affichés par le programme *Syscolag*, notamment du point de vue de la gestion de connaissances, en particulier par le souhait de la mise en place d'une *base de connaissances*, nous préconisons la mise en place d'une *ontologie*.

Il n'existe pas actuellement d'ontologies de référence dédiées à la GIZC [115, 177], cependant, dans certains sous-domaines de la GIZC, on trouve des ontologies en cours de développement

comme, par exemple, *FOS*, ontologie dédiée aux pêcheries (voir partie 2.3.2). Ce constat justifie notre proposition de réutilisation d'ontologies existantes pour *mettre en place un modèle d'ontologie spécifique à la GIZC* et contribuer ainsi aux recherches en cours dans le domaine. Concernant le développement (ou la formalisation) de l'ontologie, il existe, là encore, une variété de standards importante. Nous souhaitons permettre, par le choix du formalisme adopté, de respecter les différents langages de référence pour le Web sémantique (*OWL*, *DAML*, *RDFS*, *RDF*...), tout en restant compréhensible par les utilisateurs de la GIZC, non spécialistes de cette problématique. A nouveau, le choix d'*UML* semblait judicieux, puisque moins complexe à interpréter qu'une syntaxe *XML* et néanmoins compatible, par correspondance, avec l'ensemble des langages d'implémentation du *Web sémantique*. De plus, cela permet de limiter la variété de formalismes employés pour les présentations faites aux utilisateurs.

Standards spécifiques à la mise en œuvre du référentiel spatial En plus du développement d'une ontologie thématique et spatiale de la GIZC qui répond à la volonté de mettre en place un vocabulaire spatial contrôlé, une *interface cartographique* est proposée pour visualiser les représentations cartographiques des concepts spatiaux inventoriés dans l'ontologie [17]. Son développement nécessite également le respect de spécifications ou normes de référence pour permettre, à terme, l'interopérabilité du *serveur de données spatiales* mis en place avec d'autres serveurs du domaine. Comme le montre l'état de l'art et le suggère l'utilisation de la norme *ISO 19115*, l'interface cartographique développée doit respecter les spécifications de l'*OGC* pour la mise en place de *Web Mapping Service*. En effet cette spécification unanimement reconnue actuellement dans le monde de l'Open Source a fait l'objet d'une nouvelle norme *ISO* du *TC 211* : la norme *ISO 19139*.

Afin de proposer une *infrastructure de données spatiale* aussi conforme que possible à l'ensemble de ces recommandations, nous proposons l'emploi du *SGBDR Open Source Postgres* et de sa cartouche spatiale *Postgis* (qui implémente la recommandation *SFS* de l'*OGC*) pour la gestion de l'information géographique de référence que nous collecterons, voire centraliserons, dans le cadre du programme *Syscolag* afin de mettre en place un *cadre de référence spatiale* à l'échelle régionale. La visualisation du contenu de cette *base de données géographiques* sera possible par l'*interface cartographique* développée. Le détail de cette démarche est présenté dans le chapitre 6.

Formalisme de référence

La nécessité d'un formalisme commun nous a conduit à proposer une *approche objet*, plus adaptée à l'*approche écosystémique* préconisée en GIZC. Plus précisément, nous proposons l'utilisation d'*UML*. Spécifiquement conçu pour l'uniformisation des représentations de modélisation, *UML* permet de décrire des interactions entre éléments à travers différentes sortes de diagrammes.

De plus, l'*UML* est le formalisme utilisé pour la présentation de la norme *ISO 19115*, ce qui permet d'homogénéiser la forme de nos présentations.

Outil d'administration peu complexe

Compte tenu de la pluridisciplinarité du domaine dans lequel nous intervenons, les solutions techniques retenues pour la mise en place doivent s'adapter aux compétences des acteurs du domaine. Il s'agit donc, autant que possible, de proposer un outil dont l'administration res-

tera accessible à un public non spécialisé en informatique. Dès que possible les choix techniques pour l'implémentation des objectifs précédents ont respecté cette exigence. Par exemple, pour générer les interfaces graphiques utilisateurs, la base de métadonnées est interfacée en *PHP* (*Hypertext Preprocessor*) plutôt que *Java*, ce dernier étant moins accessible à des personnes non initiées.

De la même manière, nous proposons des interfaces d'administration simplifiées pour les référentiels thématiques et spatiaux afin d'encourager l'utilisation de l'*ontologie* et de l'*interface cartographique* en particulier.

Synthèse de la section 3.1

- Afin de proposer une solution adaptée, nous proposons de nous appuyer sur l'analyse des besoins des utilisateurs.
- Un annuaire électronique portera à connaissance du public les acteurs mobilisés dans la GIZC, leurs coordonnées et leurs activités. Il pourra également servir de source pour l'automatisation des éléments de métadonnées correspondants.
- La solution que nous proposons doit respecter les systèmes de gestion de données en place plutôt que de contraindre à leur adaptation. Pour cela, les *métadonnées* apportent une réponse pertinente à l'*interopérabilité syntaxique* des bases de données environnementales.
- Nous proposons la mise en place d'un *référentiel sémantique* utilisable aussi bien pour le *service de métadonnées* que pour le *partage de connaissances* entre acteurs.
- L'utilisation du *support cartographique* sera centrale dans les modes de gestion de l'information et de la connaissance proposés.
- Afin de ne pas proposer une solution isolée par les choix syntaxiques et sémantiques sous jacent à la gestion de l'information et de la connaissance proposée, nous préconisons l'utilisation de *normes* et *standards* de référence dans ce domaine.
- Le service de métadonnées s'appuiera sur la norme *ISO 19115*.
- Le *référentiel sémantique* sera décrit par le formalisme *UML* compte tenu de sa pertinence pour le travail collaboratif et de l'existence de passerelles avec les langages du *Web sémantique*.
- L'*interface cartographique* s'appuiera sur les spécifications de l'*Open Gis Consortium*.

Dans la section suivante nous présentons l'organisation générale de l'architecture, dans laquelle nous allons positionner les outils que nous venons d'énumérer.

3.2 Architecture du système proposé

Compte tenu du des propositions faites pour la gestion de l'information dans le programme *Syscolag* en section 3.1, nous détaillons désormais la proposition d'*architecture* qui intègre et structure les différents outils proposés.

Pour permettre la réalisation des objectifs annoncés, l'infrastructure proposée repose sur un *service de métadonnées* basé sur l'élaboration d'un profil à partir de la *norme ISO 19115* [16] accessible via un *portail Web*, avec un contrôle sémantique renforcé de la description

thématique et spatiale par le biais d'*ontologies*, et s'appuie sur l'*architecture à trois niveaux* schématisée dans la figure 3.1 ci-après :

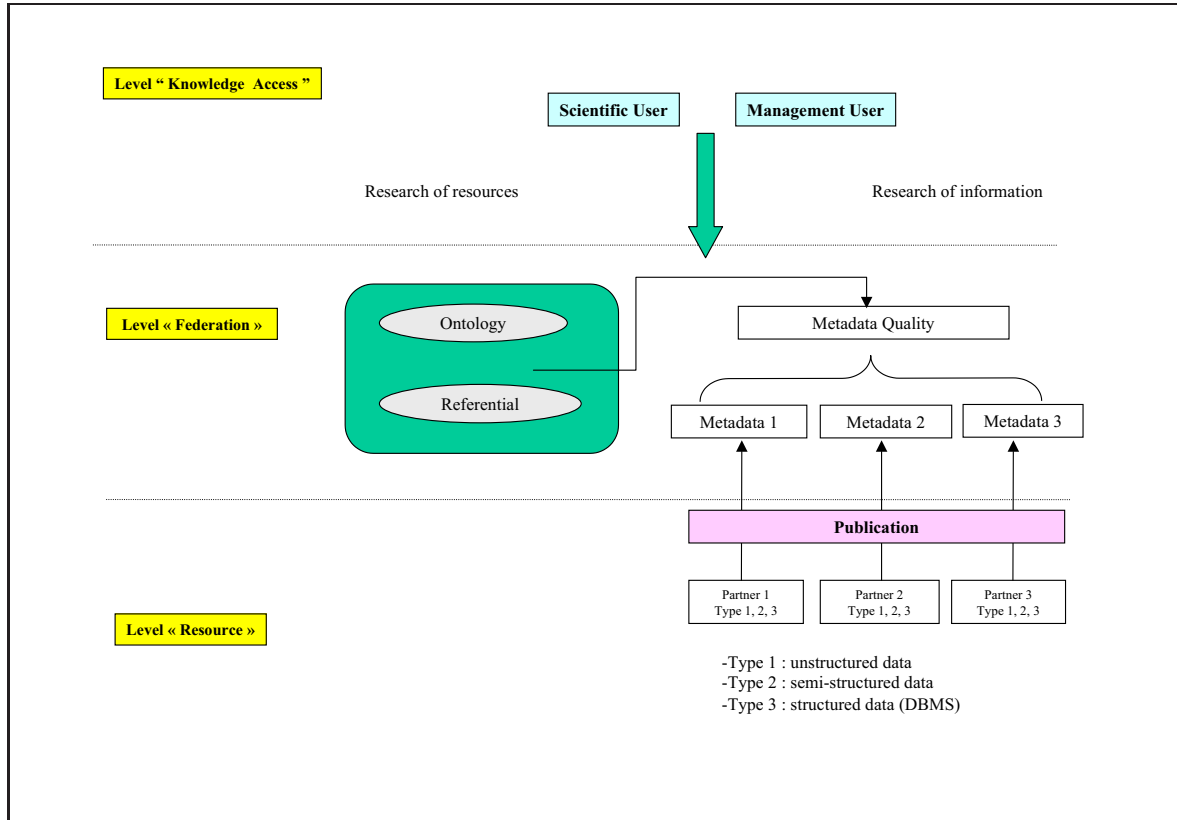


FIG. 3.1 – Les trois niveaux de l'infrastructure Syscolag [16]

3.2.1 Niveau 1 : Accès à la connaissance

Le niveau "accès à la connaissance" se fait par un *portail Web*, dont [101, 100] souligne l'intérêt particulier pour la GIZC, qui permet la *consultation* des métadonnées centralisées au niveau "fédération", qui correspondent à la *description des ressources* issues de structures réparties constituant le niveau "ressources". Cette saisie de métadonnées est effectuée en utilisant des *descripteurs thématiques et spatiaux* issues d'*ontologies* qui assurent la *qualité sémantique et syntaxique de la métadonnée*.

Conformément à la partie 3.1.2 et dans le respect des contraintes d'accès associés aux ressources (voir partie 2.1.3), nous avons donc choisi de séparer la métadonnée de la ressource (même si celles-ci peuvent être liées) afin de n'échanger que la métadonnée (c'est à dire l'information sur la ressource).

Du point de vue de l'utilisateur, la *base de métadonnée*, l'*ontologie* et l'*interface cartographique* sont interrogeables, au niveau du *portail Web*, par le biais d'interfaces graphiques (formulaires) dédiées à la recherche et à la consultation.

3.2.2 Niveau 2 : Fédération

Ce niveau intermédiaire constitue le cœur de l'*infrastructure de médiation*. Il héberge la *base de métadonnées* qui centralise les métadonnées des différents partenaires et les *référentiels thématiques et spatiaux* (en construction continue) et consultables dynamiquement au niveau supérieur (*niveau accès à la connaissance*). C'est à ce niveau que sont stockées les métadonnées et nouvelles connaissances apportées par les partenaires au niveau inférieur. C'est à travers ce niveau charnière que se réalisera l'aiguillage de la navigation des utilisateurs pour permettre la localisation des ressources réparties dans les SGBD des partenaires et l'apport de réponses par l'ontologie sur des connaissances thématiques et spatiales.

Techniquement, ce niveau implémente les recommandations de l'*ISO 19115* pour la base de métadonnées ainsi que les référentiels thématiques et spatiaux stockés dans le SGBDR spatial *Postgres / Postgis*. Le modèle conceptuel de notre base de données est détaillé dans le chapitre 4.

3.2.3 Niveau 3 : Ressources

Conformément à ce qui a été évoqué précédemment (voir partie 3.1.2) les systèmes de gestion de données en place n'ont pas été modifiés physiquement et sont répartis tels quels dans ce niveau. La localisation des ressources (médiation) se fait uniquement par la publication de métadonnées au niveau de chaque organisme impliqué (par des acteurs habilités) qui sont stockées au *niveau fédération* dans la base de métadonnées.

Les interfaces graphiques d'administration nécessaires à l'indexation des métadonnées ou à la saisie de nouveaux éléments pour les référentiels spatial et thématique sont disponibles sur le portail Web par une gestion adaptée des droits des utilisateurs. Ils utilisent les mêmes techniques que les interfaces de consultation.

La consultation de ce système est certes destinée au grand public, mais les fonctionnalités d'administration associées au service de métadonnées, à l'ontologie ou à l'interface cartographique ne sont accessibles que par l'obtention de droits d'accès réservés à certains types d'utilisateurs. Pour terminer la description de notre proposition pour le programme Syscolag, nous proposons dans la section suivante une *vision utilisateur* à travers la présentation des *cas d'utilisation*³ associés au système proposé ainsi que les conditions d'accès.

Synthèse de la section 3.2

- L'architecture proposée s'appuie sur trois niveaux.
- Le niveau *accès à la connaissance*, par le biais d'un portail Web, permet la localisation des ressources informationnelles par des *interfaces graphiques* de consultation interfacées avec le *service de métadonnées* et le *référentiel sémantique*.
- Le niveau *fédération* concentre le cœur des solutions que nous apportons : la base de métadonnées et le référentiel sémantique, garant de la qualité des indexations.
- Le niveau *Ressources* correspond aux différentes ressources informationnelles d'importance pour les processus de GIZC, réparties chez les différents acteurs du domaine, structurées ou non. Ce niveau est interfacé avec la base de métadonnée et le référentiel sémantique pour permettre, directement, sur le *portail Web* la saisie de métadonnées indépendantes des ressources qu'elles décrivent.

³Modélisation UML selon les *uses cases* de Jacobson.

3.3 Les cas d'utilisations

Nous présentons, pour finir, les cas d'utilisations principaux associés au système que nous proposons. En effet, selon le *statut de l'utilisateur* et les *droits d'accès* qui lui sont associés, les *fonctionnalités* du système qui lui sont proposées diffèrent. Il s'agit donc de répertorier, de manière non exhaustive, les principaux acteurs existants et les fonctionnalités⁴.

3.3.1 Les rôles associés aux différents acteurs humains

Conformément aux principes du DD et de la GIZC, le programme *Syscolag*, dans la perspective de mettre en place un *observatoire* du littoral souhaite diffuser, sans contraintes d'accès, l'information et les connaissances disponibles à destination du *grand public* en même temps que pour un public plus spécialisé (*scientifiques, gestionnaires* ou *décideurs*, voir détails en partie 3.3.1). Cependant, l'ensemble des fonctionnalités prévues pour le système ne va pas être ouvert à tous. Les fonctions de description et de modification de métadonnées ou de connaissances sont réservées à un public bien identifié, tout comme les fonctions plus classiques d'administration.

Les différents acteurs

Il existe deux types d'acteurs humains autour du service de métadonnées de Syscolag (voir Figure 3.2) :

- Les **acteurs primaires** sont les *utilisateurs* du système. On peut distinguer deux sous-catégories :
 - **utilisateurs** appartenant à la communauté *Syscolag*, ie droit de *consultation* / *saisie* / *modification* des *métadonnées* et des *connaissances* thématiques et spatiales,
 - **utilisateurs** hors communauté *Syscolag* avec seul le droit de *consultation* des *métadonnées* et les *connaissances* thématiques et spatiales,
- Les **acteurs secondaires** sont les *administrateurs* du système. On peut également distinguer deux sous-catégories :
 - **experts**, responsables de la *validation* des *métadonnées* et *connaissances* thématiques et spatiales saisies ou modifiées par les utilisateurs de la communauté *Syscolag*,
 - **administrateurs** du service et des principaux outils (*service de métadonnées, interface cartographique, ontologie*),

Nous allons à présent préciser la nature des différents rôles associés aux différents acteurs du service de métadonnées.

Rôle "Utilisateur"

Nous distinguons du point de vue de l'utilisateur trois *cas d'utilisation* essentiels au fonctionnement du service (voir Figure 3.2) :

- Concernant le *service de métadonnées* :
 - **rôle de saisie des fiches de métadonnées**. Nous avons en effet considéré que ce sont les utilisateurs appartenant à la communauté *Syscolag* eux-mêmes qui vont alimenter ce système. Indispensable au fonctionnement du système, la réalisation de ce rôle est essentielle, et rend possible les deux rôles décrits ci-après : consultation et modification,

⁴cette vision du système étant susceptible d'évoluer par la suite à la faveur de nouvelles spécifications.

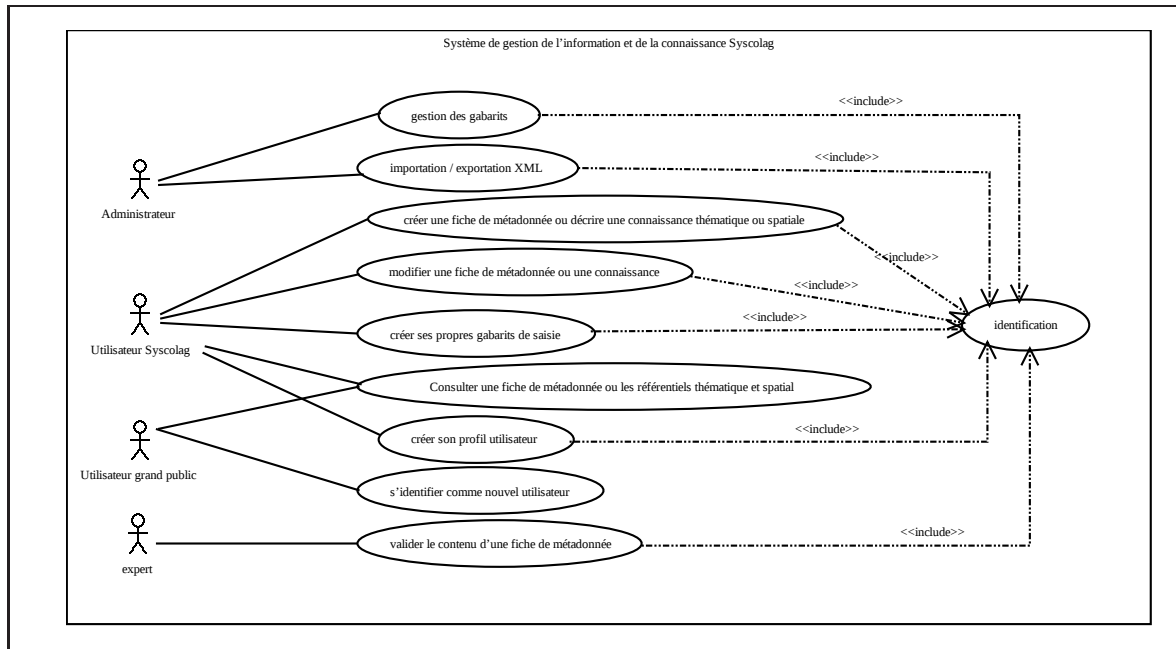


FIG. 3.2 – Synthèse des diagramme des cas d'utilisation des différents acteurs du service de métadonnées

- **rôle de modification des fiches de métadonnées**, lié au rôle de saisie, dans le cas où celle-ci n'est pas achevée ou pas satisfaisante et nécessite, avant diffusion de la fiche de métadonnée, des retouches validées par les experts,
- **rôle de consultation des fiches de métadonnées**, la consultation est un cas d'utilisation primordial, du point de vue de l'utilisateur, en terme d'attractivité. En effet, si le rôle de saisie est a priori rébarbatif (bien qu'indispensable du point de vue du fonctionnement du système et qu'alors dans ce cas l'objectif est de limiter au minimum l'effort de saisie), le rôle de consultation est, quant à lui, attractif puisqu'il s'agit alors de profiter de tout le travail de saisie pour accéder à l'information. Ce rôle concerne l'ensemble des utilisateurs.
- Concernant les *référentiels thématique et spatial* :
 - **rôle de saisie de nouveaux descripteurs et propriétés associées** (concepts et définitions, représentations cartographiques associées...). Lorsque les descripteurs nécessaires à la saisie de fiches de métadonnées ou que des connaissances nécessaires à la constitution des cadres de référence thématiques et spatiaux sont absents ou incomplets, les utilisateurs appartenant à la communauté *Syscolag* peuvent décrire de nouveaux concepts, candidats à la validation par le groupe d'experts qui intégreront le cas échéant les référentiels thématiques et spatiaux.
 - **rôle de consultation des descripteurs** inventoriés dans les bases. Il s'agit d'un service offert à tous, sans contrainte d'accès pour porter à connaissance des utilisateurs le contenu des référentiels thématiques et spatiaux. Le contenu de l'*ontologie thématique et spatiale* étant visualisable par le biais de visualisateurs adaptés (visualisation de réseau sémantique et interface cartographique).
- Puis deux autres rôles optionnels mais qui apportent un confort supplémentaires à la

réalisation des rôles précédents :

- **rôle de création de son profil utilisateur**, nous souhaitons offrir la possibilité, à terme, de "customizer" l'outil à travers l'élaboration de profils utilisateurs,
- **rôle de création de nouveaux gabarits**, propres à un utilisateur ou groupe d'utilisateurs.
- **rôle d'identification comme nouvel utilisateur**, pour ceux qui n'ont jamais utilisé le service et font une demande à l'administrateur pour un droit de saisie de fiche.

Rôle "Expert"

Il s'agira pour l'"Expert" d'évaluer la validité/convenance d'une fiche de métadonnées une fois qu'un utilisateur appartenant à la communauté *Syscolag* l'a saisie et déclarée diffusable de son point de vue (sans quoi elle reste "en cours ou modifiable" dans une table du SGBD prévue à cet effet). L'"Expert" doit alors donner un avis définitif sur la validité effective de la fiche proposée par rapport aux contraintes de l'ISO 19115 et du programme *Syscolag* sur le contenu des éléments de métadonnées avant la diffusion définitive.

Aucune personne n'est pour l'instant concrètement associée à ce rôle, ceci doit être décidé par les partenaires du projet, différentes solutions sont envisageables pour l'acteur expert : un comité de relecture, une personne désigné dans chaque organisme. . .

Cette réflexion est à encore entièrement à mener (et ne concerne pas ce travail), car les compétences nécessaires à l'évaluation de la qualité d'un travail d'indexation sont difficiles à réunir en particulier d'un point de vue thématique.

En résumé, notre proposition est de ne soumettre à l'expert que les fiches déclarées diffusables par les utilisateurs qui saisissent la métadonnée. Elle ne seront effectivement diffusées que lorsque l'expert donne son accord sans quoi elle reste dans la pile des fiches "en cours" de l'utilisateur avec une note d'explication de l'expert sur les raisons de non diffusabilité.

Rôle "Administrateur"

Nous rentrons ici dans des cas d'utilisation plus particuliers qui n'intéressent pas directement les utilisateurs, mais concerne l'administrateur du service (voir Figure 3.2) :

- **rôle de gestion des utilisateurs** (mot de passe, ajout, suppression, modification. . .),
- **rôle d'importation et d'exportation des fiches de métadonnées** (format XML), important pour faire vivre les métadonnées en dehors du contexte régional sur des services équivalents à des échelles différentes.
- **rôle de diffusion des fiches de métadonnées validées par l'expert**. Chaque fiche validée par l'expert doit être envoyée/stockée dans une table accessible uniquement à l'administrateur qui les diffusera effectivement sur le service.
- **rôle d'édition de gabarits**.

3.3.2 Les contraintes d'accès

Dans cette dernière partie nous justifions brièvement la nature des contraintes d'accès que nous associons aux utilisateurs en fonction des différentes fonctionnalités du système.

Accès à la fonction de consultation

Compte tenu de la séparation physique entre la métadonnée et la ressource qu'elle décrit, il n'est pas envisagé d'accès sécurisé à la consultation de la métadonnée, puisque la diffusion de la métadonnée n'engage pas à la diffusion de la ressource.

L'accès à la fonction de consultation est donc entièrement libre. C'est d'ailleurs le cas dans de nombreux services de métadonnées pour lesquels seule la consultation est libre.

Accès aux fonctions de saisie et de modification

A l'inverse, les droits de saisie ou modification d'une fiche de métadonnées ou de connaissance nécessitent une authentification et sont réservés aux membres de la communauté *Syscolag* dans un premier temps.

On peut, par contre, d'ores et déjà réfléchir à la possibilité pour un non membre de *Syscolag* de soumettre des propositions de fiche de métadonnées à la communauté.

Accès à la fonction d'expertise

Ce rôle, par sa complexité, nécessite logiquement une authentification, pour assurer la qualité des fiches diffusées. Mais là encore une phase préalable est nécessaire pour déterminer quelles seront les personnes impliquées dans la relecture des fiches de métadonnées pour réfléchir par la suite à ce que pourront être les modalités d'accès à cette fonction.

Accès à la fonction d'administration

L'accès à la fonction d'administration est sous contrainte d'accès pour des raisons évidentes de sécurité et de complexité des tâches réservées à ce type d'utilisateur.

Synthèse de la section 3.3

- On distingue différentes catégories d'utilisateurs.
- Selon l'utilisateur concerné, l'accès aux différentes fonctionnalités que nous allons proposer est variable.
- La consultation de la base de métadonnées et du référentiel sémantique sont ouvertes au *grand public*.
- La saisie de métadonnées ou de connaissances dans le référentiel sémantique est réservée aux *utilisateurs identifiés*, leur validation pour diffusion l'est également ainsi que l'administration.

Conclusion du chapitre 3

Nous avons présenté dans ce chapitre le contenu général de notre proposition. Nous tenons compte des directives pour la *gestion de l'information et de la connaissance* dans le domaine de la GIZC détaillées dans le chapitre 2. De plus, l'outil proposé tente de conjuguer au mieux l'application de ces directives et les besoins des *acteurs* dans notre contexte spécifique.

Nous recommandons pour atteindre ces objectifs de nous appuyer sur les *métadonnées* pour la *description des ressources informationnelles de la GIZC*. Elles permettront la *localisation des ressources* existantes tout en respectant l'autonomie des modes de gestion des *ressources physiques* par les *acteurs* qui les possèdent. En complément, la mise en place de *référentiels thématique et spatial*, interprétables par les *acteurs* comme par les *machines*, permettra la qualité des *descriptions* des ressources et, en conséquence directe, une localisation plus performante de ces dernières par le *service de métadonnées*. Nous pensons ainsi assister la *mutualisation d'information*, d'une part, et d'autre part, permettre l'*élaboration d'un langage commun* nécessaire au *partage des connaissances* entre les *acteurs*.

Par la mise en place d'un *annuaire électronique* et en contrôlant d'avantage les *descriptions spatiale et thématique*, nous pensons améliorer la description des points clés qui bloquent la gestion efficace des ressources informationnelles de ce domaine (*Qui ?, Quoi ?, Où ?, Quand ?...*) et apporter plus de lisibilité sur ces éléments autour desquels s'articulent globalement les problématiques de GIZC.

Bien qu'adaptée au domaine de la GIZC, un des intérêts de l'*architecture* que nous proposons est qu'elle possède une valeur générique pour les problématiques de *gestion de l'information environnementale*. Nous verrons que ceci est également le cas du *service de métadonnées* et du *référentiel sémantique* proposés.

En outre, nous prenons garde à utiliser des *standards de référence* aussi bien pour les *métadonnées* que pour les *outils de contrôles terminologiques ou géographiques*. Cela doit permettre, à terme, l'*interopérabilité* du système proposé avec d'autres évoluant en dehors de notre contexte régional. A ces fins, nous verrons qu'il est, en particulier, possible de s'appuyer sur des protocoles Web.

Dans les parties qui suivent nous présenterons successivement, le *service de métadonnées* proposé et les *standards de référence* sur lesquels nous nous appuyons pour sa réalisation (chapitre 4), le *référentiel sémantique*, sous la forme d'une *ontologie* (chapitre 5), la *médiation appuyée le support cartographique* (chapitre 6) et nous terminerons enfin par une partie qui illustre le prototype sur lequel nous travaillons actuellement et qui permettra d'évaluer en quoi les réponses proposées répondent à ces objectifs.

Synthèse du Chapitre 3

- Afin de proposer une solution de *mutualisation* des *données* et de la *connaissance* adaptée au contexte de la *GIZC*, nous proposons de nous appuyer sur l'analyse des besoins des utilisateurs et les conclusions du chapitre 2.
- Un *annuaire électronique* rendra compte des *acteurs* mobilisés dans la *GIZC* (leurs coordonnées, activités...).
- Afin de respecter l'autonomie des *modes de gestion de ressources informationnelles* des partenaires (plutôt que de contraindre à leur adaptation), nous proposons l'utilisation de *métadonnées* pour *décrire les ressources informationnelles* et en particulier leur *localisation*. Nous proposons également la mise en place d'un *référentiel sémantique* utilisable aussi bien pour les *descriptions thématique et spatiale* dans le *service de métadonnées* que pour le *partage de connaissances* entre *acteurs*. L'utilisation du *support cartographique* sera centrale dans les applications associées à la réalisation de ces solutions.
- Afin de permettre l'*interopérabilité syntaxique* et *sémantique* avec d'autres *systèmes équivalents* dans ce domaine évoluant en dehors de notre contexte d'application, nous préconisons l'utilisation de *normes* et *standards* de référence pour la *gestion des métadonnées* et des *connaissances*. Nous proposons donc l'utilisation du *jeu d'éléments de métadonnées* normalisé par l'*ISO 19115* et d'une *ontologie spatiale* formalisée à l'aide du *langage UML*. L'*interface cartographique* s'appuiera, quant à elle, sur les spécifications de l'*Open Gis Consortium*.
- L'architecture proposée s'appuie sur trois niveaux accessibles par le biais d'un *portail Web*. Le niveau *Accès à la connaissance* (respectivement niveau *Ressources*) permet la *consultation* (respectivement la *saisie*) des *métadonnées* sur les *ressources informationnelles* et des *connaissances* qui leurs sont associées par des *interfaces graphiques* interfacées avec le niveau *Fédération* composé d'une *base de métadonnées* et d'un *référentiel sémantique* garant de la qualité des *indexations thématiques et spatiales*.
- On distingue différentes catégories d'utilisateurs. Selon l'utilisateur concerné, l'accès aux différentes fonctionnalités que nous allons proposer est variable. La *consultation* de la *base de métadonnées* et du *référentiel sémantique* est ouverte au *grand public* mais la *saisie* de *métadonnées* ou de *connaissances* est réservée à des utilisateurs identifiés de même que l'*administration*.

Chapitre 4

Partager les ressources de la GIZC

Le domaine de la GIZC est caractérisé par l'hétérogénéité des ressources informationnelles produites et des SGBD qui leur sont associés (voir chapitre 2). Les problèmes d'*interopérabilité* n'en sont que plus conséquents et la réponse des *métadonnées* que plus pertinente. En effet, elles permettent la localisation des ressources et respectent leur autonomie étant gérées de manière indépendante. Elles répondent ainsi aux objectifs présentés en partie 3.1.2 et n'imposent donc pas le traitement physique des ressources nécessaire à la mise en œuvre d'un *médiateur*.

Afin de proposer une réponse éprouvée et pérenne à notre problématique ainsi qu'une solution d'interopérabilité au delà de notre contexte d'implémentation, l'usage de *normes de métadonnées* apparaît indispensable à la mise en place du service que nous souhaitons proposer.

En outre, l'importance de l'*espace* dans la GIZC (voir parties 2.1.2 et 3.1.3) a naturellement orienté le choix de la *norme de métadonnée* vers un standard de référence adapté à la gestion de l'information géographique et *adaptable au contexte de la GIZC*. L'étude des standards existants appuyée par les recommandations européennes de référence dans le domaine nous a conduit à choisir la *norme ISO 19115*. Nous proposons une description des recommandations de l'*ISO 19115* et montrons que cette norme spécialisée est adaptable en s'appuyant sur les propriétés des nombreux éléments métadonnées qu'elle propose, en particulier grâce à la possibilité de mettre en place des *profils* spécifiques à différents contextes d'implémentation. En l'absence d'implémentation officielle de cette norme, nous proposons une étude des caractéristiques de quelques outils existants pour l'édition et la consultation de métadonnées.

Nous nous appuyons sur ces propriétés conjuguées à la prise en compte des remarques de nos partenaires afin d'évaluer la pertinence des différents éléments de métadonnées proposés par l'*ISO 19115* pour le contexte de la GIZC en général et celui plus particulier de *Syscolag*. Cela permet la mise en place d'un *profil Syscolag* et de *gabarits* élaborés selon les types de ressources à décrire.

Nous tenons compte de ces analyses pour adapter le modèle conceptuel de l'infrastructure proposée. En particulier nous proposons d'optimiser la gestion des valuations pour les éléments de métadonnées clés dans le domaine de la GIZC pour lesquels l'*ISO 19115* n'apporte pas de réponse directe.

4.1 Intérêt des métadonnées pour le partage des ressources de la GIZC

Lorsqu'on parle d'interopérabilité des bases de données environnementales, ce n'est pas tant le partage physique des ressources contenues dans les *SGBD* qui compte que de décrire aux partenaires l'*existence* et les *caractéristiques* de ces ressources (en particulier, dans le cadre de la GIZC, ce que [47] appelle le *contexte* de production de la donnée (voir partie 2.1.3)). En effet, il s'agit, lors d'une démarche de recherche d'information, dans un premier temps, de savoir qu'une ressource pertinente pour l'usage envisagé *existe*, puis, si la description qui en est fournie convient à l'utilisateur, dans un deuxième temps, de la détailler en vue d'un éventuel partage, dans la mesure des contraintes d'accès et de diffusion qui lui sont associées. Le *partage physique* de la donnée n'est, en effet, qu'un résultat possible à l'issue de la consultation de la description qui lui en est faite.

Ce sont les métadonnées qui sont les plus pertinentes pour fournir ces *descriptions* ce *contexte*. Après avoir défini ce concept, les différentes sortes de métadonnées et les objectifs associés à leur utilisation, nous justifierons l'intérêt d'utiliser une *norme de métadonnées* et, en particulier, la pertinence du choix de la norme *ISO 19115* pour notre domaine d'application.

4.1.1 Interopérabilité des bases de données environnementales

Nous revenons dans cette partie sur les caractéristiques des bases de données environnementales et des ressources qu'elles gèrent pour mieux comprendre par la suite la pertinence de la réponse apportée par les *métadonnées*.

Contexte général

Plus encore que dans d'autres domaines, la gestion des ressources environnementales en général [131, 159] et de la GIZC en particulier est associée à des problèmes d'*interopérabilité* des bases de données, dus à l'hétérogénéité des ressources elles-mêmes (voir partie 2.1.3) et aux nombreux SGBD existants et disséminés dans les multiples organismes impliqués dans le domaine. En l'état, ces systèmes de gestion ne peuvent interopérer, et il n'est pas envisageable de contraindre les acteurs du domaine à un mode de gestion unique qui ne saurait s'adapter à la diversité des usages et ne satisferait finalement personne. Or nous avons vu que le *principe 19 de la convention de Rio* insiste sur la nécessité de s'appuyer sur des SI permettant de transférer de l'information pertinente d'un Etat à l'autre pour améliorer la décision pour la préservation de l'environnement (voir partie 2.1.1).

En pratique, il s'agit donc d'établir des correspondances entre les systèmes existants sans les modifier pour atteindre l'*interopérabilité* qui fait actuellement défaut à la démarche de DD et *intégrer* les informations et connaissances dispersées ou réparties chez les différents partenaires. En effet, les *bases de données environnementales* sont caractérisées par des contraintes lourdes qui rendent les tâches de partages habituelles particulièrement difficiles [117].

Nous rappelons dans la partie suivante quelques caractéristiques essentielles que nous avons déjà abordées en partie 2.1.3.

Caractéristiques des ressources du domaine

Les ressources informationnelles isolées ou gérées dans les bases de données environnementales sont caractérisées par [61, 16, 37, 20, 43] :

- l'hétérogénéité des ressources (thématiques, formats, systèmes de projection, précision. . .) et des systèmes de gestion associés,
- la disparition rapide de certaines ressources ou de l'information sur leur existence (en lien avec le travail temporaire ou les changements de matériel et l'inefficacité des systèmes de gestion),
- le coût des ressources élevé : données cartographiques, observations issues de capteurs variés (images de télédétection, bouées marines, forages. . .)
- les contraintes d'accès : données confidentielles (exemple des positions géographiques des récifs artificiels pour empêcher le chalutage à proximité des côtes. . .),
- la diversité des usages possibles d'une même ressource : il n'est pas toujours possible de prévoir quel usage peut être fait d'une ressource.
- le problème d'échange sémantique inhérent au partage des ressources : diversité de description liée à la diversité des acteurs et des usages associés.

Pour l'ensemble de ces raisons, on comprend que le *partage physique* des données environnementales est complexe à mettre en œuvre, voire peut être utopique à court terme. Une première étape consiste à partager l'information sur l'existence de données en des termes compréhensibles par l'ensemble des acteurs concernés, en faisant état de *contraintes d'accès* parfois complexes.

Nous montrons dans la section suivante, dans la perspective d'une médiation efficace, que les *métadonnées* apportent une réponse pertinente puisque leur rôle consiste à décrire les caractéristiques des ressources informationnelles.

4.1.2 Définition du concept de métadonnée

Le terme *métadonnée* est ancien mais l'usage des métadonnées et leur importance connaissent un regain d'attention récent [88]. Il illustre un besoin croissant de gestion des informations et des connaissances lié à la nécessité actuelle de gérer des données produites aussi rapidement qu'elles peuvent disparaître.

En particulier, les métadonnées permettent la localisation de ressources informationnelles isolées ou dispersées dans les différents *SGBD* sans imposer la révision de leurs méthodes de gestion [16]. Or, comme nous l'avons mis en avant dans nos objectifs le respect des modes de gestion des partenaires est essentiel (voir partie 3.1.2)

Dans son acception première, le terme ***métadonnée*** signifie "*données sur les données ou données qui renseignent sur certaines données et qui permettent ainsi leur utilisation pertinente*" [21].

Nous élargissons cette définition puisque pour nous le niveau "*méta*" renseignera *tout type de ressource* (voir définition en partie 2.1.3). L'importance des métadonnées va s'avérer primordiale dans le contexte qui est le nôtre, car si l'on considère, d'une part, la diversité des ressources et d'autre part, l'aspect distribué du système concerné (les différents *SGBD* des partenaires mobilisés dans la GIZC, voir partie 3.1.2), elles vont constituer une réelle *structuration partagée et homogène*. Nous affirmons que les métadonnées constituent un composant de base pour une gestion et une utilisation judicieuses des ressources [15]. Elles fournissent à tout utilisateur potentiel les moyens de savoir :

- *Quelles ressources existent ?*
- *Lesquelles correspondent à des besoins spécifiques ?*
- *Comment les localiser voire y accéder dans la limite de ses droits ?*

Les métadonnées sont un aiguillage vers la ressource pertinente pour l'utilisateur. Elles ont

un rôle d'*identification* et de *spécification* de la ressource "référence" et peuvent ainsi servir de base aux moteurs de recherche [43]. L'importance de ce rôle est telle qu'on peut considérer qu'elles sont au moins aussi importantes que la donnée elle-même, car elles préviennent des mauvais usages ultérieurs à l'appropriation physique de cette dernière.

Les métadonnées répondent donc à notre problématique puisque pensées pour la *gestion* d'une masse de données hétérogènes (*localisation*, voire *accès* et *traitement* adéquat de données). En outre, elles ont pris un intérêt particulier avec la numérisation récente¹ de l'information [155, 37].

4.1.3 Métadonnées et NTIC

Associée à l'émergence des *NTIC*², la métadonnée est d'origine ancienne, pour exemple, la gestion des bibliothèques incorpore toujours un catalogue et une indexation qui constituent l'équivalent d'une base de métadonnées actuelle, avant l'arrivée des *NTIC*. L'avènement du *document numérique* a simplement remis le concept à la mode et a conduit à l'élaboration de standards d'échange [15, 168].

Face aux contraintes d'éloignement géographique des organismes, justification passée de l'absence ou de la difficulté de collaboration entre les hommes, les *métadonnées* associées au *NTIC* permettent sans rien changer aux systèmes ou aux implantations actuelles de mutualiser en une ou plusieurs bases dédiées et interopérables les connaissances de tous ceux qui désirent partager leurs ressources informationnelles en gardant une liberté sur leur diffusion physique (voir partie 3.1.2). Elles participent ainsi au décloisonnement des organismes, condition indispensable à la mise en place d'une GIZC.

Le *Web* en particulier (qui ouvre des perspectives prometteuses pour la GIZC [100]) est, selon *Tim Berners Lee* du *W3C*, un espace privilégié pour bénéficier de l'apport des *métadonnées* pour la gestion de l'information dans le contexte du *Web sémantique* [22] :

"La métadonnée est une information compréhensible (interprétable) par la machine (ordinateur) sur des ressources d'information du *Web* ou d'autres sources de données. L'expression "compréhensible par la machine" est la clé. Nous parlons ici de l'information que des logiciels traitent pour nous faciliter la vie, nous assurer que nous obéissons à nos principes, à la loi, vérifier que nous pouvons croire en ce que nous faisons et exécuter tous les travaux régulièrement et rapidement. **La métadonnée définit bien la sémantique et la structure.** La métadonnée a été appelée ainsi à sa création et est encore actuellement une information sur des ressources du *Web*, ou des données sur des données. Plus tard, lorsque les métadonnées, les langages, les moteurs de recherches seront plus évolués, ils formeront une base plus forte pour l'information "compréhensible par la machine" du *Web*, sur n'importe quoi : les gens, les choses, les concepts et les idées. Gardons dans notre esprit ce modèle (préconisé), bien que la première étape soit de réaliser un système d'indexation de l'information en réseau."

Il est intéressant de relever dans cette explication les perspectives associées à l'utilisation des métadonnées, en particulier celles qui proposent de décrire les ressources informationnelles variées (à l'image de celles qu'on peut consulter sur le *Web*) en incluant les descriptions de *personnes*, d'*idées*...ce qui rejoint l'ouverture que nous avons souhaité donner dans notre définition de *ressource* en partie 2.1.3.

Cependant, les métadonnées dans le contexte du *Web* sont, lorsqu'elles existent, le plus fréquemment *sous format textuel libre* et se résument généralement à des *mots clés*. Ce sont

¹A l'échelle des temps géologiques.

²Nouvelles Technologies de l'Information et de la Communication.

les flux croissants d'information liés aux nouvelles technologies et les besoins d'échanges et de catalogage qui en découlent, qui ont amené les premiers travaux de structuration en vue de normaliser l'utilisation des métadonnées [15] par des jeux d'éléments de métadonnées contrôlés.

On distingue en pratique différents types d'éléments de métadonnées.

4.1.4 Typologie d'éléments de métadonnées

Il existe une hiérarchie dans la chronologie et l'importance des questions auxquelles les métadonnées doivent répondre (voir partie 4.1.2).

En effet, il s'agit, *en premier lieu*, de savoir si des ressources existent. Si tel est le cas, il faut, *dans un deuxième temps*, regarder, parmi les ressources trouvées, celles qui conviennent effectivement à l'usage envisagé. *Finalement*, une fois les ressources pertinentes localisées, il s'agit de prendre connaissance des contraintes d'accès pour un éventuel partage physique de la donnée.

Nous donnons ci-après les définitions de différents types d'éléments de métadonnées (associés à la réalisation de ces trois étapes) que nous utiliserons dans la suite du document.

Les métadonnées principales

Elles permettent à l'utilisateur de répondre à la question de l'existence d'une ressource en interrogeant la métabase par des requêtes construites autour de concepts qui répondent aux questions : *quoi, où, quand, qui, comment* ? [76, 78, 77, 131].

Pour la mise en place de *métadonnées principales*, il est souhaitable que le nombre d'éléments de métadonnées à décrire n'excède pas une dizaine d'éléments pour alléger la saisie des métadonnées [79].

Généralement, les *éléments de métadonnées principaux* se retrouvent dans tous les standards et souvent dans leurs *noyaux* (voir définition en partie 4.2.2).

Les métadonnées secondaires

Les *métadonnées secondaires*, encore appelées *métadonnées élargies*, détaillent la capacité des données à satisfaire les besoins d'une application déterminée, à obtenir les renseignements nécessaires à une utilisation adéquate de la donnée [76, 78, 131].

Les métadonnées pour fins de diffusions

Ce sont les éléments de métadonnées requis pour aider une organisation à faire connaître à l'extérieur les ressources existantes, accessibles ou disponibles pour diffusion [76, 78, 131].

Une fois déterminé le jeu d'éléments de métadonnées nécessaire pour satisfaire les besoins d'une application donnée, il est important, pour le succès d'un service de métadonnées, de respecter une certaine discipline dans la saisie des fiches.

4.1.5 Objectifs associés à l'utilisation des métadonnées

Afin d'assurer la bonne marche d'un *service de métadonnées*, il est essentiel de faciliter la tâche de description de la métadonnée et d'observer quelques règles simples :

- *Décrire les ressources en temps réel* : synchronisation de la *production* et de la *description* des ressources pour assurer la qualité de la description qui en est faite (tâche qui, idéalement, revient au(x) producteur(s) de la ressource).

- *Faciliter la description des métadonnées* sans que les utilisateurs n’aient à devenir des spécialistes des métadonnées (par une automatisation partielle ou totale, de manière transparente et par des *interfaces graphiques (GUI)* appropriées), d’où l’importance des métadonnées principales (voir partie 4.1.4).
- *Commencer au plus tôt à documenter les ressources d’information* est une priorité. Le fait que les organisations aient, pour cela, utilisé des moyens différents comporte des inconvénients mais moins que de ne pas avoir de documentation du tout [60, 40].
- *Employer un jeu d’éléments de métadonnées standardisé* pour améliorer la gestion des métadonnées.

Nous développons dans la partie suivante l’importance de ce dernier point.

4.1.6 Les normes de métadonnées

Dans cette partie, après avoir justifié l’intérêt de l’utilisation des *standards* et des *normes de métadonnées* et le choix de celle qui est appropriée à notre problématique, nous allons rappeler, les caractéristiques générales et l’historique des normes et standards d’intérêt pour notre contexte d’application. Nous justifions alors le choix de la norme *ISO 19115*.

Organisation d’une norme

On peut généralement décrire une *norme de métadonnées* comme une agrégation de *rubriques*, chacune composée d’un jeu d’éléments de métadonnées structurés (qui entretiennent des relations) relatifs à la description d’une catégorie particulière de l’information décrite : *contraintes d’accès, qualité de la donnée...*(voir Figure 4.1).

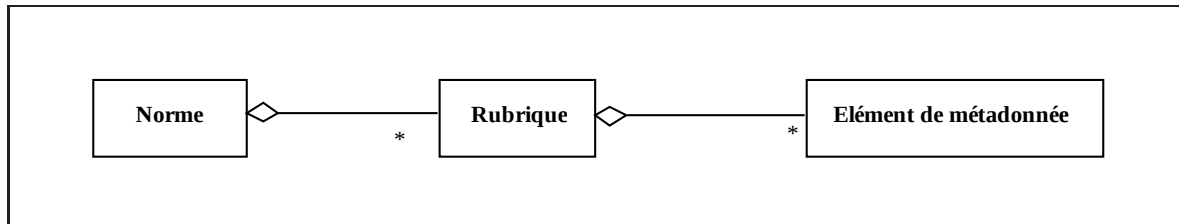


FIG. 4.1 – Vue simplifiée de l’organisation générale d’une norme

Intérêt d’une norme

Malgré l’intérêt commun évident qu’il y a à limiter au maximum la multiplication de standards de description répondant aux mêmes objectifs, il en existe, malgré tout, une diversité notable. Une *ressource* peut, en effet, être décrite de manière différente (différents éléments de métadonnées mais aussi différents *concepts* pour les valuer) selon les contextes, mais avec la même *intension* (significations identiques des différents éléments et termes employés, synonymie).

Par conséquent, les normes possèdent au moins deux avantages de taille :

- Elles évitent d’avoir à créer un nouveau jeu d’éléments de métadonnées. Il n’y a aucun intérêt à refaire des normes dont le contenu et les objectifs sont foncièrement proches, en particulier pour une *norme de découverte* (voir définition en partie 4.1.6). En effet,

l'utilisation d'un standard assure un *gain de temps et de qualité pour le jeu de métadonnées utilisé* puisque ce dernier est élaboré à la suite d'un consensus de spécialistes du domaine pour unifier un ensemble de standards préexistants.

Si chaque organisation documente ses propres ressources différemment, cela ne pose pas un grave problème si les métadonnées sont uniquement utilisées au niveau interne ; cela le devient toutefois si les utilisateurs éprouvent beaucoup plus de difficultés à chercher ce qu'ils veulent. D'où l'importance d'adopter et d'utiliser des normes [40]".

- On peut toutefois réinventer un jeu de métadonnées qui remplit les objectifs visés, mais, en sachant cependant que, si on souhaite interopérer avec d'autres systèmes à l'extérieur, cette solution devient un obstacle à l'échange et à la compréhension des métadonnées produites en dehors de son groupe de travail³. Comme le souligne [131] :

"Il est recommandé d'observer une logique dans le contenu et le style des métadonnées pour permettre aux utilisateurs de faire rapidement des comparaisons au niveau de l'adéquation des données provenant de différentes sources.[...]. Sans standardisation, il est plus difficile de trouver des comparaisons intéressantes sans lire et apprendre de nombreux styles de gestion des métadonnées. La prévoyance est aussi encouragée par la conformité aux normes".

Pour ces raisons, durant les dernières années, les efforts de description des ressources existantes qui utilisent des normes ou standards, vont croissants.

Comme nous l'avons vu, *dans un premier temps*, il est nécessaire, par l'usage de *métadonnées principales*, de porter à la connaissance de l'utilisateur l'existence de ressources indépendamment de leurs secteurs ou disciplines d'origines et sans détail de leurs caractéristiques techniques. *Dans un deuxième temps*, l'usage de *métadonnées secondaires* procure à l'utilisateur une finesse de description de la ressource informationnelle qui lui permet d'estimer sa pertinence pour une utilisation à des fins spécifiques.

A ces différents objectifs correspondent respectivement les *normes de découverte et de domaine de référence*. Nous allons désormais présenter celles d'entre elles qui permettent de répondre aux objectifs que nous nous sommes fixés.

Normes de découverte

L'usage de *normes* longues et complexes risque rapidement d'aboutir à des informations incomplètes ou dépassées et de finir par ruiner l'objectif global de la documentation [61]. La communauté internationale s'est donc davantage intéressée aux *normes de métadonnées de "découverte"*, dont le but est de permettre aux utilisateurs potentiels de trouver *qui détient quelle information* tout en permettant aux producteurs et aux intermédiaires de soutenir l'effort de description.

Par *norme de découverte*, on entend un ensemble restreint d'*éléments de métadonnées principaux* visant à *découvrir l'existence d'une ressource d'un type quelconque dans des domaines disciplinaires ou sectoriels variés* [76, 10], sans toutefois décrire ses caractéristiques de manière trop détaillée. Ces normes sont donc complémentaires de *normes spécialisées*, spécifiques à un domaine, qui interviennent à un niveau de description plus fin, voir partie suivante).

Il existe pour répondre à ce besoin de découverte un *standard* de fait qui connaît un succès croissant : le **Dublin Core**⁴ (*DC*). Le *DC* n'est pas une norme mais une initiative

³On peut rappeler le principe 9 du DD pour mettre en évidence l'intérêt d'une norme pour la GIZC pour participer au décloisonnement et à une interopérabilité des systèmes à l'échelle internationale.

⁴<http://purl.org/DC/>.

internationale des communautés de bibliothèque et de publication [41]. Elle est basée sur un consensus et est entièrement ouverte. Le *Dublin Core* existe uniquement à des fins de découverte mais il suscite un intérêt croissant chez un grand nombre d'industries [156]. Le *Dublin Core* a également été récemment repris au niveau européen par le *CEN* dans le cadre d'une convention émanant d'un groupe de travail, i.e. un mécanisme volontaire impliquant une série de partenaires issus de divers secteurs⁵.

Normes spécialisées

En complément de ces *normes de découverte génériques*, les *normes spécialisées* permettent, par l'introduction d'éléments de métadonnées complémentaires (dédiés à la *description de types de ressources particuliers*), une finesse de description supplémentaire *tout en conservant les éléments de métadonnées principaux*.

En particulier, les ressources environnementales mobilisées dans le cadre du développement durable nécessitent des descripteurs adaptés à la caractérisation de ressources fortement marquées par les *concepts spatiaux, thématiques et temporels* [131, 13]. Elles ont donné lieu à l'élaboration d'une série de standards dédiés à la description des ressources d'*information géographique*.

Jusqu'à maintenant, l'un des principaux problèmes était la diversité des standards de métadonnées existants parmi les communautés nationales et internationales, en particulier dans le *domaine de l'information géographique*. En effet, dans ce domaine, la spécificité et la diversité des productions et de la localisation de l'information géographique, la dualité producteur-utilisateur ont poussé à l'édification de divers standards : le *FGDC* étant le précurseur [69]. Cette actuelle difficulté de choix semble désormais se résoudre et pourrait progressivement devenir obsolète avec l'avènement d'un consensus international sur la manière de décrire les métadonnées pour l'information géographique (grâce au développement d'une norme internationale reconnue). Les travaux coordonnés de l'*ISO* et de l'*OGC* sont en train de proposer une norme de fait : *ISO 19115*.

Historique

En attendant qu'une norme internationale agréée soit mise en place, on a assisté, durant les dix dernières années, à l'émergence d'initiatives (parfois concurrentes) nationales, voire locales, visant à l'élaboration de standards de référence pour documenter les ressources d'information géographique [10, 76, 78] :

- Aux **Etats-Unis**, la principale initiative est la suivante : ***FGDC-CSDGM*** ou *Content Standard for Digital Geospatial Metadata*, développée en 1994 par le *Federal Geographic Data Committee* américain (*FGDC*) comme partie intégrante de l'*infrastructure des données spatiales nationales*. Il s'agit d'une norme complète de métadonnées comprenant 219 champs pour décrire une ressource. Elle est obligatoire pour les agences fédérales⁶. C'est le standard précurseur du domaine.
- En **Europe**, une initiative équivalente a été menée par le *CEN*, le *Comité Européen de Normalisation*, en réponse au *FGDC*. Il s'agit de l'initiative **CEN/TC 287 ENV 12657**. Il s'agit d'une norme préliminaire volontaire (i.e. non obligatoire) élaborée en 1997-98 par le *CEN*. C'est aussi une norme complète de métadonnées qui a servi de base

⁵<http://www.cenorm.be/news/press-notice/metadata.htm>.

⁶<http://www.fgdc.gov/metadata/constan.html>.

à de nombreuses initiatives européennes. Le *Comité Technique 287* est à présent inactif puisque ses résultats ont été pris en compte dans le développement de la norme *ISO 19115*⁷.

- En **Nouvelle Zélande** et en **Australie**, une norme a été proposée en 1995 afin de mettre en place un *système d'information géographique global* pour ces deux pays. Pour cela le standard proposé s'est appuyé sur le *CSDGM*, en limitant toutefois volontairement le niveau de détail des éléments de métadonnées retenus.

Cette période a eu le mérite de croiser différentes analyses qui ont contribué à cerner les *éléments de métadonnées* nécessaires pour la description de l'*IG*. Cependant, l'interopérabilité à l'échelle internationale entre des services de métadonnées utilisant ces différents jeu d'éléments standardisés se heurte à des problèmes de correspondances syntaxiques et justifient la synthèse de ces de ces efforts (et d'autres) en un jeu d'éléments normalisé qui correspond à la norme *ISO 19115*.

Synthèse de la norme l'ISO 19115

L'*International Standardisation Organisation (ISO)* est en train de développer, au sein de son *Comité Technique 211 ISO/TC 211*, une famille de normes liées à l'*information géospa-tiale*⁸. Parmi celles-ci, la norme *ISO 19115* [2] établit une *synthèse des précédents standards de métadonnées la description des ressources du domaine de l'information géographique* (voir partie précédente) afin de proposer un mode de description unique, unanimement partagé et reconnu de par le monde, qui permet l'*interopérabilité* des systèmes à l'échelle nationale et international.

Le travail sur cette norme a duré cinq ans, a réuni les principaux standards du domaine et a été publié en tant que norme internationale en mai 2003. En tout, 33 pays [2, 112] et 12 organisations observatrices ont participé au développement de l'*ISO 19115*. Comme certaines organisations avaient déjà proposé un standard de métadonnées, la nécessaire adaptation de ces standards préexistants a été envisagée *via* la notion de *profil* intégrée par l'*ISO* et présentée en partie 4.2.2. Par exemple, la *version 3 de la norme FGDC* est devenue un *profil* de la norme *ISO 19115*. Ainsi, le *CEN* a approuvé cette norme et en fait ainsi un point de référence pour toutes les *institutions européennes*.

En réponse à cette norme, dans le domaine de la gestion de données géographiques et environnementales, deux initiatives européennes majeures recommandent l'utilisation de l'*ISO 19115* comme norme de métadonnées.

4.1.7 Les recommandations européennes

A la suite de (voire pendant) l'avènement de l'*ISO 19115*, des recommandations européennes guident les développements en cours ou à venir en spécifiant les normes de référence pour les institutions européennes. En particulier, pour le domaine environnemental et l'information géographique, deux initiatives font référence. Elles donnent des recommandations aux organismes qui souhaitent mettre en place de telles infrastructures.

⁷<http://forum.afnor.fr/afnor/WORK/AFNOR/GPN2/Z13C/index.htm>.

⁸<http://www.statkart.no/isotc211/pow.htm>.

MADAME

Le projet *MADAME*⁹ vise la mise en place de méthodologies d'accès aux *données* et *métadonnées* européennes afin de valoriser les données publiques [40].

Entre autres, il recommande aux autorités locales pour décrire leurs ressources l'utilisation du *Dublin Core* comme *métadonnées génériques de découverte multidisciplinaire* et celle de l'*ISO 19115* comme *norme spécialisée pour l'information géographique* :

"les autorités locales doivent se conformer aux normes de métadonnées acceptées au niveau international comme la norme ISO 19115 pour les informations géographiques. En d'autres termes cela signifie que si une organisation utilise un modèle de métadonnées différent, elle doit l'adapter à la norme ISO, du moins pour les éléments obligatoires. Si une organisation part de zéro, le mieux est de commencer par utiliser les éléments obligatoires de la norme ISO".

Concrètement cela signifie qu'il faut créer des bases de métadonnées pour documenter les ressources d'information avec des éléments du *Dublin Core* et de l'*ISO 19115*, de façon à pouvoir générer automatiquement un résultat conforme aux deux selon la nature de la demande.

INSPIRE

*INSPIRE*¹⁰ est une initiative de l'*Union Européenne* pour mettre en place une *infrastructure pour l'information spatiale* en Europe afin de rendre ce type d'information plus accessible et interopérable dans le cadre de projets au service du développement durable¹¹ [159] (voir partie 2.3.1).

En effet, le constat actuel est que les informations relatives aux éléments géographiques sont recueillies en l'*absence de toute coordination* et sur la base de *méthodes et spécifications non uniformes*, ce qui aboutit à des *lacunes et des difficultés pour comparer les données* en Europe. Afin d'améliorer la cartographie en *Europe*, la *Commission européenne* a donc lancé l'initiative INSPIRE. En pratique, il s'agira d'une base de données comprenant des informations géographiques cohérentes qui appuiera ainsi les politiques de protection de l'environnement ainsi que le développement des infrastructures, l'agriculture et la navigation maritime. Une nouvelle directive récente prévoit l'obligation pour les États membres de l'*UE* de placer l'information géographique sur un réseau électronique *accessible au public* et d'en assurer l'*harmonisation progressive*.

Margot Wallström, Commissaire européen à l'environnement, s'est félicitée de cette proposition : "*INSPIRE* harmonisera les informations géographiques dans l'*UE*, ce qui contribuera largement à améliorer la conception, la mise en œuvre et le suivi des mesures environnementales. Notre capacité à protéger la biodiversité, à lutter contre la pollution et à prévenir incendies et inondations s'en trouvera améliorée. Une approche commune au niveau de l'*UE* est à l'évidence nécessaire dans ce domaine pour tirer le meilleur parti des données existantes."

Afin d'atteindre ces objectifs, *INSPIRE* demande la prise en compte des spécifications du *Dublin Core* en tant que *norme de découverte* pour les technologies de l'information et celle de l'*ISO 19115* pour décrire l'IG. A ce titre, *INSPIRE* souhaite procurer dans son catalogue une vue *Dublin Core* sur les informations et services inventoriés, ce qui requiert une *correspondance*

⁹MADAME, Methods for Access to data and Metadata in Europe (1998-2000).

¹⁰Infrastructure for SPatial Information in Europe - Infrastructure d'information spatiale dans l'Union européenne.

¹¹http://www.notre-planete.info/actualites/actu_387.php.

entre les éléments de métadonnées de l'*ISO 19115* et ceux du *Dublin Core*.

Pour la documentation des données, les standards recommandés sont les suivants :

- *ISO 19115* comme normes spécialisés pour la description de l'*information géographique*,
- *Dublin Core* comme standard de découverte,

En outre, *INSPIRE* est responsable de la mise en œuvre d'un profil européen de l'*ISO 19115* (voir détail en partie 4.2.2).

Sur la base de ces constats et recommandations pour la gestion de l'*IG* dans les projets de *DD* et conforté par des choix similaires dans des systèmes de gestion de l'information sur les zones côtières ou marines [115, 177, 137], nous avons donc choisi pour le projet de *GIZC Syscolag*, de nous conformer aux spécifications de l'*ISO 19115*. Nous allons maintenant présenter cette norme, de manière à mettre en avant ses qualités ainsi que ses lacunes pour l'adapter à notre domaine d'application.

Synthèse 4.1

- Compte tenu des caractéristiques des ressources du domaine, l'interopérabilité des bases de données environnementales est complexe à mettre en œuvre. Un premier niveau de réponse à cette problématique est de permettre la localisation des ressources informationnelles existantes, en vue, à terme, de leur éventuel partage physique.
- Les métadonnées permettent une description fine des ressources informationnelles et sont particulièrement adaptées pour la gestion des ressources informationnelles associées aux *NTIC* (au Web en particulier).
- Les *métadonnées* apportent donc une réponse pertinente à la problématique générale de description des ressources informationnelles nécessaires pour la *GIZC* et donc à leur localisation. Il existe différentes catégories de métadonnées destinées à décrire différentes caractéristiques des ressources informationnelles.
- Les *normes de métadonnées* présentent un grand intérêt pour l'efficacité du jeu d'éléments de métadonnées utilisé et l'interopérabilité avec d'autres services qui implémentent les mêmes recommandations.
- Une *norme de découverte* suffit à *localiser* les ressources informationnelles, cependant, compte tenu de l'*importance de l'espace et de l'information géographique dans le domaine de la GIZC*, nous recommandons plus particulièrement une *norme de métadonnées pour l'information géographique*.
- Compte tenu des standards existants, des recommandations européennes actuelles (*MADAME* et *INSPIRE*) et de leur adhésion consensuelle à la norme *ISO 19115*, nous proposons d'implémenter cette norme pour la mise en place de notre service de métadonnées.

Dans la section suivante, nous présentons comment le choix de l'*ISO 19115* permet une adaptation à notre contexte d'implémentation.

4.2 Présentation de l'ISO 19115

Comme nous l'avons vu, l'*ISO 19115* est une norme fédératrice, synthèse "exhaustive" et récente des différents standards préexistants, qui permet d'envisager le passage à une interopé-

tabilité des bases de données environnementales au plan international en unifiant les modes de description des ressources informationnelles qui y sont contenues.

A ces fins, nous allons montrer que cette norme propose un jeu d'éléments de métadonnées conséquent et largement adaptable (quantitativement et qualitativement) à différents contextes ou niveaux de précision, par la mise en place de *profils*.

4.2.1 Description des rubriques de l'ISO 19115

Comme de nombreux standards, la norme *ISO 19115* propose de stratifier les éléments de métadonnées en *rubriques* ou *sections* (voir partie 4.1.6 et Figure 4.1). Elle est composée de *douze rubriques principales* (voir Figure 4.2).

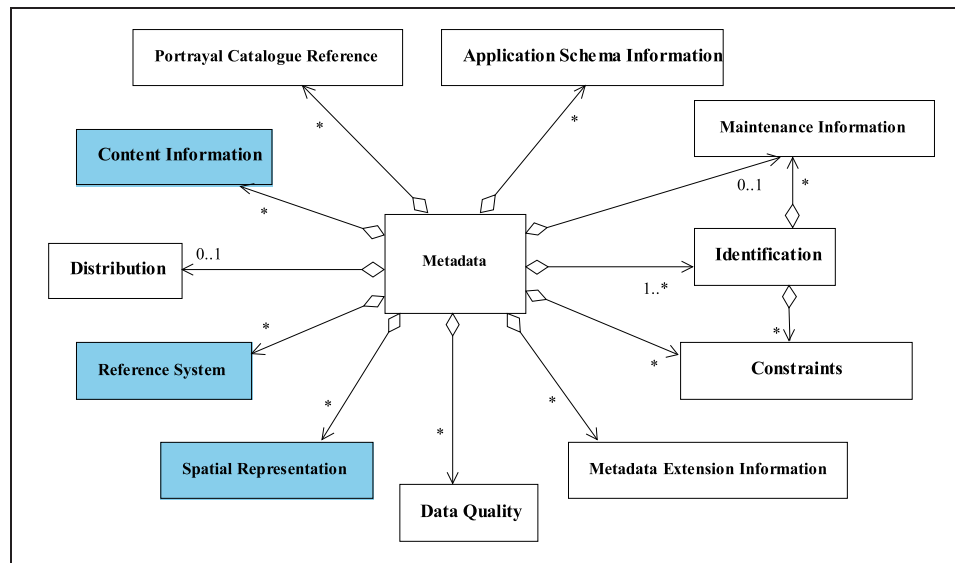


FIG. 4.2 – Les différentes rubriques de l'ISO 19115 [16, 2]

Neuf rubriques sont communes à tout type de ressource :

- **Metadata** : Information pour l'identification des principales caractéristiques de la métadonnée. Cette rubrique regroupe des *éléments de métadonnées principaux* pour la gestion des différentes *fiches de métadonnées* (date de création, standard et langage utilisé...),
- **Identification** : Information pour l'identification des principales caractéristiques de la ressource. Ce sont des éléments de métadonnées principaux pour décrire la *ressource* (comme le *Titre*, la caractérisation spatiale et thématique...). *Seules ces deux premières rubriques sont obligatoires, les suivantes sont optionnelles*,
- **Constraints** : Information sur les contraintes d'accès à la ressource. Ces éléments de métadonnées permettent de savoir si une ressource est en *accès libre* ou *limité* à certains utilisateurs et organismes, cette rubrique peut également être appliquée pour décrire les contraintes d'accès associées à la fiche de métadonnée elle-même (existence d'un lien d'agrégation entre *MD_Metadata* et cette rubrique, voir Figure 4.2),
- **Data quality** : Information sur la qualité de la ressource. Dans cette rubrique, il s'agit de rendre compte du *contexte* dans lequel la ressource a été produite, en particulier

de sa *généalogie* : la *méthodologie employée*, la *qualité du résultat* (en particulier d'un point de vue spatial - la précision d'une carte par exemple - ou temporel...), d'attribuer une note/un indicateur de qualité. Ceci permet de savoir si une ressource est compatible avec les attentes d'un utilisateur pour une application donnée (traitements envisagés après acquisition...).

- **Maintenance information : Information sur la maintenance et la mise à jour des données/métadonnées.** Cette rubrique informe sur la fréquence de mise à jour de la ressource (*régulière* ou non), son état (ressource devenue *obsolète*) et les parties de la ressource¹² concernées par d'éventuelles mises à jour,
- **Distribution : Information sur les modalités de distribution de la ressource.** Ces éléments permettent de savoir *qui* contacter et *où* se procurer la ressource (le *nom*, *adresse*, *téléphone*... du *distributeur*) mais aussi sur quel support la ressource est disponible (*numérique*, *papier*, le *format* sur lequel elle est enregistrée ou disponible...), par quel moyen elle est accessible (*en ligne*, *courrier*...), éventuellement à quel *prix*...,
- **Metadata extension information : Information pour l'ajout d'éléments de métadonnées.** Dans le cas où les éléments proposés par l'ISO ne suffisent pas à décrire les caractéristiques de certaines ressources, cette rubrique permet d'*étendre le modèle de métadonnées* en proposant de *nouveaux éléments de métadonnées*. Pour cela, il faut renseigner les propriétés des éléments créés. La manipulation de cette rubrique présente moins d'intérêt pour l'utilisateur que pour l'administrateur. Par contre, elle est capitale pour l'*adaptation de la norme à un contexte donné*.
- **Portrayal Catalogue Reference : Présentation.** Cette classe contient un unique attribut dont l'objectif est de décrire le catalogue de règles de (re)présentations utilisées dans la ressource,
- **Application Schema Information : Schéma conceptuel de la ressource.** Ce paquet contient la description du *schéma conceptuel* et du *logiciel* utilisés pour générer la ressource décrite.

Remarque – En général, ces deux dernières rubriques disparaissent des profils sont difficiles à renseigner et peu claires pour les utilisateurs. De même une partie de la rubrique *qualité* n'est exploitable que pour de l'information géographique au sens strict [37, 20, 38].

Trois rubriques servent uniquement à la *description des ressources de type information géographique* (en bleu sur la Figure 4.2) :

- **Spatial representation : Information sur le mode de représentation spatiale de la ressource.** Selon la nature de la ressource décrite, les éléments de métadonnées de cette rubrique permettent la description détaillée des représentations *vecteur* ou *raster* à travers des attributs tels que la *résolution*, les caractéristiques du *géoréférencement*, la nature des *objets géométriques* présents...(voir Figure 4.3),
- **Reference system : Information sur le système de référence.** On trouve dans cette section les informations relatives à la description des *systèmes de référence* spatial (géographique et vertical) et temporel utilisés, le *système géodésique*, *ellipsoïde*...,
- **Content information : Information sur le contenu de la ressource.** Ces éléments portent sur des caractéristiques plus techniques, en particulier pour le cas d'images (*aérienne*, *satellite*...) pour lesquelles la compréhension fait appel à des notions plus fines de *télédétection* (caractéristiques des *capteurs*, *longueurs d'ondes*, catalogue des *objets*,

¹²L'ISO 19115 parle de *Scope* (en français *portée*) de la maintenance, qui signifie, en informatique : partie du programme à laquelle s'applique la déclaration [149].

description du contenu des données raster...),

A ces 12 rubriques s'ajoutent deux **data types** communs à de nombreuses rubriques :

- **Geographic extent** : **Information sur l'étendue (spatiale, temporelle) de la ressource.** Cette rubrique contient les éléments qui décrivent l'étendue spatiale (verticale et horizontale) et temporelle de la ressource. Avec cette catégorie, nous commençons donc à choisir des éléments qui caractérisent le contenu de la ressource,
- **Citation** : **Information sur la citation et le/les responsable(s) d'information.** Ce paquet de type de données contient les éléments de métadonnées qui concernent la manière de citer la ressource (référence bibliographique...) et les personnes à contacter pour des informations complémentaires sur la ressource (nom, adresse, fonction, moyen de les contacter, ...période de présence au bureau).

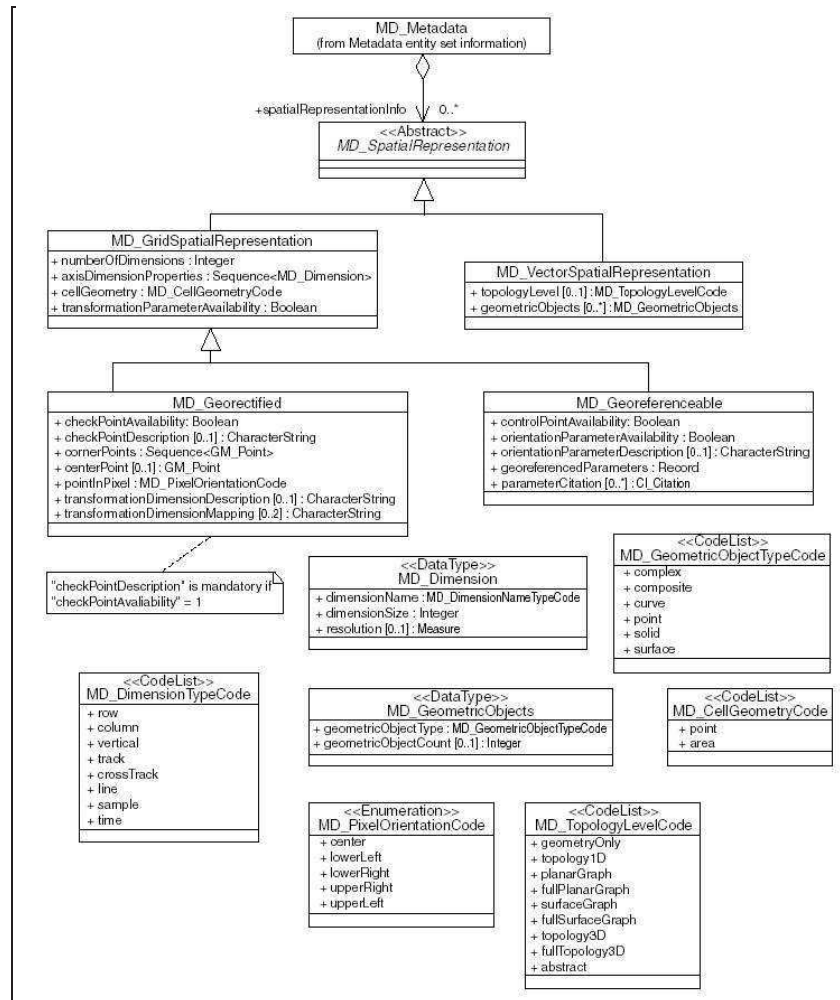


FIG. 4.3 – Détail du diagramme de classes UML pour la rubrique *Spatial representation* de l'ISO 19115 [2]

Chacune de ces rubriques comporte diverses sous-rubriques, obligatoires ou optionnelles, qui structurent plus finement les relations entre les éléments de métadonnées (voir exemple en

Figure 4.3).

En pratique, la présentation du standard est effectuée en tant que *méta-modèle* décrit en formalisme *UML* (*Unified Modeling Language*) instance du *MOF*, (*Meta Object Facility*) [24]. Ce méta-modèle présente les rubriques ou sections sous forme de *packages* pouvant partager des *classes* communes. Ces *classes* entretiennent des *relations* de différentes natures : *agrégation*, *spécialisation/généralisation*, *association*...

A présent, nous allons présenter les différents types d'éléments de métadonnées proposés par la norme *ISO 19115* afin d'évaluer leur intérêt et de justifier les choix que nous avons effectués pour implémenter cette norme dans le contexte de la GIZC, en prenant en compte les remarques de nos partenaires sur leur pertinence pour décrire les ressources de la GIZC.

4.2.2 Description des éléments de métadonnées de l'ISO 19115

A l'ensemble des *rubriques* que nous venons de décrire est associé un *dictionnaire de données* de 409 éléments. Une partie seulement des éléments de métadonnées qui y apparaissent est *obligatoire*, l'autre est *optionnelle* et son utilité est soumise à l'appréciation des utilisateurs en fonction du domaine et du contexte d'application. (voir notion de *profil* et *gabarit* en parties 4.3.1 et 4.3.2).

Nous proposons de décrire, dans cette partie, la notion de *profil* et les propriétés des éléments sur lesquelles nous allons pouvoir nous appuyer par la suite pour améliorer l'adaptation de l'outil aux spécificités du domaine de la GIZC (voir parties 4.3.1 et 4.3.2).

Définition d'un profil de métadonnées

Un *profil*, ou encore *adaptation*, est une *extension* d'un *méta-modèle* selon la terminologie de l'*OMG* qui a pour objectif de définir le *méta-modèle de norme spécifique à une communauté*. Un profil précise donc les options de mise en place d'une norme afin de répondre à un besoin particulier. La norme *ISO 19115*, décrite dans un *méta-modèle UML*, a donc repris la notion de *profil* de l'*OMG*.

Par essence, un *profil* ne contredit pas la norme à laquelle il se réfère, pour cela il en respecte toujours le *cœur* (ou *noyau*¹³) composé d'*éléments obligatoires* généralement peu nombreux. En outre, il est possible d'ajouter, en complément, des *éléments optionnels* (pris dans la liste des éléments génériques recommandés par la norme) voire d'en créer de nouveaux (spécifiques au contexte d'implémentation). Cette propriété est capitale pour la norme *ISO 19115*, elle va nous permettre de conjuguer une normalisation agréée à l'échelle internationale avec la nécessaire adaptation régionale au programme de GIZC *Syscolag* (voir partie 4.3.1). Par contre, il peut intégrer des éléments n'existant pas initialement dans la norme, ou *éléments étendus* (voir Figure 4.4 et rubrique *Metadata extension information* présentée en partie 4.2). Ils viennent compléter celle-ci en y apportant des éléments de description utiles dans le contexte précis pour lequel est utilisé le *profil*. Ainsi, un *profil* permet d'adapter *culturellement et linguistiquement* une norme internationale *aux particularités d'un pays ou d'une région*. Dans notre cas, cette propriété assouplit, le cadre a priori rigide de la norme *ISO 19115* [16, 37, 155].

Dans la partie suivante nous détaillons la nature des quelques *éléments de métadonnées obligatoires* qui composent le *noyau* de l'*ISO 19115*.

¹³ *Core* en anglais.

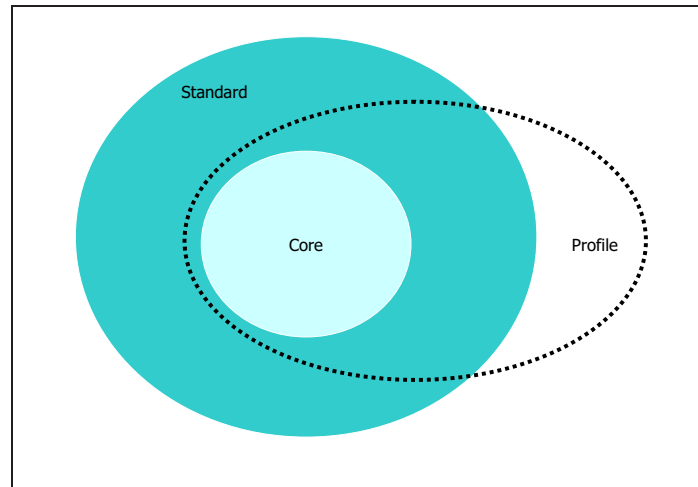


FIG. 4.4 – Profil de l'ISO 19115 [2]

Les éléments du noyau de l'ISO

Comme nous l'avons vu, l'ISO 19115 est issue d'une synthèse des caractéristiques des standards préexistants. Elle bénéficie, de ce fait, d'un nombre conséquent d'éléments valables, a priori, pour la description de n'importe quel type de ressource.

De plus, cette norme n'impose qu'une petite partie des éléments de métadonnées qu'elle propose : ce sont les éléments du *noyau* qui sont des *éléments de métadonnées principaux* (ou *métadonnées de découverte* voir partie 4.1.4) généralement communs à tous les standards. La lourdeur présumée de l'ISO 19115, en tant que *norme spécialisée*, est donc une fausse contrainte largement gommée par la possibilité de supprimer les éléments jugés inutiles dans un contexte donné (et à l'inverse la possibilité d'en créer de nouveaux).

En effet, au sens strict, le noyau de l'ISO 19115 regroupe 12 éléments [112, 37] qui sont des *éléments de métadonnées obligatoires* provenant des deux seules rubriques obligatoires de la norme (*Metadata* et *Information*, voir Figure 4.2), soit les éléments de métadonnées suivants :

- Dans la rubrique **MD_Metadata** :
 - **contact**, qui est une classe du "Data Type" *Cl_Contact* ne contenant pas d'éléments obligatoires,
 - **dateStamp**,
- Dans la rubrique **MD_Identification** :
 - **citation**, qui est une classe du "Data Type" *Cl_Citation* contenant deux éléments obligatoires :
 - **title**,
 - **date**, qui est une classe du "Data Type" *Cl_Date* contenant deux éléments obligatoires :
 - **date**,
 - **dateType**,
 - **abstract**,
 - **language**,

Ce noyau restreint est finalement très proche des 15 éléments optionnels constitutifs de l'initiative *Dublin Core (DC)* comme le montre [112, 133] avec la publication d'un *tableau*

Dublin Core Element	ISO 19115 Element
DC.Creator	Metadata.mdContact.RespParty (role=principal investig.)
DC.Contributor	Metadata.mdContact.RespParty (role=resource provider)
DC.Publisher	Metadata.mdContact.RespParty (role=publisher)
DC.Date	Metadata.Ident.Citation.resRefDate
DC.Date.created	Metadata.mdDateSt
DC.Title	Metadata.Ident.Citation.resTitle
DC.Title.alternative	Metadata.Ident.Citation.resAltTitle
DC.Description.abstract	Metadata.Ident.idAbs
DC.Description.tableOfContents	
DC.Subject	Metadata.Ident.descKeys
DC.Type	Metadata.Ident.Citation.citIdType
DC.Format	Metadata.Ident.Format.formatName
DC.Format.extent	Metadata.Ident.distTransOps.transSize
DC.Format.medium	Metadata.Ident.distTransOps.medName
DC.Identifier	Metadata.Ident.Citation.citId
DC.Source	Metadata.Ident.Citation.citRespParty
DC.Relation	Metadata.Ident.mdParentId
DC.Language	Metadata.contInfo.catLang
DC.Coverage.spatial (DCMI Box)	Metadata.dataIdInfo.geoBox
DC.Coverage.spatial (TGN)	Metadata.dataIdInfo.geoDesc.geoId
DC.Coverage.spatial (polygon)	Metadata.dataIdInfo.datExt.geoEle.polygon
DC.Coverage.spatial (vertical)	Metadata.dataIdInfo.datExt.vertExtent
DC.Coverage.temporal	Metadata.dataIdInfo.datExt.exTemp
DC.Rights	Metadata.Ident.mdConst.useConsts

FIG. 4.5 – tableau de correspondances entre éléments de métadonnées du *DC* et de l'*ISO 19115* [112]

de correspondance entre les éléments de ces deux standards (voir Figure 4.5). L'utilisation de cette correspondance permet à partir d'une implémentation minimale de l'*ISO 19115* de générer une sortie *DC*.

Ainsi, il est possible de répondre aux recommandations de *MADAME* et *INSPIRE* qui préconisent l'emploi du *DC* en *norme de découverte* et celle de l'*ISO* comme *norme spécialisée* (voir partie 4.1.7).

En dehors de ces quelques éléments de métadonnées constitutifs du noyau de l'*ISO 19115*, il existe une grande variété d'éléments composant les autres rubriques qu'il est possible d'adapter à notre contexte.

Autres propriétés remarquables des éléments de l'ISO

D'après les recommandations de la norme *ISO 19115*, en plus d'être *obligatoire* ou *optionnel*, chaque élément de métadonnée possède des propriétés particulières qui concernent son mode de *valuation*.

Idéalement, le *contrôle de la valuation* assure la *qualité de l'indexation* et, par la suite, de la *recherche des métadonnées* produites (terminologies et significations partagées). Cependant, selon la nature des éléments considérés, les modes de valuation proposés par l'*ISO 19115* diffèrent.

On peut, en fonction de leur mode de valuation, distinguer les différentes catégories d'éléments suivantes :

- **les éléments de valuation libre.** En pratique, ces éléments de métadonnées corres-

pondent aux attributs des classes de la norme de type *CharacterString* (voir exemples en Figure 4.3). Après analyse, on peut distinguer deux catégories parmi ces éléments :

- Pour certains d’entre eux, il n’est pas possible, dans l’absolu, d’automatiser le remplissage par des valeurs prédéfinies : par exemple, pour des éléments tels que *Résumé*, *Titre*... La valeur de ces éléments est en effet propre à chaque ressource.
- Pour d’autres, la proposition de valeurs prédéfinies est potentiellement envisageable mais, compte tenu de la valeur générique de la norme *ISO 19115*, cette tâche revient à ceux qui implémentent la norme selon leur contexte d’application. Parmi ces éléments, on peut citer, par exemple, les *contacts* (*acteur*, *organisme*...) ou les *mots clés* pour lesquels, la norme se contente de suggérer la mise en place d’un contrôle de la valuation en faisant référence à un *Thésaurus d’origine* pour les *mots clés* choisis.
- **les éléments de valuation contrôlée.** Pour ces éléments, l’*ISO 19115* propose, soit des *CodeLists*¹⁴ proposant différentes valeurs possibles pour la description d’un élément de métadonnées, soit des *types d’attributs* particuliers (*DataType*) faisant parfois appel à d’autres normes ou formats. Il s’agit notamment de la norme *ISO 19104*, qui porte spécifiquement sur la terminologie usitée dans la communauté qui manipulent l’*IG*, et de la norme *ISO 19108* "GI-Temporal schema" pour l’indexation temporelle.

En outre, on remarquera que la *visibilité* des *attributs* de toutes les *classes* présentées est une propriété dont la valeur est toujours "public" (indiqué par le signe + devant les attributs des classes, voir exemple en Figure 4.3). Lors de l’implémentation, il est néanmoins possible de moduler la visibilité des attributs.

Dans la partie suivante, nous allons détailler quelques choix possibles pour implémenter les recommandations de l’*ISO 19115*, permettre la valuation de ces différents gabarits ainsi que leur stockage et restitution à l’utilisateur.

4.2.3 Implémentation de l’ISO 19115 : état de l’art

La norme *ISO 19115* ne formule que des recommandations [2] mais l’implémentation officielle de cette norme (la norme *ISO 19139*) n’est pas encore finalisée¹⁵. En attendant que cette dernière soit disponible, de nombreux organismes en charge de la gestion de ressources spatiales ont implémenté les recommandations de l’*ISO 19115*. Certains ont opté pour l’implémentation à l’aide de *SGBD (SQL)* et d’autres à l’aide d’une approche semi-structurée (*XML* et *DTD*). Les logiciels sont soit disponibles via un réseau (généralement via le *Web*) soit installés en local sur un *PC*.

Or, compte tenu de notre contexte d’implémentation, nous avons vu que nous souhaitons favoriser l’utilisation d’outils dans un *environnement de développement libre*, d’administration peu complexe (par les langages de développement utilisés), mais aussi d’utilisation simple par des *interfaces graphiques* qui simplifient les tâches de saisie a priori complexes pour le public ciblé et permettent notamment d’exploiter au mieux le *support cartographique* et les outils de *contrôle du vocabulaire*.

Nous proposons à présent une description représentative des choix techniques possibles concernant les *outils d’édition de métadonnées* qui implémentent l’*ISO 19115*.

¹⁴Les *CodeLists* sont des classes *UML* de type *énumération* mais ouvertes à l’introduction de nouvelles valeurs. Ils sont propres à l’*ISO 19115*.

¹⁵Ceci devrait être mis en place en collaboration avec l’*OGC* [37].

ArcCatalog

ArcCatalog est un logiciel propriétaire destiné aux *géomaticiens* pour une utilisation sur poste fixe.

Contexte et objectifs *ArcCatalog* est un produit de la gamme *ArcGis* développé par la société *ESRI* pour optimiser la gestion de l'information géographique à travers l'emploi de normes de métadonnées telles que *FGDC* et *ISO 19115*. L'utilisateur bénéficie de processus d'automatisation pour la description de certains éléments de métadonnées, en particulier ceux gérés par *ArcGis* (système de référence spatial, Bounding Box...), et de fonctions d'exportation et de stockage de métadonnées directement au format *XML* (voir illustration en Figure *refarccatalog*).

ArcCatalog est une application orientée données pour la *gestion*, la *localisation* et la *navigation* dans les données spatiales. Cet outil permet de créer et d'administrer les bases de données spatiales. Les utilisateurs y définissent leur schéma de données, spécifient et utilisent les métadonnées (voir Figure 4.6).

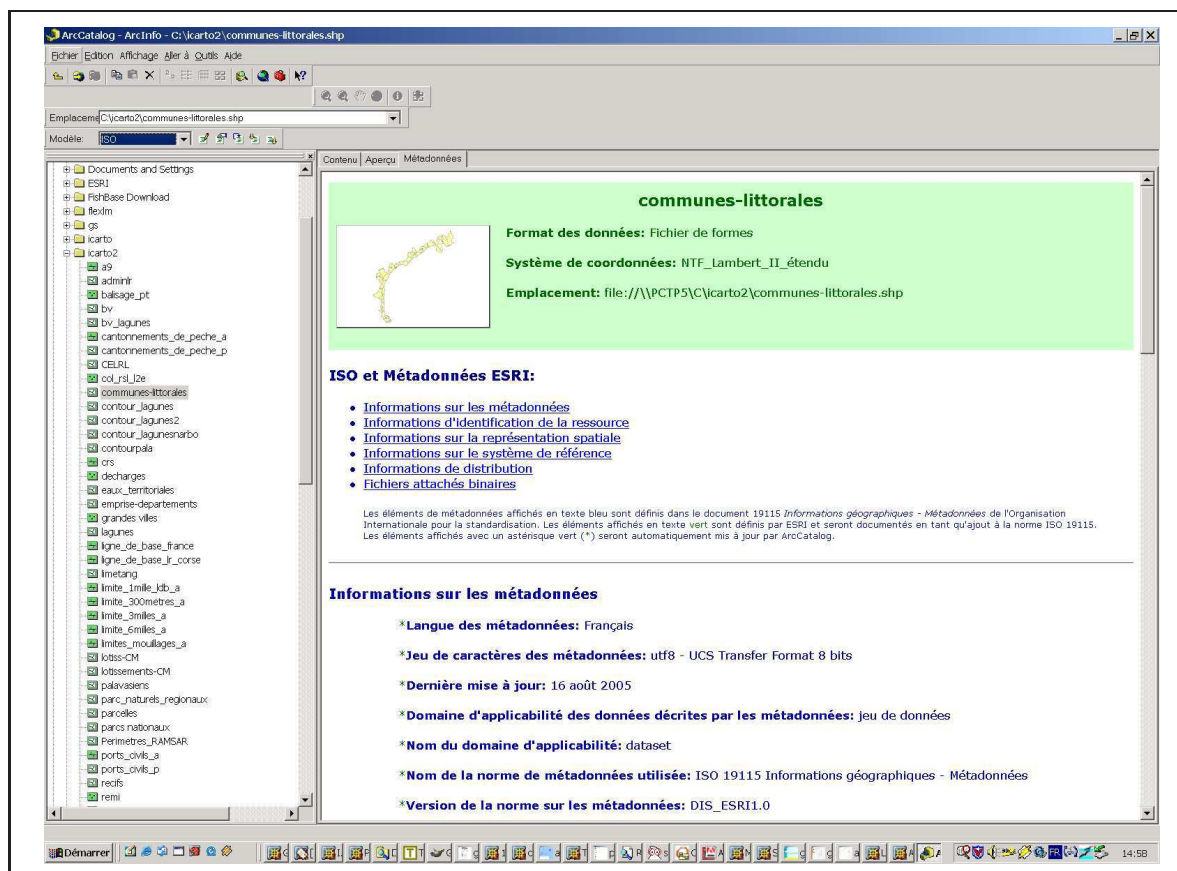


FIG. 4.6 – Exemples de fichiers de métadonnées dans ArcCatalog

Caractéristiques techniques Les *caractéristiques techniques* d'*ArcCatalog* sont les suivantes :

- *Environnement de base de données* : ESRI (fichiers *.shapes, ArcSDE ou connexion à d'autres *SGDB* comme *Oracle*),
- *Statut de développement* : stable,
- *Public visé* : utilisateurs des produits ESRI,
- *Licence* : propriétaire (ESRI),
- *Système d'exploitation* : Windows,
- *Langages de programmation* : *VBA*,
- *traductions* : multi-lingues,
- *Interface utilisateur* : logiciel indépendant¹⁶.

Si la solution apportée par *ArcCatalog* est intéressante, en particulier en permettant un lien direct entre production de la donnée et automatisation de la métadonnée, elle n'est cependant, contrairement à d'autres, ni accessible *via* le Web ni gratuite.

MDWeb

MDWeb est une application Web *multi-norme et multi-lingue* dans un *environnement de développement libre*.

Contexte et objectifs Le programme "*Réseau d'Observatoires de Surveillance Ecologique à Long Terme*" (ROSELT) a pour objectif de promouvoir et de soutenir des programmes de surveillance environnementale à long terme dans des zones arides soumises à la dégradation des terres, en prenant en compte la gestion de la biodiversité et le rôle de ces zones, en relation avec les changements climatiques globaux et la lutte contre la désertification.

La zone concerne l'Afrique de l'Ouest et un programme de gestion de l'information est en cours pour partager les ressources existantes dans cette zone.

En réponse à cette nécessité, *MDWeb* est une application logicielle multi-norme, multi-lingue accessible sous forme d'un *portail Web* (développé en *PHP* et *SQL*, utilisation de *POST-GRESQL* et *Postgis*)¹⁷. L'environnement de développement est entièrement libre.

Les développements en cours s'intéressent à l'optimisation de la gestion par la mise en place d'outils de contrôle du vocabulaire mais aussi à l'optimisation de l'utilisation d'une interface cartographique pour la résolution de requêtes spatiales complexes sur les métadonnées.

Caractéristiques techniques Les *caractéristiques techniques* de *MDWeb* sont les suivantes :

- *Environnement de base de données* : Postgres - Postgis, MySQL,
- *Statut de développement* : stable,
- *Public visé* : Communauté du programme *ROSELT*,
- *Licence* : GNU General Public License (GPL),
- *Système d'exploitation* : Windows, Mac, Linux,
- *Langages de programmation* : *PHP*, *SQL*, *XML*,
- *traductions* : français, anglais, portugais,
- *Interface utilisateur* : application Web (Web-based),

¹⁶Logiciel *stand-alone* en anglais.

¹⁷Ce logiciel est développé à Montpellier pour l'*US désertification* de l'*IRD*.

M³Cat

M³Cat est une application Web *multi-norme et multi-lingue*.

Contexte et objectifs *M³Cat* (*outil de catalogage de métadonnées multilingue et multi-norme*) est un outil libre, développé par *Intelec Géomatique*¹⁸, qui permet l'entrée et la gestion de métadonnées de jeux de données géographiques.

Cet outil de *catalogage* de métadonnées offre les fonctionnalités suivantes :

- la possibilité de cataloguer les métadonnées en utilisant plusieurs profils de normes (FGDC, GILS, Biodiversité, gouvernement du Québec et ISO 15046-15 du TC 211...);
- le soutien multi-lingue, en utilisant les directions préconisées par JTC1 d'ISO (Information Technology) et la traduction semi-automatique des métadonnées;
- des aides au catalogage;
- l'intégration d'un outil d'analyse sémantique pour la gestion des mots-clés;
- le fonctionnement dans plusieurs environnements technologiques;
- l'interface aux services de recherche distribués ANSI/NISO Z39.50, notamment ISITE et MetaManager.

Caractéristiques techniques Les *caractéristiques techniques* de *M³Cat* sont les suivantes :

- *Environnement de base de données* : Access, Oracle,
- *Statut de développement* : stable,
- *Public visé* : géomaticiens,
- *Licence* : GNU General Public License (GPL),
- *Système d'exploitation* : Windows,
- *Langages de programmation* : ASP (IIS), SQL, XML,
- *Traductions* : français, anglais,
- *Interface utilisateur* : Web-based,

L'environnement de développement n'est donc pas entièrement libre.

Geonetwork

*Geonetwork*¹⁹ est une application Web développée par la *FAO* qui implémente la norme *ISO 19115*.

Contexte et objectifs *Geonetwork* est un logiciel libre multi-lingue développé par la *FAO*. Il permet de disposer d'un catalogue de métadonnées accessible depuis un portail Web pour la saisie et la recherche de métadonnées à travers l'implémentation de la norme *ISO 19115* et de la norme *ISO 23950 (Z39.50)*.

Ce logiciel s'appuie sur des fichiers *XML* et *XSL* et une *interface cartographique (InterMap OpenSource)* pour la génération de formulaires de saisie ou de consultation de métadonnées (voir Figure 4.7).

Les **objectifs** de *GeoNetwork* sont de faciliter le partage de l'information thématique à référence spatiale entre différentes institutions pour :

¹⁸ *M³Cat*, http://www.intelec.ca/technologie_f.html.

¹⁹ *Geonetwork*, <http://www.fao.org/geonetwork/srv/fr/main.search>.

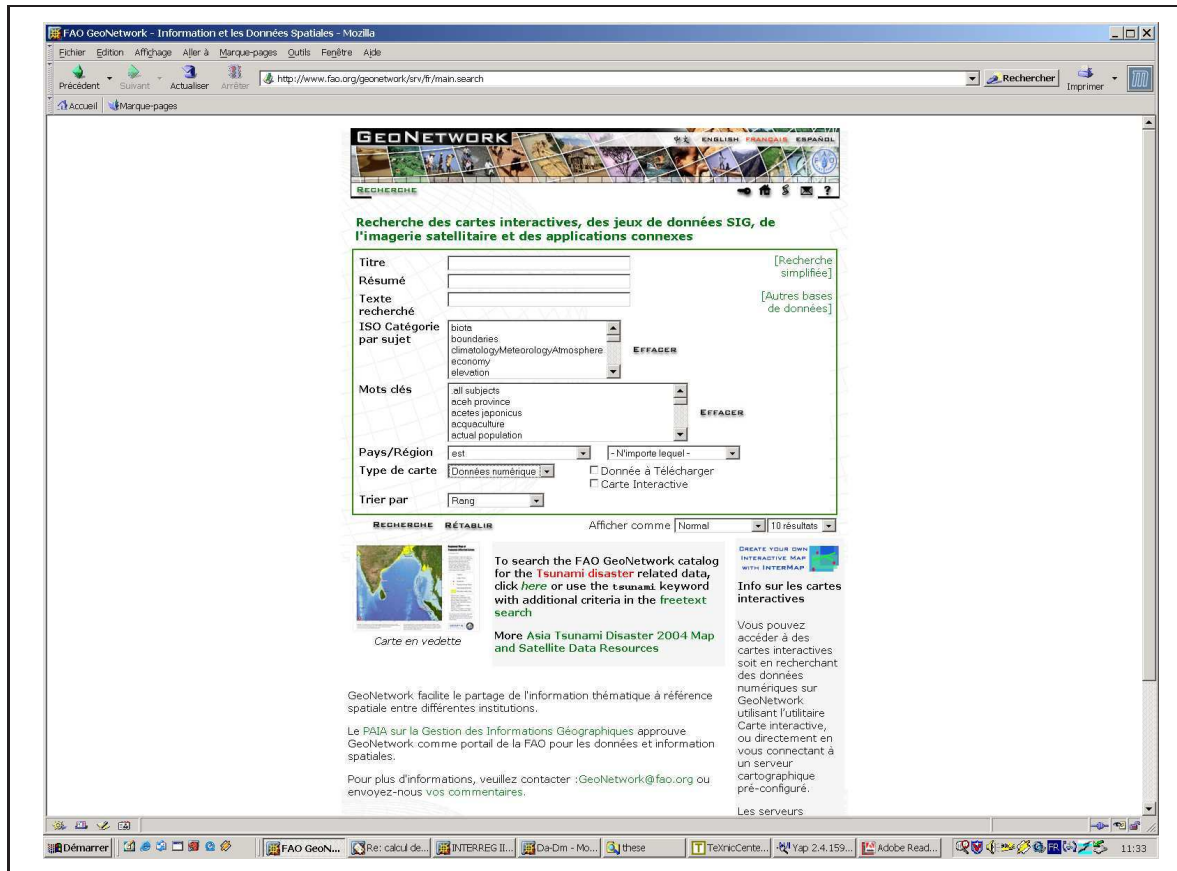


FIG. 4.7 – Interface graphique utilisateur de recherche de métadonnées avec Geonetwork

- Améliorer l'accès aux données spatiales et à l'information et leur utilisation intégrée,
- Aider à la prise de décision en agriculture, foresterie, pêche, et sécurité alimentaire,
- Promouvoir une approche multidisciplinaire du développement durable,
- Renforcer la prise de conscience des avantages de l'information géographique.

Caractéristiques techniques Les *caractéristiques techniques* de *Geonetwork* sont les suivantes :

- *Environnement de base de données* : JDBC ²⁰,
- *Statut de développement* : Stable (version *GeoNetwork* 2.0 en cours de développement, statut *GeoNetwork* 2.0.0 Beta4),
- *Public visé* : utilisateurs sur poste fixe, éducation, administrations publiques, ONG, organismes scientifiques, autres. . . ,
- *Licence* : GNU General Public License (GPL),
- *Système d'exploitation* : indépendant,

²⁰ JDBC (Java DataBase Connectivity) est une API fournie avec Java (depuis sa version 1.1) permettant de se connecter à des bases de données. Pour cela JDBC fournit un ensemble de classes permettant de développer des applications capables de se connecter à des serveurs de bases de données (SGBD). L'API JDBC a été développée de façon à permettre à un programme de se connecter à n'importe quelle base de données en utilisant la même syntaxe (l'API JDBC est donc indépendante du SGBD).

- *Langages de programmation* : Java, XSL (XSLT/XPath/XSL-FO),
- *Thèmes* : partage de fichiers, contenu dynamique, SIG, archivage,
- *Traductions* : chinois (simplifié), anglais, français, espagnol,
- *Interface utilisateur* : application Web ²¹.

Pour améliorer sa portée et sa capacité en tant que portail géospatial orienté utilisateur, *Geonetwork* va bénéficier de **nouvelles fonctionnalités** qui vont permettre de satisfaire les objectifs suivants :

- supporter la définition de différents types de métadonnées, ce qui rendra le catalogue et l'éditeur de métadonnées indépendants du standard de métadonnées. Cela permettra le support natif d'autres standards (FGDC, Dublin Core par exemple) mais aussi la norme d'implémentation de l'ISO 19115 ainsi que l'ISO 19139 qui valide la métadonnée sur un schéma XML.
- étendre les recherches de métadonnées localement et sur d'autres nœuds équivalents pour améliorer les recherches sur des types de métadonnées interopérabilité entre systèmes équivalents à l'aide de la norme *ISO 23950*).
- permettre l'usage de thésaurus pour l'indexation et la recherche par mots clés.
- permettre la synchronisation des archivages existants de par le monde (miroir automatiques de catalogues distants),
- supporter l'extraction de données en utilisant les services *WCS* et *WFS* de l'*OGC*.

Ces développements seront incorporés dans un futur proche dans *Geonetwork 2.0*, développé courant 2005.

Nokis

Contexte et objectifs Nous ne revenons pas sur le contexte et les objectifs de ce programme qui ont été décrits en partie 2.3.2. Concernant l'éditeur de métadonnées, Nokis propose un téléchargement gratuit jusqu'à la version actuelle, les versions suivantes seront payantes ou sous restrictions d'usage. La volonté d'utiliser le support cartographique est confirmée par l'implémentation d'une interface cartographique pour la saisie et la consultation du service de métadonnées (basée sur l'application Mapserver) [112].

Caractéristiques techniques Les *caractéristiques techniques* de *NOKIS* sont basées sur un environnement de développement libre [113] :

- *Environnement de base de données* : MySQL et PostgreSQL accès par JDBC,
- *Statut de développement* : stable,
- *Public visé* : au départ la communauté NOKIS (liée à la GIZC) puis communauté de la géomatique dans son ensemble,
- *Licence* : libre selon le numéro de version,
- *Système d'exploitation* : Windows,,
- *Langages de programmation* : Java, XML-XSL,
- *traductions* : allemand, anglais,
- *Interface utilisateur* : Web-based (JavaServer Pages, pages XHTML et serveur Web Apache avec Tomcat),
- *Interface cartographique* : MapServer (Université du Minnesota),
- *Contrôle du vocabulaire* : GEMET,

²¹ Web-based application en anglais.

- *Autres outils* : *ht://Dig search engine* comme outil de recherche plein texte,

En outre on peut préciser que NOKIS a mis au point des filtres d'import pour ArcCatalog [111] et permet de gérer des systèmes distribués à l'aide la norme *ISO 239.50*.

Ce tour d'horizon donne un aperçu concret du panorama d'outils disponibles actuellement pour décrire l'information géographique à l'aide de la norme *ISO 19115*. Les développements associés à ces outils évoluent rapidement, d'autant plus que la période sur laquelle s'est effectué ce travail de thèse a été associée à un regain d'intérêt pour les métadonnées conjugué à la finalisation de la norme *ISO 19115* et à celle à venir de l'*ISO 19139*.

Compte tenu de nos objectifs (environnement de développement libre et langage d'interfaçage données Web simple) et de la collaboration avec l'équipe *IRD ROSELT*, nous proposons de nous appuyer sur *MDWeb* pour proposer un outil spécifique à notre contexte qui s'appuie sur un nouveau *MCD* et de nouvelles fonctionnalités.

Synthèse 4.2

- La norme *ISO 19115* offre un jeu d'éléments de métadonnées très riche. Ces éléments sont structurés en rubriques thématiques. Ils sont obligatoires ou optionnelles et leur valuation peut être libre ou contrôlée.
- Cette norme est adaptable via la mise en place de *profils*.
- Le *noyau* de l'*ISO 19115* est restreint à une quinzaine d'éléments. La correspondance entre les éléments de l'*ISO 19115* et ceux du *Dublin Core* est facile à mettre en place. Ceci permet, par l'implémentation d'un jeu d'éléments unique, d'être en accord avec les recommandations européennes (*MADAME*, *INSPIRE*).
- Il n'existe pas d'implémentation de référence de la norme *ISO 19115*, celle-ci sera normalisée par l'*ISO 19139* en cours de finalisation.
- Certaines solutions proposées par les outils existants permettent la mise en place de *services de métadonnées libres et en ligne*. En outre, il y a une tendance générale à développer des outils *multi-normes*^a et *multi-lingues*.
- Les systèmes de gestion des données utilisent soit des *SGBDR libres* ou *propriétaires* soit *XML*.
- Nombreux sont ceux qui mettent en oeuvre ou envisagent l'adjonction de *référentiels thématiques et géographiques* par l'emploi respectivement d'*outils de contrôle du vocabulaire* et d'*interfaces cartographiques*.

^aSi on se tient à l'implémentation de la norme *ISO 19115* et du *Dublin Core*, une approche générique multi-norme n'est donc pas nécessaire, cependant, il y a eu tellement de standards différents que cette approche paraît prudente.

Nous allons voir par la suite comment nous avons adapté ces recommandations générales en triant parmi les éléments proposés ceux qui correspondent à notre contexte d'implémentation et, parmi les éléments retenus, comment nous avons reconsidéré certaines de leurs propriétés (obligation, contrôle de la valuation, visibilité) au regard des attentes de nos utilisateurs.

4.3 Proposition d'implémentation de la norme ISO 19115 pour la gestion des ressources de la GIZC

Compte tenu des caractéristiques de la norme *ISO 19115*, nous proposons dans cette section de présenter l'adaptation effectuée pour le projet *Syscolag*. A ces fins, les consultations réalisées auprès de nos partenaires ont permis de distinguer différents types d'éléments de métadonnées d'importance variable pour la GIZC qui conditionnent le *profil* de la norme réalisé et la réponse apportée pour l'implémenter.

4.3.1 Réalisation d'un profil Syscolag sur la base d'entretiens

Nous décrivons dans cette partie la manière dont nous avons procédé pour mettre en place un *profil* par le biais d'*entretiens* auprès des partenaires du projet *Syscolag*.

Entretiens auprès des partenaires du programme *Syscolag*

Les éléments de la norme *ISO* ont été évalués en fonction de notre domaine d'application et du contexte de notre projet avec les partenaires du programme *Syscolag*.

Un *profil spécifique* (ou *profil Syscolag*) résultat d'un tri sur les éléments de métadonnées optionnels proposés par l'*ISO 19115* a été mis en place (sur la base des réponses à des *questionnaires* traduisant en français les définitions des éléments de métadonnées de cette norme et permettant l'évaluation de leur intérêt par les futurs utilisateurs).

Cette enquête auprès des partenaires du projet a contribué à l'objectif d'élaborer un outil de travail commun par la concertation (voir partie 3.1.1). En outre, cette approche plus concrète que des explications en réunion a permis de mieux sensibiliser ce public non spécialisé au concept de métadonnée.

A cette occasion, nous avons également anticipé les déclinaisons possibles de ce *profil* en fonction des types de ressources manipulées dans notre projet (voir notion de *gabarit* en partie 4.3.2).

Bilan des entretiens – attentes des utilisateurs

On peut distinguer deux sortes d'éléments du point de vue des utilisateurs : les éléments *rejetés* et ceux *retenus* dans le *profil Syscolag*, parmi lesquels on distingue différents types.

Nature des éléments rejetés par les utilisateurs La première remarque qui peut être faite à l'encontre de l'*ISO 19115* est la (trop) grande richesse de ses éléments. En effet, c'est une norme *générique*, pensée pour des usages privés ou publics, commerciaux ou non et destinée à des structures de toutes tailles (de l'association à la multinationale).

Or, certains usages sont exclus du domaine de la GIZC en général, et de *Syscolag* en particulier, qui regroupe essentiellement des structures de recherche, à but non lucratif. Il est clair que certains éléments proposés par la norme n'ont alors que peu d'intérêt dans notre contexte. En particulier, l'*ISO 19115* prévoit un ensemble élaboré d'éléments de métadonnées pour décrire la *commercialisation* et la *distribution* des ressources alors que, dans notre cas, cet usage reste anecdotique. Les éléments associés à ces catégories ou sous-catégories ont donc été exclus du *profil Syscolag* puisqu'ils parasitent plus qu'ils ne servent le partage d'information dans notre contexte où ils ne sont pas pertinents.

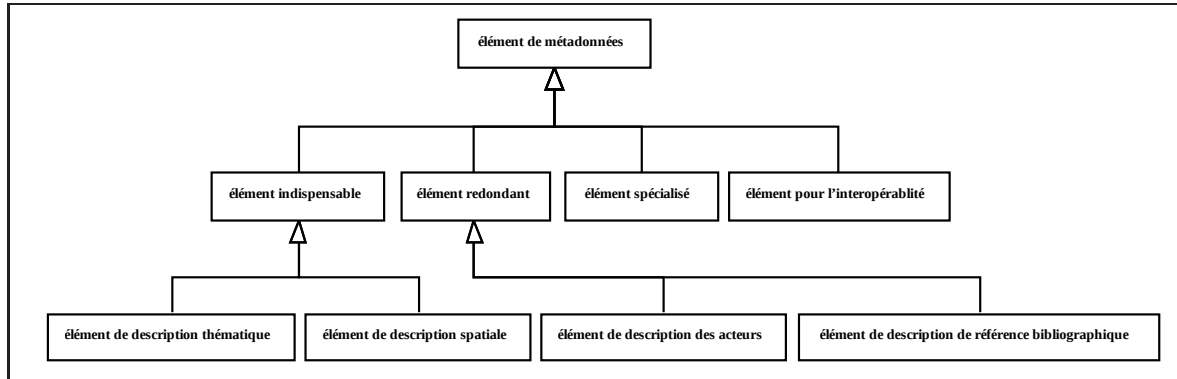


FIG. 4.8 – Les différents types d'éléments de métadonnées vus par les partenaires de Syscolag

De même, aux dires de nos partenaires, il n'est pas nécessaire de demander aux futurs utilisateurs de décrire leurs horaires de présence bureau quand ceux-ci se connaissent assez bien pour décrocher le téléphone. Dernier exemple, l'élément *topics* regroupe des thématiques trop générales qui chapeautent ou composent le domaine de la GIZC et sont donc utiles seulement à l'extérieur de Syscolag.

L'ensemble de ces éléments, non pertinents pour la description des ressources de GIZC dans notre contexte, ont donc été rejetés. Il reste un certain nombre d'éléments sur lesquels on va faire des adaptations pour notre domaine.

Les différents types d'éléments retenus du point de vue des utilisateurs Nous proposons d'adapter les propriétés des éléments de métadonnées que nous avons retenus (modes de *valuation* et *visibilité* proposés par défaut dans l'*ISO 19115*) au contexte spécifique de notre application de GIZC en LR, en nous basant sur la synthèse des remarques des utilisateurs ci-après (voir illustration en Figure 4.8) :

- Les *éléments de métadonnées principaux pour la GIZC*. En complément des *éléments de métadonnées principaux* obligatoires de l'*ISO 19115*, les utilisateurs considèrent que les éléments dédiés aux *descriptions thématique* et *spatiale* (*indexation* à l'aide de **mots-clés** ou de *descripteurs cartographiques*) sont également indispensables à la *mutualisation d'information et de connaissance* dans notre domaine. En effet, ces modes *descriptions* du contenu des *ressources informationnelles* de la GIZC concentrent les problèmes d'*interopérabilité sémantique* entre les *acteurs* ou les *machines*. Nous avons donc choisi de les ajouter au *noyau* de notre *profil* et de contrôler leur *valuation* (voir partie suivante).
- Certains *éléments de métadonnées secondaires*, bien que nécessaires, alourdissent la saisie et sont redondants dans les différentes fiches de métadonnées, voire au sein d'une même fiche. Par exemple, le *Data Type Citation* correspond à une citation du type *référence bibliographique*. Il regroupe 42 éléments et est potentiellement présent à dix reprises dans sept rubriques différentes. On peut également citer le *Data Type ResponsibleParty* (auquel fait appel le *Data Type Citation*) qui contient 22 éléments et concerne la description des *acteurs* (dans des rôles de *distribution*, *production*, *utilisation* de ressources ou de métadonnées). Ce type de données est potentiellement présent à sept reprises dans six rubriques différentes. On imagine sur la base de ces deux exemples le volume de saisie

que ces types de données représentent potentiellement (sachant que leur cardinalité est parfois multiple).

Or, dans le domaine de la GIZC, le réseau d'*acteurs* est restreint, stable dans le temps (pour une période donnée, les *acteurs* possèdent une même *adresse*, même *téléphone*, même *e-mail*...) et gérable par le biais d'un *annuaire électronique* (voir partie 3.1.1). Il paraît alors judicieux d'enregistrer les *éléments de métadonnées* relatifs aux *coordonnées* des *acteurs* une fois pour toutes et d'exploiter le stockage de cette information pour proposer des *valeurs pré-enregistrées* et *éditables*, qui *automatisent* et donc *allègent la saisie* du point de vue de l'utilisateur.

Le cas des *références bibliographiques* est tout à fait similaire. En effet, les *champs* constitutifs d'une *référence bibliographique* (*titre*, *auteur*, *date d'édition*, *mots clés*, *résumé*...) correspondent à des *éléments de métadonnées principaux* gérés par l'ISO 19115. En enregistrant ces valeurs qui leur sont associées dans une base qui regroupe les *références bibliographiques* des partenaires, il est alors possible d'*automatiser la valuation* des *éléments de métadonnées* correspondants lors de la description d'une *ressource informationnelle* qui est déjà décrite par le biais d'une *référence bibliographique*.

- Certaines disciplines impliquées dans la GIZC sont peu familières de l'utilisation d'IG et de SIG. En conséquence, les trois rubriques (et éléments associés) dédiés à l'IG (voir partie 4.2.1) ne présentent pas vraiment d'intérêt de leur point de vue. Pour d'autres, utilisateurs ou producteurs d'IG, l'intérêt de ces rubriques est évident²².

Néanmoins, *tous* portent un intérêt minimum à la spatialisation des ressources et utilisent les descripteurs spatiaux comme les *toponymes* ou *mots-clés spatiaux*, et les éléments qu'ils souhaitent retenir profilent plus l'ISO dans le sens d'une *norme de découverte* que vers une *norme spécialisée*. Le reste pourrait être masqué. L'information environnementale est spatialisée sans être forcément de l'information géographique au sens strict.

- Certains éléments ne répondent pas aux attentes des utilisateurs dans le contexte d'utilisation interne à Syscolag (ni pour la saisie, ni pour la consultation). Ils sont, néanmoins, indispensables car ils permettent l'*interopérabilité* au delà de notre contexte d'implémentation avec d'autres systèmes à l'échelle nationale ou à l'étranger. Ces éléments décrivent, par exemple, la *norme utilisée* (ISO 19115), la *langue utilisée* (français pour notre part), le *profil* employé (*profil Syscolag* dans notre cas), le *jeu de caractères*... Ils constituent des critères importants dans la construction de requêtes de consultation inter-systèmes ; en outre, ils possèdent des valeurs généralement constantes dans un contexte d'application donné.

Nous proposons donc de les conserver dans notre *profil Syscolag* mais en les masquant pour l'utilisateur (en leur attribuant une *visibilité* de type *privé* et une *valuation par défaut*).

Cette approche de *concertation* sous forme de *réunions* et *questionnaires* a donc permis de formuler des remarques et de s'entendre sur la nécessité d'adaptations supplémentaires afin de rendre plus spécifique à notre domaine cette norme générique prévue pour le domaine de l'IG. Nous proposons ainsi de modifier les propriétés associées aux éléments de métadonnées (automatisations, valuations par défaut ou masquées...) pour ne présenter à l'utilisateur que les éléments qui lui semblent pertinents pour répondre aux questions de GIZC qu'il se pose

²²La question de *profils utilisateurs* au sein du même *profil Syscolag* se pose donc pour la *visibilité* ou l'*obligation de remplissage* de tels éléments.

tout en restant, de manière transparente, conforme aux descriptions nécessaires à une meilleure interopérabilité des systèmes.

Par ce biais, nous souhaitons résoudre le manque d'adaptation de la norme *ISO 19115* au public visé et la lourdeur de remplissage qui dans d'autres projets ont été fatals à la réussite de l'outil mis en place.

Nous revenons à présent sur les deux éléments de métadonnées principaux (d'après les utilisateurs interrogés) pour la description des ressources impliquées dans la GIZC.

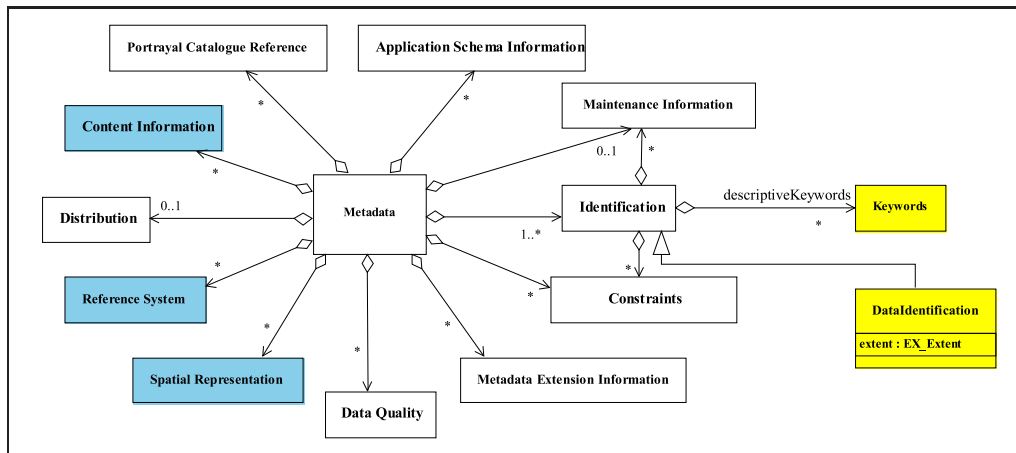


FIG. 4.9 – Importance particulière des descripteurs relatifs aux mots clés et à l'étendue géographique

Eléments de métadonnées incontournables pour la GIZC Nous revenons dans cette partie sur l'importance des *descriptions thématique* et *spatiale* dans le contexte pluridisciplinaire de la GIZC, que les utilisateurs ont bien identifiée.

En effet, dans notre domaine [68], il ne faut pas perdre de vue qu'un des intérêts majeurs de l'*ISO 19115* est de permettre en même temps que la *description thématique* d'une ressource, la *description de son étendue géographique* (y compris pour une ressource quelconque, information géographique au sens strict ou non). Ceci permet, par la suite, une recherche de métadonnées qui conjugue critères thématiques et / ou spatiaux sur lesquels l'essentiel de la connaissance contenue dans la requête est concentrée.

En effet, pensée pour décrire l'*IG*, cette norme intègre aux descriptions proposées la composante spatiale nécessaire dans la *GIZC* pour utiliser le *support cartographique* comme un moyen privilégié pour échanger l'information (voir partie 3.1.3). Ce sont les éléments du *data type "EX_Extent"* qui rendent possible, indépendamment de la nature de la ressource, une recherche par critère spatial (voir Figure 4.9).

Pour respecter l'importance des critères spatiaux et thématiques dans la *GIZC*, quel que ce soit le type de la ressource décrite, nous avons choisi d'imposer la description de ces éléments de métadonnées. Ils s'ajoutent donc à la courte liste d'éléments strictement obligatoires pour notre *profil*.

La figure 4.9 illustre leur position dans la norme *ISO 19115*. Nous verrons par la suite, dans les chapitres 5 et 6, comment nous prêtons une attention particulière à leur valuation.

Une fois que les éléments de métadonnées pertinents dans le contexte général de la *GIZC*

ont été connus, nous avons remodelé notre *profil* selon le type des différentes ressources inventoriées.

4.3.2 Réalisation des gabarits associés au profil Syscolag

Le *profil Syscolag*, obtenu à la suite d'entretiens avec les partenaires du programme, constitue un premier niveau de tri qui avait pour but d'estimer l'utilité des éléments optionnels dans l'absolu. Si l'élaboration d'un *profil* est une étape nécessaire à l'adaptation d'une *norme spécialisée* telle que l'*ISO 19115*, elle n'en est pas moins insuffisante pour répondre à l'ensemble des cas d'utilisation d'un *service de métadonnées* pour la *GIZC*. En particulier, selon la nature des ressources informationnelles décrites il est possible de s'appuyer sur la mise en place de *gabarits*.

Définition et intérêt des gabarits

Selon le *type de ressource* à décrire, les éléments de métadonnées retenus dans le *profil* ne sont pas toujours pertinents ou suffisants. Il faut donc enlever ou ajouter des éléments au *profil* selon la nature des ressources décrites. Ceci s'effectue par la mise en place de *gabarits*. On peut donc considérer que *les gabarits sont des déclinaisons du profil Syscolag (un profil du profil Syscolag) selon la typologie des ressources concernées par le service de métadonnées* [13].

Lors de l'élaboration des *gabarits* pour *Syscolag*, nous éliminons ou ajoutons des éléments dans le *profil Syscolag* en fonction de leur pertinence pour la description de chaque type de ressource d'intérêt dans notre contexte.

A ces fins, compte tenu de la diversité des *acteurs* impliqués dans ce domaine et de la diversité des *ressources* manipulées qui en découle (voir partie 2.1.3), une analyse préliminaire de la *typologie des ressources informationnelles* et des modalités de description associées a été nécessaire [14].

Typologie de ressources et gabarits associés dans Syscolag

La Figure 4.10 présente une typologie des *ressources informationnelles* d'intérêt potentiel pour *Syscolag* et plus généralement pour la *GIZC*. Il est donc intéressant d'associer un *gabarit spécifique* à ces différents types de ressources.

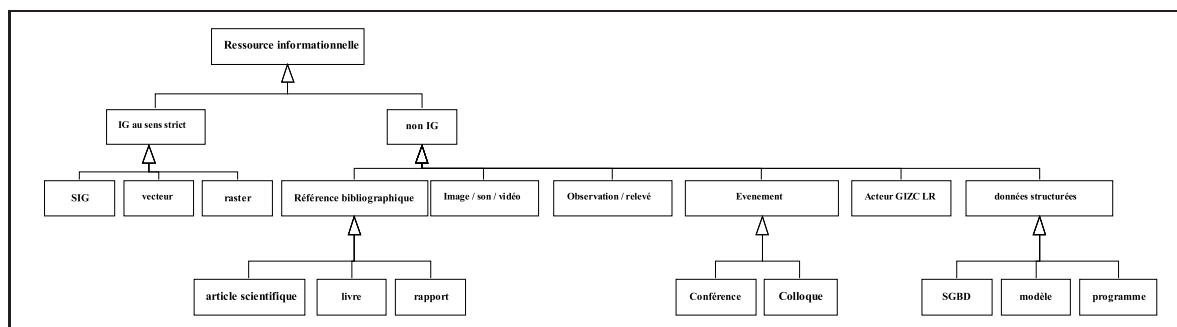


FIG. 4.10 – Proposition de typologie de ressources informationnelles pouvant faire l'objet d'un gabarit

En particulier, il ressort, pour notre communauté d'utilisateurs, qu'on peut distinguer les ressources informationnelles du type d'*information géographique au sens strict* qui respectent le cadre strict de cette norme, de celles d'information non géographique qui nécessitent une adaptation plus poussée de l'*ISO*, en particulier en éliminant les éléments trop ciblés *IG* (soit trois rubriques qui disparaissent, voir partie 4.2.1).

Quoiqu'il en soit on ne peut pas considérer actuellement que l'ensemble des ressources à renseigner ait été complètement cerné. De nouveaux types de ressources devront alors faire l'objet de nouveaux *gabarits*. Par exemple, il n'a pas été décidé de renseigner de métadonnées pour des événements de types *réunion*, *colloque*, *séminaire*...et de leur attribuer un *gabarit* particulier. De même il n'a pas été explicitement décidé de considérer comme un *type de ressource* un *acteur* ou une *connaissance*, même si les informations correspondantes existeront dans la base (voir notre définition de ressources en partie 2.1.3 et définition de *Tim Berners Lee* en partie 4.1).

Pour cette raison, la mise en place d'un *outil d'édition de gabarits* facilitera la création de gabarits pour de nouveaux types de ressources (voir partie 7.2.2).

Nous venons d'expliquer, dans les parties 4.3.1 et 4.3.2, en quoi les propositions de l'*ISO* sont adaptables à notre contexte de GIZC à l'aide d'un *profil* et de *gabarit*. La démarche de saisie restera néanmoins assez lourde compte tenu de la quantité moyenne d'éléments à renseigner par *gabarit*. Nous verrons dans le chapitre 7.2 comment nous facilitons et optimisons la saisie de métadonnées par le biais d'outils qui assistent l'utilisateur et l'administrateur lors des tâches de saisie les plus délicates.

A présent nous allons voir comment le *modèle conceptuel* proposé pour l'implémentation d'un *service de métadonnées* dans *Syscolag* tient compte de l'ensemble de ces remarques.

4.3.3 Modèle conceptuel proposé

Une fois que le *profil* et les *gabarits* ont été élaborés, ils doivent être implémentés au sein d'un système qui tiendra compte des remarques des utilisateurs et visera à réduire la lourdeur du *service de métadonnées* proposé en facilitant la *saisie* aussi bien que la *recherche* des métadonnées.

Pour cela, en l'absence actuellement d'implémentation normalisée de l'*ISO 19115*, différents *modèles conceptuels* de métadonnées et différentes implémentations techniques sont possibles (voir section 4.2.3). Au regard des solutions existantes, nous présentons dans cette section notre proposition pour notre contexte d'implémentation.

Modèle conceptuel pour un service de métadonnées

Nous proposons à présent une synthèse sur les *concepts* que nous avons définis dans ce chapitre afin d'explicitier les *relations* qu'ils entretiennent. Pour cela, le diagramme *UML* de la Figure 4.11) *structure* les classes *norme*, *profil* (qui pourrait aussi bien être remplacé par *gabarit*), *fiche de métadonnée élément de métadonnée* et *valeur* selon les définitions qui en ont été données dans les sections précédentes.

Nous nous appuyons sur cette Figure pour présenter les différentes approches envisageables afin de justifier et faire comprendre la pertinence des choix que nous préconisons pour la réalisation d'un *service de métadonnées* adapté à notre domaine.

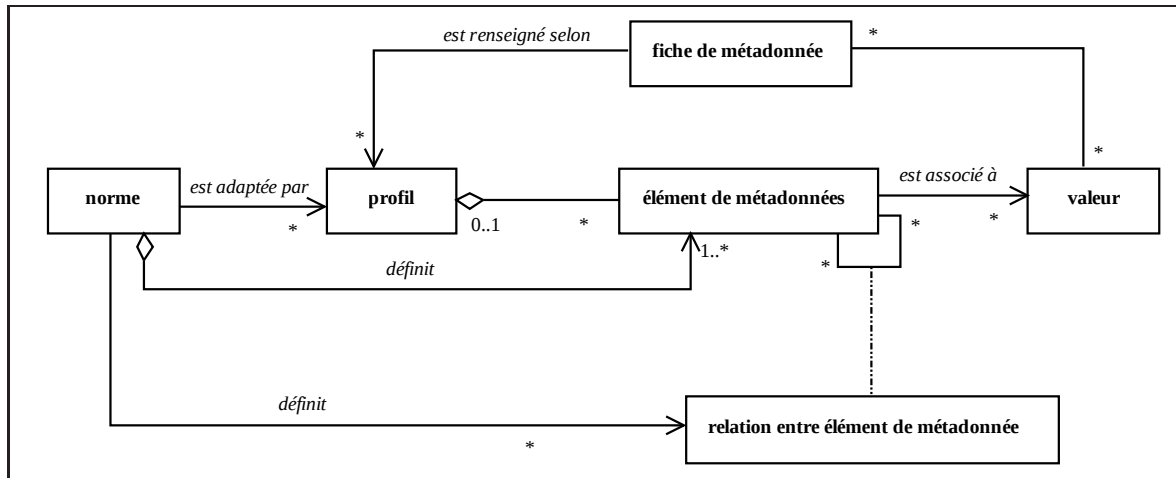


FIG. 4.11 – Approche générique d'une norme de métadonnées

Les différentes approches possibles

Parmi les approches possibles pour implémenter les recommandations de la norme *ISO 19115*, nous en distinguons ici deux qui sont représentatives des applications actuellement disponibles. La première est basée sur l'utilisation de *SGBD*, la seconde sur l'utilisation de *XML*.

Gestion des métadonnées à l'aide de *SGBD(R)* (*SQL*) Elle peut être réalisée en s'appuyant sur deux types de *modèle conceptuel* :

- une *approche générique multinorme*, basée sur l'interprétation de la Figure 4.11, qui présente l'avantage d'être robuste (indépendante des normes, de leur contenu et de leurs évolutions).

Cette approche est dite *générique* car la structure de toute *fiche de métadonnée* est *calculée* et *valuée* à partir des *extensions* des tables associées aux classes "*valeur*" et "*relation entre éléments de métadonnées*".

Les tables "*valeurs*" et "*relation entre éléments de métadonnées*" ont des *attributs génériques* comme "*nom_norme*", "*nom_element*", "*valeur*"... "*type_relation*" et ce sont uniquement les *tuples* de ces deux tables qui retranscrivent l'information sur la nature des *éléments de métadonnées* (la *norme* qui les définit, leur type...), le *profil* et les *valeurs* associées ainsi que la manière de les *structurer* (la nature des *relations* étant déduite des *éléments* et de la *norme de métadonnées* concernés).

Cette approche est idéale pour des contextes d'implémentation *multi-culturel* (*multi-lingues*, multi-domaines...). Cependant, dans ce cas, les valeurs de l'ensemble des *fiches de métadonnées* sont gérées dans une table "*valeur*" unique, sans distinction sur la nature des éléments auxquels elles sont associées et donc sans gestion des redondances d'information. Par exemple, si l'*étang de Thau* est utilisé pour la description spatiale de *n* fiches de métadonnées, il sera alors stocké *n* fois dans la table "*valeur*" (dans des *tuples* stockant également *n* fois le nom des descripteurs spatiaux de l'*ISO 19115*). Ceci pose donc un problème d'efficacité pour l'exécution des requêtes lorsque le volume de la table devient conséquent. Or, ce problème est d'autant plus important que le nombre de

fiches de métadonnées est élevé (ce qui, au moins à long terme, doit être pris en compte). Le *modèle conceptuel* sur lequel s'appuie la réalisation du logiciel *MDWeb* utilise une approche de ce type [44].

- une *approche matérialisée* qui associe directement aux classes *UML* de la norme *ISO 19115* un ensemble de tables (que nous nommerons *tables en dur*). Toute *fiche de métadonnée* est alors structurée et évaluée d'après les relations et les valeurs contenues dans ces différentes *tables*. En effet, par opposition à l'approche précédente, les tables d'un *MPD* issues d'une telle approche traduisent donc, par leur structure, la nature de la norme implémentée (*ie* ce sont les *colonnes* –et non plus les *tuples*– qui portent directement le nom des *éléments de métadonnées* et retranscrivent leurs relations). Le modèle physique est donc fonction de la norme implémentée.

Cette approche présente moins de souplesse que la précédente pour d'éventuelles évolutions des *gabarits*, du *profil*, voire de la *norme*, et nécessite une forte stabilité du contexte d'implémentation : nature des ressources décrites, thématiques et acteurs impliqués. ... Par contre, elle permet de gérer la *redondance d'information* et d'optimiser l'exécution des *requêtes* en distinguant la nature des différents *éléments de métadonnées* qui peuvent être gérés dans différentes tables. Pour répondre à l'exemple donné dans l'approche générique, l'*étang de Thau* serait, dans ce cas, stocké dans un unique *tuple*, associé à différentes *fiches de métadonnées* structurées par différentes normes (dans le respect des règles d'*intégrité référentielle*).

Gestion des métadonnées à l'aide de fichiers *XML* Du point de vue de l'implémentation, *XML* constitue l'alternative au *SQL*. Ses avantages sont la facilité de publication des documents *XML* sur le Web *via* l'utilisation de fichiers *XSL*, l'adaptation à la traduction des modèles objets de la norme *ISO 19115* et l'adéquation avec le pendant *XML* de la norme *ISO 19115* : la norme *ISO 19139*²³. Par contre, en terme de stockage, de recherche, de multilinguisme et d'évolution des *profils* ou *gabarits* élaborés les limites de ce type d'implémentation par rapport au *SQL* sont plus fortes et certains *éditeurs de métadonnées* mélangent une approche *SQL* et *XML* (voir partie 4.2.3). En effet, dans ce cas, l'information est stockée à plat dans des fichiers *XML*, structurés par une *DTD* qui doit être parcourue ("*parsée*") entièrement pour résoudre une requête.

Nous présentons dans la section suivante quelques outils permettant la mise en place de service de métadonnées et répondant aux objectifs d'implémentation de notre contexte.

Le modèle mixte proposé

Le constat est le suivant : l'*approche générique* en *SQL* telle que celle proposée par *MDWeb* (basée sur le type de modèle présenté en Figure 4.11) ne permet pas, sans adaptations (voir Figure 4.12), de distinguer les différents types d'éléments de métadonnées proposés par l'*ISO 19115* conformément aux remarques que nous avons recueillies auprès des futurs utilisateurs du *service de métadonnées* de *Syscolag* (voir Figure 4.8 en partie 4.3.1). Or, dans notre contexte, certains de ces éléments sont potentiellement redondants au sein d'une même fiche et d'une fiche de métadonnées à l'autre, en particulier ceux que nous appelons *élément indispensable* et *élément redondant* (voir Figure 4.8). En effet, les valeurs possibles pour décrire la thématique, l'étendue géographique ou les acteurs associés à une ressource informationnelle

²³ *XML* est le format d'échange pour les systèmes qui implémentent cette norme.

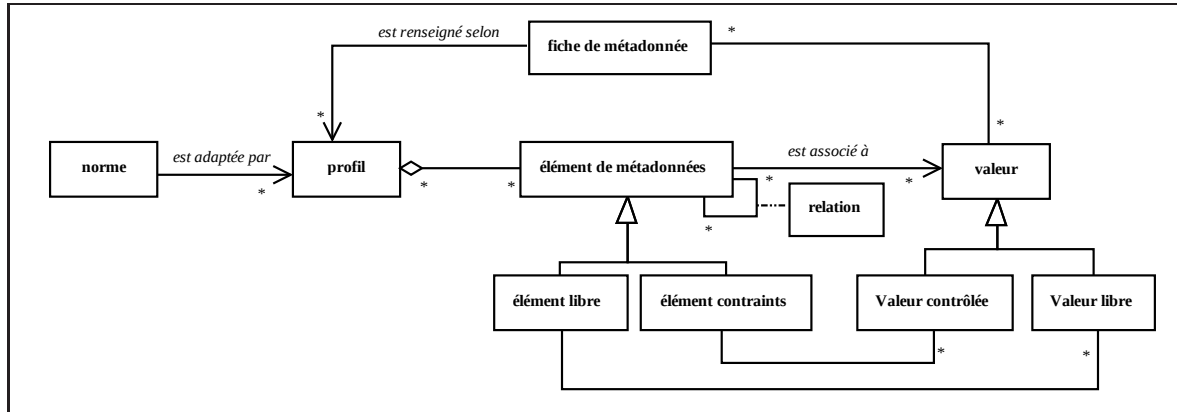


FIG. 4.12 – Adaptation pour Syscolag du modèle générique multi-norme en modèle mixte qui distingue le mode de valuation selon la nature des éléments de métadonnées

de notre domaine, appartiennent à un univers limité par le contexte régional et la thématique d'application. Ceci constitue une limite de l'*approche générique* pour la gestion des *redondances d'information* liée à la valuation de certains types d'éléments (voir partie 4.3.1) et donc une limite pour la *résolution des requêtes* mais aussi pour d'éventuelles *automatisations* dans notre contexte. Par contre, à l'aide de *tables en dur*, la redondance est limitée et les automatisations sont possibles à l'aide de *tables de jointures* (pour la gestion des *relations n-n*) mais cette approche trop figée ne supporte pas la gestion multi-normes...

Nous proposons donc une adaptation du *modèle conceptuel de données (MCD) générique* présenté en Figure 4.11 qui permet de distinguer différents modes de valuation selon la nature des *éléments de métadonnées* à gérer (voir Figure 4.12). A ce modèle est associé un *modèle physique de données (MPD)* constitué de tables de différentes natures qui conjuguent *approche générique* et *approche matérialisée* :

- une *approche générique* qui utilise le mécanisme, décrit dans la partie précédente, pour calculer la *structure* de la *fiche de métadonnées* et la *valuation* des *éléments* dits *libres* (*ie* dont on ne peut pas anticiper la valeur : *valuation libre*). L'approche générique développée par *ROSELT* (sur laquelle nous nous appuyons) ou *M³Cat* est donc conservée pour ces éléments. Cela se justifie par le fait que ce ne sont pas des *éléments clés* de description ou de recherche et que leurs valuations ne sont a priori pas redondantes et pas automatisables. De plus, ces éléments répondent chronologiquement à la dernière question posée lors d'une recherche de métadonnées : comment se procurer la ressource décrite ? Ce sont des *éléments de métadonnées secondaires ou pour fins de diffusion* (voir partie 4.1.4). Or à cette étape la ressource est déjà identifiée par le système qui peut alors reconstituer simplement la fiche détaillée.
- une *approche matérialisée* appliquée aux classes *UML* de la norme *ISO 19115* relatives aux *éléments* dits *contraints* permettant de compléter la *valuation* de la *fiche de métadonnée* par des *requêtes* portant sur les *tables en dur* correspondant à ces classes. A l'aide de ces tables, nous permettons donc la *valuation contrôlée* des *descriptions thématiques* et *spatiales* (*éléments incontournables* voir partie 4.3.1) et l'*automatisation* des *éléments redondants* et évitons ainsi la redondance d'information (voir Figure 4.8, partie 4.3.1).

En particulier, comme nous l'avons évoqué, ces *éléments de métadonnées principaux* qui répondent aux questions, *qui ?*, *quoi ?*, *où ?* et *quand ?* sont prépondérants dans la construction de requêtes lors de la *consultation* de *métadonnées*. Ils sont de plus, dans notre contexte, potentiellement redondants dans une même fiche ou dans différentes fiches de métadonnées.

A ce titre, nous avons intérêt à les gérer dans des *tables en dur* pour l'efficacité des requêtes en plus des facilités de saisie.

Les bénéfices attendus d'une telle approche sont les suivants :

- La *saisie de éléments de métadonnées clés* est améliorée par de possibles *automatisations* grâce à l'information stockée dans les *tables en dur* (qui enlève les redondances dans une table *unique générique*, l'optimisation de la saisie se fait concrètement en indiquant un nouvel enregistrement dans la jointure plutôt qu'une nouvelle saisie). Le *contrôle de la terminologie* employée pour la *valuation des éléments clés* participe à la *qualité de l'indexation* lors de la saisie ainsi qu'à un *contrôle sémantique* bénéfique pour la *qualité des recherches d'information*. En outre, cela permet d'envisager la mise en place de *profils utilisateurs*.
- La *recherche de métadonnées*, est améliorée d'une part, par la possibilité de résoudre les requêtes primordiales dans des *tables en dur* dédiées à la gestion des *éléments de métadonnées principaux* (en effet, il n'est plus nécessaire de parcourir une table unique et volumineuse qui mélange *éléments principaux clés et secondaires* pour résoudre les requêtes mais plutôt des *tables de jointures* restreintes), et d'autre part, par la gestion de la redondance d'information qui *améliore les temps de réponse du système*²⁴.
- La gestion multi-normes n'est pas compromise puisque les éléments gérés dans des tables matérialisées sont des *éléments de métadonnées principaux* qu'on retrouve dans tous les standards (en particulier le *Dublin Core*). Il est donc possible d'utiliser des *correspondances*.

Nous apportons, par de possibles automatisations, une réponse à la *lourdeur de la saisie qui est un verrou majeur pour l'utilisation d'un service de métadonnées*. Mais surtout, dans notre contexte, nous pouvons nous appuyer sur un *contrôle de la valuation d'éléments clés* en lien avec la nécessité pour les *acteurs* la GIZC de *partager des descriptions thématiques et spatiales*. Cette approche assure finalement, côté *système*, une meilleure *gestion de l'information* tant sur le plan *quantitatif* que sur le plan *qualitatif*.

4.3.4 Réalisation de la base de métadonnées

Afin de développer un *prototype* pour mettre en œuvre cette *approche mixte*, nous avons bénéficié d'une collaboration avec l'équipe *ROSELT* qui développe le logiciel *MDWeb* (voir partie 4.2.3). Sur la base de ce logiciel nous avons donc collaboré pour effectuer des adaptations complémentaires à l'approche *multi-norme* et *multi-lingue* employée à l'origine afin de permettre les *contrôles* et *automatisations* d'éléments de métadonnées essentiels dans le domaine de la GIZC.

²⁴Le problème de l'optimisation du stockage pour ménager l'espace disque par une éradication des redondances est plus ou moins obsolète compte tenu des capacités actuelles de stockage, cependant le problème de l'efficacité d'exécution des requêtes sera toujours d'actualité et il reste en partie lié à des modèles performants du point de vue de la conception.

Le SGBDR utilisé

Nous avons, par commodité, souhaité gérer l'ensemble des éléments de métadonnées (thématiques, spatiaux...) au sein d'un même *SGBDR* (voir Figure 4.12). Au regard de ce que nous avons exposé, le *SGBDR* mis en place doit donc permettre la gestion :

- d'un *annuaire électronique*,
- d'un *outil de contrôle du vocabulaire*,
- d'un *outil de contrôle des descriptions géographiques en mode cartographique*,
- d'un *outil de saisie des références bibliographiques*.

Le mélange dans un même *jeu d'éléments de métadonnées* de *descripteurs textuels* et *cartographiques*, normalisés par l'*ISO* et l'*OGC*, oriente préférentiellement le choix d'un *SGBDR* vers ceux qui disposent d'une *cartouche spatiale*. Or, dans un *environnement de développement libre*, le *SGBDR* de référence est actuellement *Postgres* et sa *cartouche spatiale Postgis*. En outre, le choix de ce *SGBDR* qui implémente les recommandations de l'*OGC* est opportun compte tenu de la collaboration étroite entre l'*ISO* et l'*OGC*, qui laisse à terme espérer un bénéfice pour l'utilisation des outils libres développés pour exploiter le contenu des bases réalisées avec *Postgis* (*SIG clients libres, interfaces cartographiques...*) (voir détail en partie 6.2.3).

Ainsi, les *référentiels thématique* et *spatial* sont gérés au sein du même système que les métadonnées. Cela permet une simplification des requêtes tant du point de vue de la *saisie* et de la *recherche* de *métadonnées* que du point de vue de la génération de fichiers *XML*.

Du point de vue de l'administration du système, nous pensons que l'utilisation d'un système unique (basé ici sur *SQL*, l'alternative du choix de *XML* étant tout à fait justifiable, voir détail complémentaire en partie 8.2.1) présente des avantages en terme de lisibilité et de prise en main de l'application.

Par contre, cela impose, dans notre cas, le choix de *SQL* pour exploiter le modèle du *référentiel terminologique* que nous souhaitons mettre en place, et nous verrons dans le chapitre suivant que beaucoup d'applications de contrôle sémantique sont basées sur l'utilisation de langages dérivés du *XML*.

Le modèle physique de la base de métadonnées

Le *modèle physique* de la *base de métadonnées* conjugue donc deux types de tables, des *tables en dur* pour la gestion des éléments essentiels à la GIZC et des *tables génériques* pour la gestion d'autres éléments pour lesquels il n'est pas envisageable de gérer les redondances des valuations. Le *MPD* définitif étant fonction des modes de contrôle et de gestion des *descriptions spatiales* et *thématiques* qui vont être décrits par la suite, nous présentons ici, à titre illustratif, la Figure 4.13 qui montre un extrait de *MPD* possible focalisé sur les *tables en dur* utilisées pour la gestion des *descriptions spatiales* (*Table OG*), par exemple dans la cartouche *Postgis* de *Postgres*, des *descriptions thématiques* (*Table mot clé*), et des *contacts* (*Table Acteur et Organisme*). Une table dédiée à la gestion des *descripteurs des références bibliographiques* aurait également pu figurer sur ce modèle.

Les chapitres 5 et 6 vont présenter la nature des *modèles conceptuels* associés aux *référentiels* qui permettent la gestion des *descriptions thématiques* et *spatiales* (afin de valuer les *éléments de métadonnées incontournables pour la GIZC*). Ils permettront de mieux comprendre la nature de la réalisation proposée, actuellement en cours de finalisation.

- Le deuxième niveau d'adaptation concerne la déclinaison du *profil Syscolag* en *gabarits* correspondant aux types de ressources susceptibles d'être décrits par des métadonnées.
- Parmi les solutions proposées par les outils existants, compte tenu de l'absence d'implémentation de référence de la norme *ISO 19115* et des contraintes de notre contexte d'implémentation, nous avons opté pour l'adaptation d'un outil existant.
- Une approche mixte alliant généricité du *modèle conceptuel de données* proposé à des tables en dur pour des types d'*éléments de métadonnées clés* dans notre domaine permet de gérer différemment la valuation des différents types d'éléments de métadonnées retenus pour la description des ressources informationnelles de la GIZC.
- La tâche de saisie est ainsi allégée tout en améliorant la qualité d'indexation et par conséquent la qualité des recherches de métadonnées (*ie* la localisation de l'information).

Conclusion du chapitre 4

Compte tenu des problèmes d'interopérabilité des bases de données environnementales en général et des caractéristiques du domaine de la GIZC, les métadonnées apportent une réponse pertinente pour le partage des ressources informationnelles existantes. En effet, elles permettent de localiser les informations existantes, sans changer les structures de gestion des données dispersées chez les différents acteurs. En effet, la gestion séparée de fiches de métadonnées dispense du partage et du traitement physique des ressources pour permettre la localisation de l'information et résout les problèmes de confidentialité liés aux ressources décrites.

Le caractère spatial marqué de l'information et le bénéfice lié à l'usage d'un jeu d'éléments de métadonnées normalisé et spécialisé, nous ont conduits à choisir la norme ISO 19115 dédiée à la description de métadonnées pour l'information géographique. En outre cette norme répond aux recommandations européennes en matière de normalisation de la description des ressources existantes. Cette norme propose un jeu d'éléments de métadonnées très riche et modulable par le biais d'adaptations qui permettent d'envisager son usage pour la GIZC. Il existe des outils d'édition de métadonnées qui implémentent les recommandations de cette norme, cependant l'*ISO 19139* qui normalise l'implémentation officielle est en cours de finalisation.

L'adaptation générale de la norme *ISO 19115* à notre domaine passe par la nécessaire mise en place d'un *profil*. A partir de là, nous avons élaboré un *profil Syscolag* que nous avons décliné en autant de *gabarits* correspondants que de types différents de ressources à décrire dans le *service de métadonnées*. Cette étape a été réalisée en collaboration avec les partenaires du projet *Syscolag* pour répondre au souci d'*élaboration concertée* des outils proposés et à la nécessaire adaptation de cette *norme générique* au contexte particulier de la GIZC en Languedoc Roussillon. Ceci n'est cependant pas suffisant, des adaptations complémentaires sont nécessaires à la bonne marche du service de métadonnées proposé afin d'optimiser le partage des ressources informationnelles de la GIZC.

En particulier, l'adaptation à notre domaine passe par la description obligatoire (en plus des *éléments de métadonnées principaux* traditionnels) de la *portée thématique* et *géographique*

des ressources. Ces deux critères sont, en effet, essentiels à l'échange de connaissance en GIZC et doivent donc apparaître en priorité dans les descriptions des ressources du domaine. L'*ISO 19115* suggère l'utilisation de terminologies de référence de même que l'utilisation de *SGBDR* à *cartouches spatiales* (en lien avec ses travaux avec l'*OGC*) pour gérer ces deux types d'éléments de métadonnées particuliers. Cependant, les valeurs des descripteurs de référence (concepts et objets géographiques) sont à définir par les utilisateurs du domaine concerné.

Dans les deux chapitres suivants, nous rendons compte des différents types d'*outils de contrôle du vocabulaire* et de ceux existants dans le domaine de la GIZC puis nous proposons une solution adaptée à une *valuation contrôlée* des éléments clés de descriptions thématique et spatiale.

Synthèse du chapitre 4

- Les métadonnées apportent une réponse pertinente aux problèmes d'interopérabilité des bases de données environnementales compte tenu de la variété des ressources informationnelles, de leur dispersion et de leurs contraintes d'accès.
- En particulier, elles permettent de partager la localisation des ressources informationnelles existantes et la gestion de leur diffusion, condition qui sera nécessaire à un éventuel partage physique des ressources à plus long terme.
- L'utilisation d'un jeu d'éléments normalisés permet l'interopérabilité syntaxique du service de métadonnées proposé avec d'autres systèmes similaires. Le choix d'une norme de métadonnées adaptée à la nature des ressources qui vont être décrites (en particulier la description spatiale des ressources) est nécessaire dans le cas de la GIZC.
- La norme *ISO 19115* dédiée à la description d'information géographique est une norme riche et adaptable qui respecte les exigences de notre domaine et répond aux recommandations européennes (*MADAME* et *INSPIRE*).
- Il existe différents outils d'édition de métadonnées qui permettent d'implémenter la norme *ISO 19115*.
- Nous adaptons la norme *ISO 19115* par la mise en place d'un *profil Syscolag*, en vue d'une implémentation spécifique à la *GIZC*, puis par la mise en place de *gabarits* adaptés aux différents types de ressources à décrire.
- Parmi les éléments de métadonnées essentiels à la description des ressources de la *GIZC*, les éléments qui permettent les descriptions thématique et spatiale sont fondamentaux. Ils doivent être contrôlés lors de la saisie ou de la consultation du service de métadonnées pour optimiser la localisation de l'information.
- D'autres éléments de métadonnées de moindre importance (contacts des acteurs ou références bibliographiques) nécessitent des automatisations pour faciliter la saisie ou la consultation de métadonnées,
- Nous proposons de nous appuyer sur une approche mixte qui adapte les modes de gestion des éléments de métadonnées en fonction de leur importance pour la saisie ou la consultation.
- Pour cela nous devons donc nous appuyer sur des outils de contrôle des descriptions thématiques et spatiales.

Chapitre 5

Partage de la connaissance en GIZC

Nous allons à présent nous intéresser à l'optimisation de l'utilisation du *profil Syscolag* élaboré à partir des recommandations de la norme *ISO 19115*. En particulier nous allons proposer de gérer au mieux *les éléments clés pour l'indexation des ressources de la GIZC* : ceux qui concernent les descriptions *thématique* et *spatiale*. En effet, en tant que *norme générique pour décrire l'information géographique*, l'*ISO 19115* ne propose pas, pour les éléments thématiques et spatiaux, de *contrôles des valuations* adaptés au cas spécifique de la GIZC. Elle suggère tout au plus, de manière optionnelle, la mise en place de *vocabulaires (thésaurus de référence)* ou *méthodologies contrôlées* pour décrire les portées thématique et spatio-temporelle de la ressource (le mode de contrôle restant à l'appréciation des utilisateurs selon le contexte d'implémentation).

Nous montrons dans un premier temps en quoi cette problématique du *contrôle de la sémantique de ces éléments* est centrale pour le partage des connaissances dans la GIZC et qu'elle rejoint celle émergente du *Web sémantique* liée à l'évolution des *NTIC*. Nous proposons donc, en complément de l'*ISO 19115*, de nous appuyer sur la perception des *mots clés* par les utilisateurs de notre domaine pour la mise en place d'un *contrôle adapté de la valuation des éléments de métadonnées relatifs aux descriptions thématique et spatiale*.

De la même manière que la mise en œuvre d'un *jeu de métadonnées* peut bénéficier de l'utilisation de standards, il existe, pour permettre le contrôle de l'indexation des ressources de notre service de métadonnées, des *terminologies standardisées relatives aux concepts thématiques ou géographiques de référence dans des domaines connexes de la GIZC*. Cependant, il n'existe pas d'outil de référence pour le contrôle du vocabulaire dédié spécifiquement à ce domaine. Un tel outil permettrait de faciliter le travail de *saisie* tout en permettant la nécessaire maîtrise sémantique des concepts employés pour valuer les *mots clés*. En conséquence, la qualité des résultats de *consultation* et la *gestion des connaissances* portées par ces concepts seraient améliorées, ces derniers étant interprétables par la machine.

Nous proposons donc de mettre en place un tel *outil de contrôle du vocabulaire* qui permet, autant que possible, d'intégrer et de restituer la connaissance experte associée aux concepts essentiels de notre domaine à l'aide d'un *formalisme* approprié. Nous décrivons les différentes sortes d'outils de contrôle du vocabulaire (*glossaire, thésaurus, ontologie*) et discutons de leur adéquation par rapport aux objectifs visés. Nous nous appuyons sur des exemples d'outils de ce type existant dans des domaines proches.

Compte tenu de ces présentations et de la complémentarité idéale entre *ontologies* et *métadonnées*, nous proposons la *mise en place d'une ontologie spécifique à la GIZC en Languedoc*

Roussillon qui, comme nous le souhaitons, apporte une réponse au *contrôle de la terminologie* ainsi qu'à la *gestion des connaissances* des partenaires de notre projet.

Comme pour le *service de métadonnées*, la mise en place de ce *référentiel sémantique*, a fait l'objet d'un *travail collaboratif* sur la base d'*entretiens* ou de *réunions*. Les choix techniques ont tenu compte d'un *formalisme adapté* au travail collaboratif souhaité et aux choix informatiques annoncés dans les chapitres précédents pour l'implémentation de notre *service de métadonnées*.

En nous appuyant sur les *sources de référence du domaine* et en mobilisant l'*approche écosystémique*, nous proposons un *méta-modèle d'ontologie* pour l'*intégration de la connaissance liée à la GIZC*. Ceci permet de réutiliser les briques existantes dans des domaine connexes et sur lequel peut s'appuyer le développement de notre outil de contrôle de la terminologie.

Nous exposons finalement comment les *connaissances* intégrées dans le référentiel sémantique sert aussi bien à optimiser la *gestion de la description des ressources informationnelles* en fournissant un *outil de contrôle du vocabulaire* qu'à partager les *concepts* et la *connaissance* qui leur est associé en permettant aux acteurs de consulter le contenu du *référentiel sémantique*. Nous proposons également un diagramme de synthèse pour faire le lien entre *métadonnées* et *ontologie* de la GIZC. Cette dernière peut, en effet, être formalisée comme une *extension* de la norme *ISO 19115*.

5.1 Intérêt d'un vocabulaire contrôlé

Dans un premier temps, nous mettons en évidence la nécessité de *maîtriser la sémantique associée à la valuation des éléments de métadonnées clés pour la GIZC* (descripteurs *thématiques* et *spatiaux*), compte tenu de la forte *polysémie* ou *synonymie* des concepts employés pour les renseigner. En cela, notre problématique s'apparente à des problématiques plus générales du *Web sémantique* et des *NTIC*.

Puis, en discutant sur les définitions de différents *outils de contrôle du vocabulaire* (*glossaire*, *thésaurus* et *ontologie*) et en réfléchissant sur leurs apports respectifs pour répondre aux problématiques soulevées dans le cadre du projet *Syscolag*, nous allons préciser les caractéristiques de l'outil que nous souhaitons mettre en place.

5.1.1 Contexte et problématique du Web, de la GIZC et de l'ISO 19115

Le *service de métadonnées*, accessible par le biais d'un *portail Web*, est destiné à des acteurs impliqués dans la GIZC. Le contexte pluridisciplinaire de la GIZC et du *programme Syscolag* donne une importance particulière au *verrou sémantique*. Le *partage des connaissances* entre les *acteurs* sous une forme interprétable par la *machine* apparente notre problématique à celles du *Web sémantique*. Nous illustrons ceci par des exemples sur la diversité de descripteurs possibles pour une même *indexation* et les différences d'interprétations possibles pour un même descripteur (respectivement propriétés sémantiques de *synonymie* et de *polysémie*). Enfin nous regardons plus particulièrement comment nous pouvons insérer ce contrôle terminologique dans un *service de métadonnées* qui s'appuie sur les recommandations de l'*ISO 19115*.

Problématiques générales des NTIC et du Web

La *maîtrise de la sémantique du vocabulaire employé pour produire, décrire ou consulter* des ressources dans les systèmes informatisés est un point clé pour optimiser le bénéfice que

peuvent en tirer leurs utilisateurs [22].

Les *métadonnées* participent à cet effort de contrôle terminologique avec le renfort *complémentaire* d'*outils de contrôle du vocabulaire* tels que les *ontologies* [98, 142]. Les exemples de systèmes implémentant de tels outils sur le Web vont croissants en particulier dans les sciences de la vie comme les domaines de la *génétique* (exemples *geneonto...*), de la *médecine* mais aussi de l'industrie où la terminologie plus cartésienne supporte plus facilement la formalisation des connaissances du domaine.

Il s'agit concrètement d'assurer que la méthode utilisée pour décrire les éléments de métadonnées, permet de retranscrire fidèlement la sémantique que le producteur de la métadonnée souhaite associer à ses *valuations par mots clés* (en respectant l'*intension* du producteur de la métadonnée en évitant équivoques ou amalgames ultérieurs sur la signification de l'indexation lors d'une consultation). Ceci permet, dans un deuxième temps, un gain en qualité dans la pertinence des réponses à des recherches de métadonnées [7, 57].

Compte tenu des lacunes actuelles des ressources en ligne, *métadonnées* et *outils de contrôle du vocabulaire* contribuent donc à la mise en place du *Web sémantique* [22].

Le *Web* étant perçu comme un apport majeur des *NTIC* pour la mise en œuvre de la GIZC [101], l'enjeu de la maîtrise sémantique des concepts clés de ce domaine est donc un *verrou* supplémentaire pour l'amélioration des systèmes d'information qui lui sont consacrés. On peut d'ailleurs constater dans la littérature actuelle un recoupement des problématiques du *Web sémantique* et d'élaboration de *référentiels sémantiques* dans des domaines associés à la GIZC et à l'*IG*, avec, par exemple, des titres de publications comme : "*Towards a Fishery Semantic Web*" [74] et "*Toward the semantic geospatial Web*" [54].

Nous pouvons donc nous appuyer sur les réponses techniques génériques du *Web sémantique* pour enrichir la mise en œuvre d'une solution au verrou sémantique dans la GIZC (voir partie 2.3.1).

Problématique générale de la GIZC

Nous avons vu, dans la partie 2.1.2, que la GIZC (et le programme *Syscolag*) est marquée par des problèmes d'échanges terminologiques dus à la richesse humaine et thématique de notre domaine et que la constitution de *langages communs* (ou *référentiels sémantiques*) est alors un verrou majeur pour la réussite des processus de GIZC (voir partie 2.1.3). Ceci est en effet essentiel pour le *partage d'information et de connaissances* entre les acteurs, et dépasse donc le cadre de la mise en place de systèmes informatisés, puisque, dans l'absolu, l'échange interdisciplinaire passe par une maîtrise satisfaisante des concepts communs essentiels à la description de problématiques transversales [56] (de manière à limiter l'ambiguïté sémantique de certains termes selon leur contexte d'utilisation).

Il s'agit donc, dans un premier temps, d'*inventorier les concepts essentiels à partager* et de leur donner une *lisibilité* par l'*attribution de définitions et propriétés communes et consensuelles* autant que faire se peut (ou *particulières* à certaines disciplines si nécessaires pour respecter la *polysémie* légitime de certains concepts).

Dans un deuxième temps, il s'agit de *partager une structuration de ces concepts* qui rende compte d'une *perception commune des éléments principaux et de leurs interactions au sein du système étudié* (voir partie 2.1.3 sur la nécessité d'une approche commune). Par ce dernier point, il devient possible de *modéliser une part, même élémentaire, de la connaissance experte* et de contribuer à la mise en place de *représentations partagées* dans notre domaine.

Ceci permettra finalement de restituer l'intension avec laquelle un acteur utilise un *concept* en

identifiant sans ambiguïté le sens qui lui est donné à l'aide un *référentiel sémantique*. Ainsi, au sein des ressources informationnelles ou des métadonnées que produisent ou consultent les acteurs, il sera possible de mettre en place une *traçabilité* de la signification qui a été donnée aux *concepts* parmi les *définitions* ou les *synonymes* possibles.

Pour illustrer plus en détail la nécessité de cette approche dans le domaine de la GIZC, nous allons montrer qu'un même *terme* peut avoir, selon son *contexte*, des *significations* distinctes, et des termes différents, une même signification. Ceci correspond respectivement aux *propriétés sémantiques* de *polysémie* et de *synonymie*. Voici quelques exemples de concepts caractéristiques de la richesse sémantique dans le contexte de la GIZC :

- Le terme *ressource* signifie implicitement *ressource informationnelle* dans ce document, d'un point de vue *informatique*, alors qu'un *économiste* le percevra comme une *ressource naturelle* (sorte de *ressource économique*) de même qu'un *écologiste* comme *Théodore Monod* qui comprend clairement le sens implicite que l'économiste à ce terme mais qui réproouve l'emploi du terme *ressource* pour désigner un *élément naturel* et lui préfère le terme *valeur* [127]. Il y a ici *polysémie* du terme *ressource* selon sa signification *économique* ou *informatique* et *valeur* est un *synonyme* possible de *ressource* en *écologie*. Si on souhaite maîtriser la sémantique du terme *ressource*, en distinguer les différentes interprétations possibles lors de l'indexation ou de la construction de requêtes pour la recherche d'information sur la GIZC dans notre service de métadonnées, il s'agit donc d'"étiqueter le sens" (ou *intension*) dans lequel ce terme est employé à l'*indexation* et de le restituer en *consultation* afin de ne pas générer d'équivoques. D'où la notion de *traçabilité* évoquée auparavant.
- *Bar*, *loup de mer*, *Dicentrarchus labrax* sont trois termes différents ou *synonymes* qui désignent la même *espèce* dans le *langage vernaculaire* pour les deux premiers ou dans le *langage scientifique* pour le dernier (qui sert normalement d'identifiant universel mais n'est connu que des initiés, or tous les acteurs mobilisés dans la GIZC ne le sont pas). On note que s'il fallait un jour prendre en compte la *connaissance autochtone* dans l'élaboration de notre référentiel sémantique, il s'agirait alors de connaître les *synonymes* de *Dicentrarchus labrax* en *occitan*, en *catalan* ou dans le *jargon de la pêche* ¹. Ceci est fastidieux mais nécessaire si on veut respecter le langage d'un utilisateur qui recherche *toutes les ressources disponibles sur cette espèce* (quelles que soient ses connaissances et les termes employés pour décrire les ressources lors de l'indexation) et qui, dans la construction de sa requête de consultation, n'emploiera pourtant qu'un seul de ces termes. La machine doit être capable d'interpréter que la requête porte en réalité sur l'ensemble des *synonymes* existants pour le concept demandé (expansion automatique des requêtes).

On comprend mieux ainsi la nécessité d'*inventorier ces concepts*, de *décliner leur signification* en fonction des disciplines constitutives de la GIZC (les liens de *polysémie* ou *synonymie* existants entre eux) puis de *stocker cette information* sous une forme interprétable par la machine pour qu'elle puisse en tenir compte en *saisie* comme en *recherche de métadonnées*.

Nous présentons dans la partie suivante les *éléments de métadonnées* de l'*ISO 19115* qui peuvent être valués par de tels *termes* ou *mots clés* et que nous souhaitons donc renseigner à l'aide d'un *outil de contrôle de la valuation*.

¹Parmi les noms vernaculaires attribués à l'espèce on peut citer : *loup* ou *loup blanc* en *Méditerranée*, *loubarrou* en *Roussillon*, *loubassou* en *Provence*, *ragnola* en *Corse*...

Problématique propre au service de métadonnées

Les concepteurs de la norme *ISO 19115*, conscients des problèmes sémantiques que nous venons d'évoquer, encouragent explicitement le contrôle de l'élément de métadonnée *mot clé*. Mais il n'existe pas de solution idéale dans l'absolu. La pertinence de la solution dépendra des utilisations potentielles de l'élément *mot clé* pour l'indexation des ressources, au cas par cas, selon le contexte d'implémentation. Nous détaillons dans cette partie quelques éléments de réponses génériques donnés par la norme *ISO 19115* et d'autres, plus spécifiques à notre contexte issus de la consultation des utilisateurs du projet *Syscolag* au sujet de leur perception de l'utilité de l'élément *mot clé*.

Point de vue général sur l'ISO 19115 Si la norme *ISO 19115* (que nous utilisons pour la mise en place de notre *service de métadonnées*) contrôle le remplissage d'une partie de ses éléments (voir partie 4.2.2), il reste cependant un certain nombre d'éléments qui restent en *texte libre* à la valuation et parmi ceux-ci l'élément *mot clé* (élément 52 de la norme). Or cet élément joue un rôle pivot dans l'indexation des ressources et donc, a posteriori, pour la recherche d'information.

En outre, comme le souligne [121], si la description générale des caractéristiques techniques des ressources peut se faire de manière *objective*, l'indexation à l'aide de **mots clés** pour la description des *caractéristiques thématiques spatiales* est, quant à elle, une tâche *subjective* puisqu'il est possible, au sein d'une organisation ou d'une discipline, de trouver des descriptions différentes d'une même ressource. À l'inverse, on peut trouver une même description employée dans différents sens (*polysémie*) pour décrire des ressources informationnelles distinctes. Il y a donc ici une source potentielle de confusion que l'on doit canaliser, en particulier dans le contexte plurithématique de la GIZC [15] (voir exemples de la partie 5.1.1).

Les concepteurs de la norme *ISO 19115* sont toutefois conscients de l'importance de ces problèmes de contrôle du vocabulaire et suggèrent la *maîtrise sémantique* de l'élément de métadonnées *mot clé* et préconisent donc l'usage d'un *outil de contrôle du vocabulaire* adapté au contexte d'implémentation. À ces fins, l'élément 53 de la norme permet de spécifier l'origine des *concepts* clés employés pour renseigner l'élément *mot clé* en indiquant le *thésaurus de référence* dont ils proviennent. Ils proposent également, comme alternative aux mots *mots clés* spatiaux, de maîtriser la description de l'*étendue géographique de la ressource* par l'utilisation de descripteurs géométriques spécifiques conformes aux recommandations de l'*OGC* (voir détail en Figure 6.2, partie 6.1.1).

L'exemple donné par l'*ISO 19115* (voir partie 4.2) pour l'utilisation de la rubrique *Metadata Extension Information* confirme la nécessité de cette démarche. En effet, il s'agit de contrôler le vocabulaire pour renseigner l'élément *mot clé* en ajoutant de nouveaux éléments de métadonnées et par l'extension du *CodeList MD_KeywordTypeCode*. L'exemple donné concerne des types de *mots clés* pour implémenter des *classifications hiérarchiques* basées sur des *taxonomies* [2] (dans l'exemple il s'agit de la mise en place d'une taxonomie relative à la *flore*, voir exemple 2, partie I.3 de l'*ISO 19115*). Nous y reviendrons par la suite en partie 5.1.3.

Voyons désormais comment ceci peut être mis à profit dans le cas de la GIZC.

Point de vue particulier des utilisateurs de la GIZC Il convient, dans notre cas, de s'intéresser à la perception que peuvent avoir les utilisateurs de l'élément de métadonnée *mot clé* afin de mieux en contrôler la saisie. En effet, derrière le terme *mot clé*, nous avons pu constater, notamment au cours d'une expérience de *mutualisation de références bibliogra-*

phiques, que les valeurs qui renseignent cet élément appartiennent généralement à différentes catégories. Nous pouvons en distinguer principalement quatre qui constituent logiquement des *critères clés pour indexer des ressources ou formuler des requêtes de recherche d'information dans le service de métadonnées* :

- *mots clés de description thématique* (océanographie, bathymétrie, sédimentologie...).
- *mots clés de description spatiale* (Languedoc Roussillon, étang de Thau, réserve marine de Banyuls...),
- *mots clés de description temporelle* (épisode géologique : orogénèse pyrénéenne, historique, rythmes saisonniers : printemps...).
- *mots clés de description méthodologique* (méthodes statistiques : ACP., modèles économiques...).

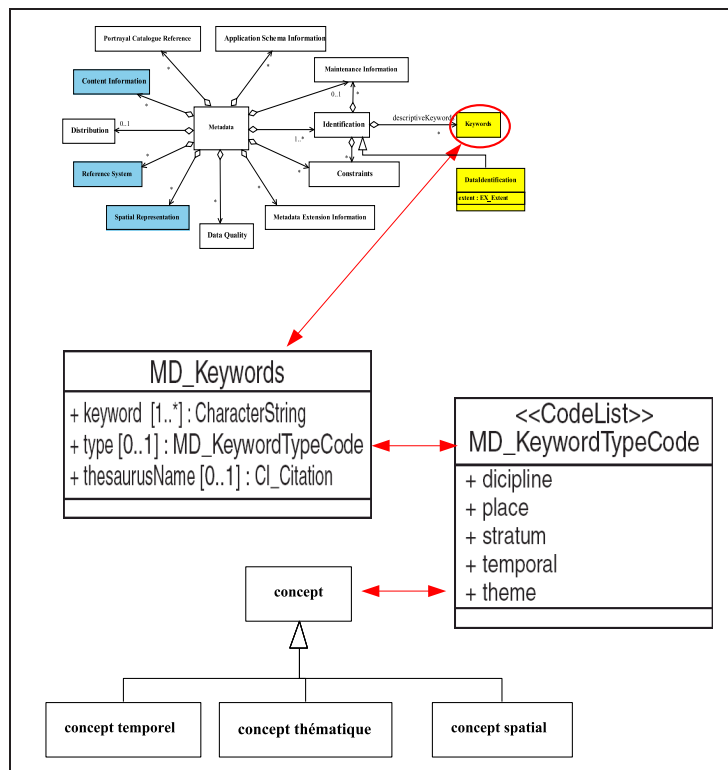


FIG. 5.1 – Les différentes sortes de concepts pour renseigner l'élément *mot clé* de l'ISO 19115

Si, dans les références bibliographiques, ces quatre catégories de *mots clés* sont traitées sans distinction dans un seul élément *mot clé*, la norme *ISO 19115*, quant à elle, permet de les différencier (en précisant la nature de chaque *concept* employé pour valuer un *mot clé*, voir Figure 5.1) mais elle n'en contrôle cependant qu'une partie.

Pour illustrer notre propos, nous prenons comme exemple une "*requête type*" de recherche d'information qui pourrait être formulée de la manière suivante :

"Je cherche une ressource informationnelle sur une thématique précise (1 - quoi), sur une zone bien localisée (2 - où), pendant une période donnée (3 - quand) et définie par telles caractéristiques techniques (4 - de quel type)".

Ainsi peut-on formuler une requête :

1. *Bathymétrie, pêche au chalut, loup de mer, substrat rocheux,*
2. *Côte catalane française,*
3. *1980-2000,*
4. *Cartes, mode vecteur, échelle x...*

Les contrôles suggérés par la norme concernent chacun de ces points mais avec plus ou moins de rigueur :

- La partie relative à la description du *type de ressource* recherchée (4) est codifiée par le biais de "*CodeList*" (voir partie 4.2.2).
- La partie relative à l'*intervalle de temps* (3) est contrôlée par le biais d'une autre norme : norme *ISO 19108 : GI-Temporal schema*.
- Le contrôle des parties (1) et (2) est plus lâche puisque les valeurs concernées ne peuvent relever que de l'élément "*mot clé*" (laissé en *texte libre*) et de l'*emprise spatiale* qui est soit définie également par un *mot clé* de type *spatial* (en *texte libre* également), soit évaluée par les positions en *degrés décimaux* d'un polygone décrivant l'emprise au sol de la ressource.

Pour rendre cette *requête type* efficace, ie *interprétable par la machine*, nous avons donc choisi de compléter les propositions de l'*ISO 19115* et de mettre en place un *vocabulaire thématique et spatial contrôlé* afin de proposer des listes de *mots clés* thématiques et d'*objets géographiques* d'intérêts pour la GIZC en LR.

La mise en place d'un *outil de contrôle du vocabulaire* pour la résolution de ces objectifs (*inventaire de différentes sortes de concepts* avec gestion de la *polysémie* et de la *synonymie*) dépasse très largement l'élaboration d'un simple *CodeList* de l'*ISO 19115* (voir partie 4.2.2). Nous distinguons trois sortes d'outils de contrôle du vocabulaire : *glossaire*, *thésaurus* et *ontologies*. Nous présentons dans la section suivante les caractéristiques de ces outils et leur *capacité en terme de retranscription de la connaissance portée par les termes qu'ils décrivent*.

5.1.2 Glossaire

Le glossaire est l'outil le plus simple de contrôle du vocabulaire. Néanmoins, sa réponse est limitée en terme de formalisation de la connaissance associée aux concepts du domaine qu'il permet d'inventorier.

Quelques définitions de glossaire

- **Définition édition** : Répertoire des termes tirés d'un **corpus** pour leur difficulté de compréhension et pour lesquels il est donné un **synonyme** connu ou une explication [149].
- **Définition informatique** : Répertoire électronique où sont consignés des mots ou expressions utilisés pour un **rappel rapide** [149].
- **Définition linguistique** : Répertoire qui définit ou explique des termes anciens, rares ou mal connus [149].

Nous considérerons ici le *glossaire* selon le sens que lui donne l'*informatique* et l'*édition*, pas uniquement pour des termes rares (puisque en GIZC les termes courants sont eux aussi sujets à des difficultés d'interprétations selon les utilisateurs) mais pour l'ensemble des *termes* ou *concepts* pouvant être utilisés pour renseigner l'élément de métadonnées *mots-clés* lors de la description d'une ressource. Un tel glossaire dépasse alors le simple *CodeList* de l'*ISO 19115*

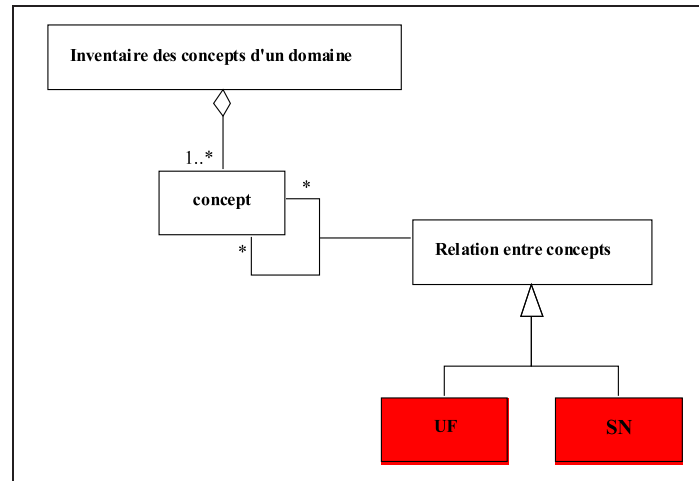


FIG. 5.2 – Les différentes relations standard d'un glossaire

(voir partie 4.2.2) puisque, dans notre cas, il est impératif d'éclaircir la ou les significations possibles des termes inventoriés.

Intérêt et limite du glossaire dans le cadre de Syscolag

Le *glossaire* répond donc à la nécessité d'*inventorier les concepts de notre domaine* et de décliner d'éventuelles *définitions multiples* selon les disciplines (gestion de la *polysémie* possible par des relations **SN** - *Scope Note* - entre concepts). Toutefois, il ne permet pas de réelle *structuration* des concepts, bien que la définition du domaine de l'*édition* lui attribuent la capacité de gérer la *synonymie* (par l'introduction de relation de type **UF** - *Used For* - entre les concepts), comme cela est indiqué en Figure 5.2.

Il permet néanmoins la mise en place rapide d'une *base terminologique* pour porter à connaissance les termes de référence dans notre domaine et d'un *outil de contrôle du vocabulaire* au sein de notre *service de métadonnées*.

Concrètement, un *glossaire* consiste généralement en une *liste organisée de termes déclinés par ordre alphabétique* (sous la forme d'un *menu déroulant* par exemple) mais son utilisation nécessite une recherche linéaire (non structurée) parmi les termes répertoriés, sans assistance dans l'aiguillage ou la gestion de liens (hormis éventuellement la *synonymie*), ce qui reste peu confortable pour l'utilisateur. Néanmoins, il est possible de faciliter l'usage des *concepts* pour la *saisie* ou la *consultation* de *métadonnées* par des fonctionnalités du type *complétion automatique*².

Cet outil de contrôle du vocabulaire est donc limité et les termes inventoriés ne sont pas interprétables par la machine. L'absence de formalisation des liens entre concepts interdit en particulier l'*expansion des requêtes* vers des concepts liés. La participation de l'utilisateur est donc assez lourde puisqu'il peut avoir à parcourir le *glossaire* pour y trouver le concept qui correspond à sa description ou son *synonyme* si le concept recherché n'est pas répertorié.

La capacité du système à fournir des résultats à une recherche de métadonnées est donc entiè-

²Un éditeur de texte fait de la complétion automatique quand il remplace une abréviation par la forme complète d'une expression lorsque vous tapez cette abréviation suivie d'une combinaison de touches magique. Cela permet d'économiser énormément les frappes au clavier [168].

FIG. 5.3 – Illustration d’interfaces graphiques du glossaire *UNESCO*

rement soumise à la saisie préalable de fiches indexées par le ou les termes employés dans la requête de consultation. En d’autres termes, les expansions de requêtes vers des termes liés ne sont pas possibles dans un glossaire (sauf si on lui attribue la gestion des liens de synonymie).

En dehors des *glossaires génériques* (comme *Le grand dictionnaire terminologique*³, que nous utilisons largement pour définir les concepts de ce document) ou propres à d’autres domaines, nous focalisons la présentation qui suit sur des outils de référence qui intéressent le domaine de la GIZC.

Glossaires pour la GIZC

Les glossaires sont les outils de contrôle du vocabulaire les plus nombreux dans des domaines impliqués dans la GIZC. Le *glossaire international d’hydrologie de l’Unesco*⁴ et le *glossaire de l’océanographie et des disciplines associées*⁵ sont deux outils de référence utilisables pour les termes du domaine de l’océanographie et des zones côtières au sens large. Le premier recense environ 1400 termes en 14 langues et constitue un outil complet pour le contrôle de la terminologie dans ce domaine (voir Figure 5.3). Il est, en particulier, utilisé dans le *service de métadonnées* de *NOKIS* (voir partie 2.3.2).

Dans les parties suivantes nous décrivons des outils de contrôle plus perfectionnés, en particulier par l’introduction de liens sémantiques entre les concepts inventoriés. Nous commençons par le *thésaurus*.

5.1.3 Thesaurus

Le *thésaurus* est un *outil de contrôle du vocabulaire* plus élaboré que le *glossaire*. En effet, il intègre une partie de la connaissance relative à un domaine par une *structuration des termes inventoriés* au moyen de liens de différentes natures. Nous détaillons son intérêt pour notre

³<http://www.granddictionnaire.com/> [149].

⁴<http://www.cig.ensmp.fr/~hubert/glu/aglo.htm>

⁵Glossary of Physical Oceanography and Related Disciplines

application à la GIZC et nous présentons quelques thésaurus existants dans des domaines proches.

Quelques définitions de thésaurus

- **Définition informatique** : *Le Thésaurus est un dictionnaire hiérarchisé comprenant les termes génériques, des termes associés et "employés pour". Il permet une homogénéité de vocabulaire lors de l'analyse de documents et de la saisie de références et une aide à la recherche/interrogation [81].*
- **Définition linguistique** : *Se dit des répertoires (lexiques ou nomenclatures) dans lesquels figurent des relations plus ou moins complexes entre les notions [149].*
- **Définition sciences de l'information** :
 - *Vocabulaire contrôlé de termes ayant entre eux des relations sémantiques et qui s'applique à un ou plusieurs domaines particuliers de la connaissance [149].*
 - *Vocabulaire normalisé et structuré de façon à éliminer les ambiguïtés du langage naturel utilisé pour l'analyse du contenu informatif des documents et la recherche automatisée de ces documents. Liste exclusive de termes à utiliser obligatoirement pour la caractérisation du contenu des documents à analyser, s'applique à un ou plusieurs domaines particuliers de la connaissance [149].*
 - *Définition de l'ISO 2788-1986 : "vocabulaire d'un langage d'indexation contrôlé organisé formellement de façon à expliciter les relations a priori entre les notions (par exemple relations générique-spécifique)" [5].*

Pour finir, une définition du *Thésaurus thématique* : *Thésaurus dont les relations entre des termes sont établies après regroupement de ceux-ci par domaines [149].*

Nous retenons la deuxième définition du domaine des sciences de l'information et celle de l'ISO 2788-1986 qui correspondent le mieux à notre perception du thésaurus. Pour ce qui nous concerne, il est important de relever dans les définitions de *thésaurus* par rapport à celles de *glossaire* le caractère organisé à travers la *structuration des termes inventoriés par des relations sémantiques* (voir Figure 5.4).

Les relations standards d'un thésaurus

Chaque terme est potentiellement en relation avec d'autres termes du *thésaurus* par des *relations* de différentes sortes (ceci est illustré en Figure 5.4), nous nous sommes appuyés sur les normes ISO 5964 [5] et ISO 2788 [4] pour le développement de *thésaurus multilingues et monolingues* :

- **BT - Broader Term/ TG - Terme Générique** : mise en relation avec des termes génériques,
- **NT - Narrower Term/ TS - Terme Spécifique** : mise en relation avec des termes spécifiques,
- **UF - Used For/ EP / ET - Employé pour** : mise en relation avec des termes équivalents / synonymes,
- **RT - Related To/ TA - Terme associé** : mise en relation avec des termes liés dans le fonctionnement du domaine considéré,
- **SN - Scope Note** : une note sur l'utilisation du terme ou sa définition [172],
- **SR** : la source de référence pour cette définition.

Remarque – **BT** et **NT** regroupent en fait deux types de liens : *spécialisation* et *agrégation*, respectivement **BTG** et **NTG** (*G* pour *generic*) et **BTP** et **NTP** (*P* pour *partitive*).

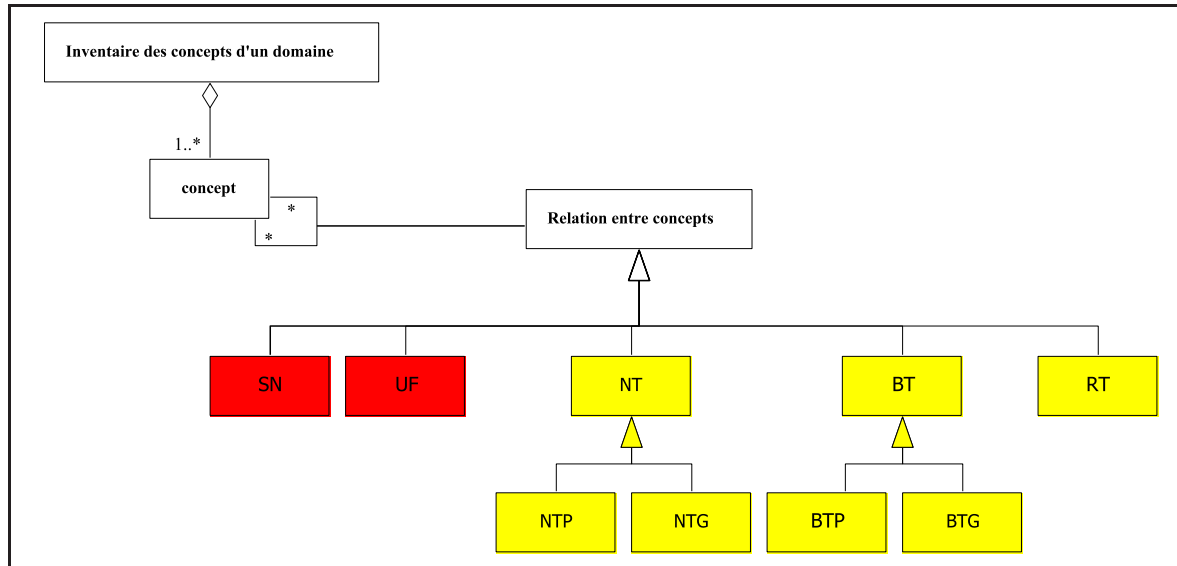


FIG. 5.4 – Les différentes relations standards d'un thésaurus

C'est dans la diversité de ces relations que réside l'intérêt du *thésaurus* par rapport au *glossaire*.

Intérêt du thésaurus dans le cadre de Syscolag

L'organisation d'un *thésaurus* permet d'introduire une *logique d'accès aux termes* recherchés selon une *connaissance* et une *représentation* partagées du domaine sur lequel il expose un point de vue, contrairement à la simple liste alphabétique issue de la constitution d'un *glossaire*. On peut donc assister la navigation dans les termes en limitant l'univers de recherche à ceux plus restreints des catégories du domaine correspondant au concept recherché. De même, pour la *synonymie*, il est possible de limiter la recherche en regroupant les synonymes dans leur catégorie thématique et non plus par ordre alphabétique.

Ces structurations introduisent donc plus de souplesse pour l'utilisateur qui navigue plus efficacement dans des sous-ensemble de l'inventaire des termes collecté (par opposition au *glossaire* qui présente l'univers complet des termes sans structuration), puisque les concepts sont regroupés (par des hiérarchisations thématique, des liens de spatialisation ou d'agrégation...). De plus, le thésaurus est également capable de gérer les *synonymes* ou les définitions multiples comme les *glossaires* au sens le plus large (voir partie précédente).

En outre, cette organisation de termes par des *relations sémantiques* rend cette structure *interprétable par la machine* et permet notamment des *expansions de requêtes*. Ainsi, imaginons que le concept *Loup de mer*, qui spécialise *poisson*, sert à indexer une ressource informationnelle. Il est donc possible si la requête "Sélectionner toutes les fiches de métadonnées qui concernent des ressources informationnelles indexées par le concept poisson" est muette, de proposer un *élargissement de la recherche* utilisant ce type de lien de spécialisation sur la classe poisson pour générer une nouvelle requête du type "Sélectionner toutes les fiches de

métadonnées qui concernent des ressources informationnelles indexées par le concept *Loup de mer*". En procédant de même, de manière itérative sur tous les liens identifiés dans le thésaurus entre poissons et d'autres concepts.

La représentation des connaissances de la communauté est donc valorisable par la proposition automatique de termes identifiés comme proches du thème à rechercher, ce que précisément le *glossaire* ne permet pas de réaliser (voir partie précédente).

A présent nous proposons l'illustration des ces caractéristiques par des thésaurus existants dans des domaines d'intérêts pour la GIZC.

Thésaurus pour la GIZC : état de l'art

Il existe de nombreux thésaurus de référence dans des domaines proches de la GIZC qui s'ajoutent aux glossaires existants pour établir un inventaire pertinent des concepts impliqués dans la GIZC ainsi que de leurs définitions. Ils permettent en outre de les structurer par l'introduction des relations standardisées propres au thésaurus.

Thésaurus des Nations Unies La *FAO*⁶ et plus généralement les *Nations Unies* sont très actives dans le domaine de la mise en place de *terminologie contrôlée* et multi-lingue. Un site Web spécifique est dédié au domaine de la terminologie et recense différents outils de référence de par le monde [63]. Nous en présentons quelques exemples dans les parties suivantes.

Thésaurus ASFA Le thésaurus *ASFA*⁷ est un outil en ligne de contrôle du vocabulaire dédié aux *sciences, technologies, gestion et conservation de l'environnement et des ressources aquatiques marines et continentales* en incluant les aspects *légal* et *socio-économique* [74]. Il comporte plus de 6000 termes ou *descripteurs* qui *entretiennent les relations classiques des thésaurus* (*NT, BT, RT, UF...*) (voir Figure 5.6) dont il donne des définitions. Il est également téléchargeable et disponible sous forme d'un fichier texte tabulé (voir Figure 5.5).

Thésaurus FIGIS Le système *FIGIS* est un *réseau global d'information intégrée sur les pêcheries*. Le *découpage thématique* du thésaurus consiste en quatre *sections* [74] :

- *espèces aquatiques* (information d'ordre biologique),
- *objets géographiques* (aires continentales et aquatiques, entités politiques),
- *ressources marines* (information sur l'état des ressources mondiales, données sur les stocks régionaux de poissons et les principaux paramètres affectant les stocks),
- *pêcheries marines* (données et cartes d'exploitation des principales espèces et information de gestion liées) et *technologies de pêche associées* (information sur l'identification des navires de pêches hauturières, sur la sélection des technologies, sur la législation internationale).

Ce thésaurus comprend approximativement 200 concepts haut niveau (*top-level concepts*), avec une profondeur maximale de 4 niveaux pour un total de 30 000 objets (*individus et concepts mélangés*) ainsi que des relations (*spécialisation* des catégories haut niveau et *instanciation*). Il est disponible en version multi-lingue [75].

A noter que *FIGIS* a défini ses propres *DTD XML* pour organiser ses bases de données. Ces *DTD* migrent à présent vers *RDFS*. Le jeu original comprend 823 éléments avec une riche structure d'attributs.

⁶Food and Agriculture Organization, <http://www.fao.org/>.

⁷Aquatic Sciences and Fisheries Abstracts (ASFA), <http://www4.fao.org/asfa/asfa.htm>.

ASFA THESAURUS		
Acoustic tags USE: Sonic tags	Acoustic wave refraction USE: Sound refraction	RT: A Trar
Acoustic telemetry BT: Telemetry RT: Acoustic command systems Acoustic tracking systems	Acoustic wave scattering USE: Sound scattering	Actiniu BT: A RT: R:
Acoustic tomography BT: Acoustic imagery RT: Acoustic holography	Acoustic wave transmission USE: Sound transmission	Actinor UF: Py Pyrg BT: R: RT: M
Acoustic tracking USE: Tracking	Acoustic waves USE: Sound waves	
Acoustic tracking systems UF: Underwater tracking systems BT: Detectors RT: Acoustic equipment Acoustic telemetry	Acoustics UF: Underwater acoustics BT: Physics NT: Bioacoustics Ultrasonics RT: Acoustic equipment Acoustic models Acoustic properties Echoes	Activat USE: S Activat BT: A NT: N Active i

FIG. 5.5 – Echantillon du thésaurus *ASFA* sous format texte tabulé

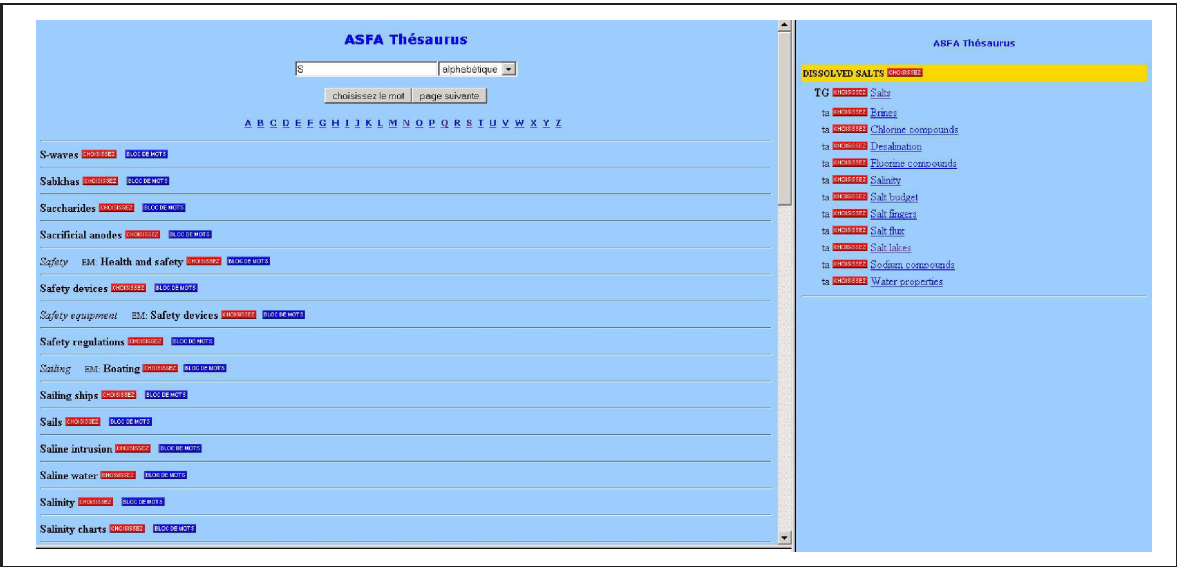


FIG. 5.6 – Illustration d'interfaces graphiques du thésaurus ASFA [66]

Thésaurus AGROVOC Le thésaurus *AGROVOC*⁸ a été développé par la *FAO* et la commission de la *Communauté Européenne* au début des années 80 et est utilisé pour l'*indexation* et la *recherche de documents*. C'est un vocabulaire multi-lingue structuré et contrôlé, pensé pour couvrir la terminologie de tous les champs thématiques qui concernent l'agriculture, la foresterie, les pêcheries, l'alimentation et les domaines liés [74]. Il existe différentes *relations hiérarchiques et associatives* entre les termes (*BT/NT*, *RT*, *ET*, *UF*). Sur 16000 termes, *AGROVOC* contient environ 2000 termes dédiés aux pêcheries. *AGROVOC* est téléchargeable sous la forme d'une base de données pour des usages non commerciaux sur demande à l'équipe de développement. Il est, de plus, possible de faire des propositions pour la prise en compte de nouveaux termes [75].

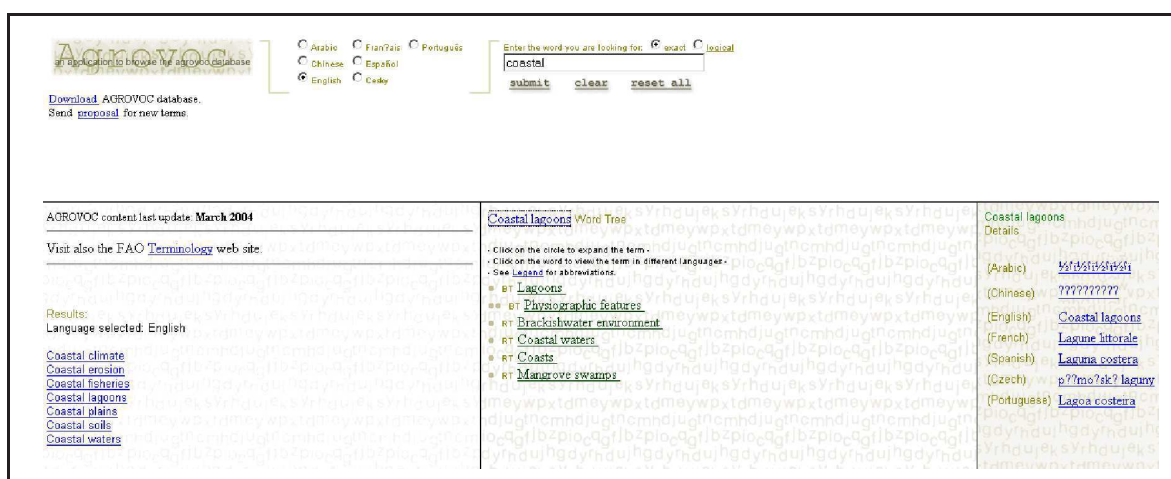


FIG. 5.7 – Illustration du thésaurus AGROVOC [64]

Thésaurus GEMET de l'Union Européenne [132][103]

*GEMET*⁹ (*General Multilingual Environmental Thesaurus*) a été développé comme outil d'*indexation*, de recherche et de contrôle pour l'*ETCCDS*¹⁰ (the European Topic Centre on Catalogue of Data Sources) et pour l'*EEA* (*European Environment Agency*¹¹). Les pays leaders de cette initiative sont l'Allemagne, l'Autriche, l'Italie, la Suède en collaboration avec d'autres pays de l'*UE* ainsi qu'avec *Infoterra* de l'*UNEP*.

Le principe de *GEMET* repose sur la fusion des meilleurs thésaurus multi-lingues existants de manière à économiser de l'énergie, du temps et des financements. *GEMET* se définit comme un *thésaurus générique* ayant pour objectif de mettre en place un langage commun général, un noyau de terminologies générales pour l'environnement.

Il est intéressant de noter que les thésaurus plus spécifiques (*conservation de la nature*, *déchets*, *énergie*...) ont été pour le moment exclus de la première étape de développement de *GEMET* et n'ont été pris en compte que du point de vue de leur structure et de leurs concepts les plus génériques (catégories de plus haut niveau).

⁸<http://www.fao.org/agrovoc/>.

⁹http://www.eionet.eu.int/gemet/index_.html.

¹⁰http://directory.eoportal.org/pres_EuropeanTopicCentreonCatalogueofDataSourcesETCCDS.html.

¹¹ Agence Européenne de l'Environnement <http://www.eea.eu.int/>.

GEMET est disponible en ligne ou téléchargeable (en *pdf*, en *XML*, sous formes de fichiers *RDF*, de base de données (Access©)), ou encore sous forme d'un *CD-Rom*. En outre, *GEMET* est aussi disponible en tant que Web Service ("*GEMET Web Service*") pour ceux qui veulent l'interfacer avec leur application (un service de métadonnées par exemple).

GEMET respecte les normes *ISO* sur les thésaurus monolingues et multi-lingues [5]. Les relations entre les termes sont de type *BT*, *NT*, *UF* et *RT*.

Afin d'améliorer la qualité terminologique de l'outil, une autre initiative en cours doit faire de *GEMET* un contenu ouvert permettant un travail collaboratif dans le jeu de *Web Services WIKI* (Wiktionary). Du point de vue de la GIZC, bien que non dédié à ce domaine, l'intérêt de cet outil est mis en avant dans les projets *COASTBASE* et *NOKIS* [103, 112] (voir parties 2.3.2 et 2.3.2).

European Environment Information and Observation Network (EIONET)
Networking improving Europe's environment

EIONET **GEMET Thesaurus**

English | Thematic listings | Alphabetic listings | Hierarchical listings

coastal ecosystem

Concept definition:
Marine environments bounded by the coastal land margin (seashore) and the continental shelf 100-200 m below sea level. Ecologically, the coastal and nearshore zones grade from shallow water depths, influenced by the adjacent landmass and input from coastal rivers and estuaries, to the continental shelf break, where oceanic processes predominate. Among the unique marine ecosystems associated with coastal and nearshore waterbodies are seaweed-dominated communities, coral reefs and upwellings. (Source: FAO/IOC)

broader terms
aquatic ecosystem

related terms
coastal environment

Scope note:
scope note is not available

Themes:
biology
natural areas, landscape, ecosystems
water

Groups:
BIOSPHERE (organisms, ecosystems)

Bulgarian:	Крайбрежна екосистема
Czech:	ekosystém pobřežní
Danish:	kystekosystem
German:	Küstenekosystem
Greek:	ποτάκιο οικοσύστημα
English (United States):	coastal ecosystem
Spanish:	ecosistemas costeros
Estonian:	ranikkuekosüsteem
Basque:	kostaldeko ekosistema
Finnish:	ranikon ekosysteemi
French:	écosystème côtier
Hungarian:	(tengerparti) ökoszisztéma
Italian:	ecosistema litoraneo
Dutch:	kustecosysteem
Norwegian:	kystekosystem
Polish:	ekosystem wybrzeży
Portuguese:	ecosistemas costeiro
Russian:	прибрежная экосистема
Slovak:	pobřežný ekosystém
Slovenian:	obalni ekosistem
Swedish:	kustens ekosystem

EEC Plus | Information | Support | About GEMET | Link website | Reference sources
2004 GEMET Thesaurus version 1.0

FIG. 5.8 – Illustration du thésaurus GEMET de l'Union Européenne [59]

Thésaurus du pôle relais Lagune Méditerranée Le comité interministériel français d'orientation et de suivi du plan d'action pour les zones humides a décidé le 2 octobre 2000 de créer six pôles relais, un par grand type de zone humide en France métropolitaine et en particulier un *pôle relais lagunes méditerranéennes* [150]. Son objectif est de relayer le plan d'action vers les gestionnaires de terrain avec des objectifs communs :

- recueillir et mettre à disposition la connaissance sur les divers aspects de la gestion des lagunes en établissant un *annuaire des compétences* et une *base de données bibliographiques*,
- promouvoir une gestion durable,
- favoriser les liens à l'échelle nationale.

Cette initiative est liée à l'initiative internationale *MedWet* (*Mediterranean Wetlands Initiative*) [125].

Le *pôle relais lagunes méditerranéennes* a notamment pour objectif d'améliorer l'accès à l'information sur les ressources existantes par la constitution d'une base de références bibliographiques sur les lagunes de méditerranée en général et du *Languedoc Roussillon* en particulier.

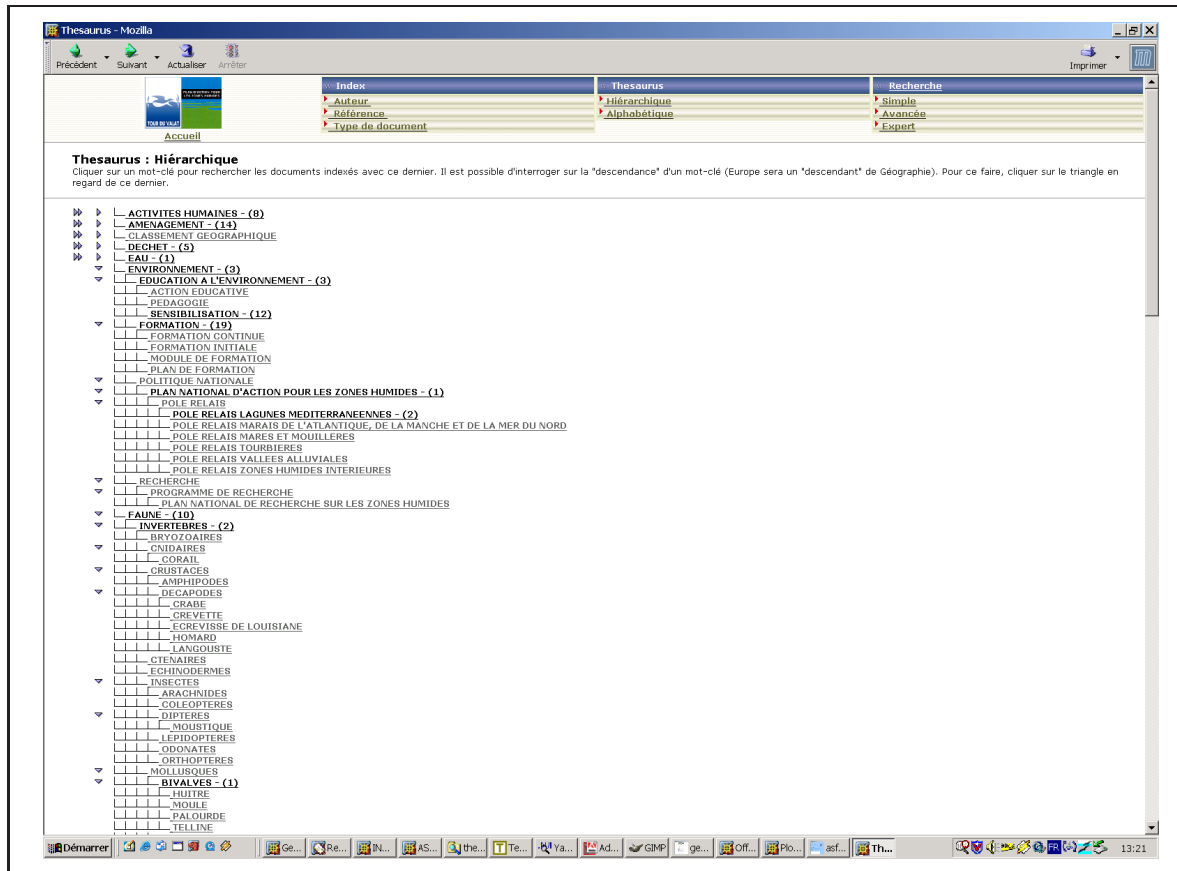


FIG. 5.9 – Illustration du thésaurus du pôle relais lagunes méditerranéennes [150]

A ces fins, l'utilisation d'un logiciel propriétaire (*Alexandrie*) a permis de développer et de mettre en ligne un *thésaurus*¹² dédié à ce domaine.

Ce *thésaurus* présente un inventaire conséquent de concepts relatifs aux lagunes méditerranéennes en général et à celles du LR en particulier. Ces concepts ont été inventoriés d'après les indexations de références bibliographiques sur le sujet. L'apport de cet outil est donc précieux pour l'inventaire d'une partie des termes de notre domaine. En effet, il peut aider à la mise en place d'un outil plus général pour le domaine de la gestion intégrée de la zone côtière en *Languedoc Roussillon* puisque celle-ci doit prendre en compte les lagunes.

Exemple de taxonomie donné par l'ISO 19115 Pour illustrer l'ajout de nouveaux éléments de métadonnées (en utilisant la rubrique *Metadata Extension Information*, voir partie 4.2.1) et l'extension des valeurs possibles d'un *CodeList*, l'ISO 19115 montre un exemple de classification basée sur une *taxonomie* (que nous entendons ici dans le sens de *thésaurus*). Aux nouveaux éléments de métadonnées correspond une extension du dictionnaire de données qui leur est dédiée ainsi que le diagramme de classes UML présenté en Figure 5.10.

Cet exemple est important puisqu'il montre une manière possible de faire le lien entre l'implémentation de l'ISO 19115 et celle d'un outil de contrôle du vocabulaire par une forma-

¹²<http://pole-lagunes.org/web/index.php>.

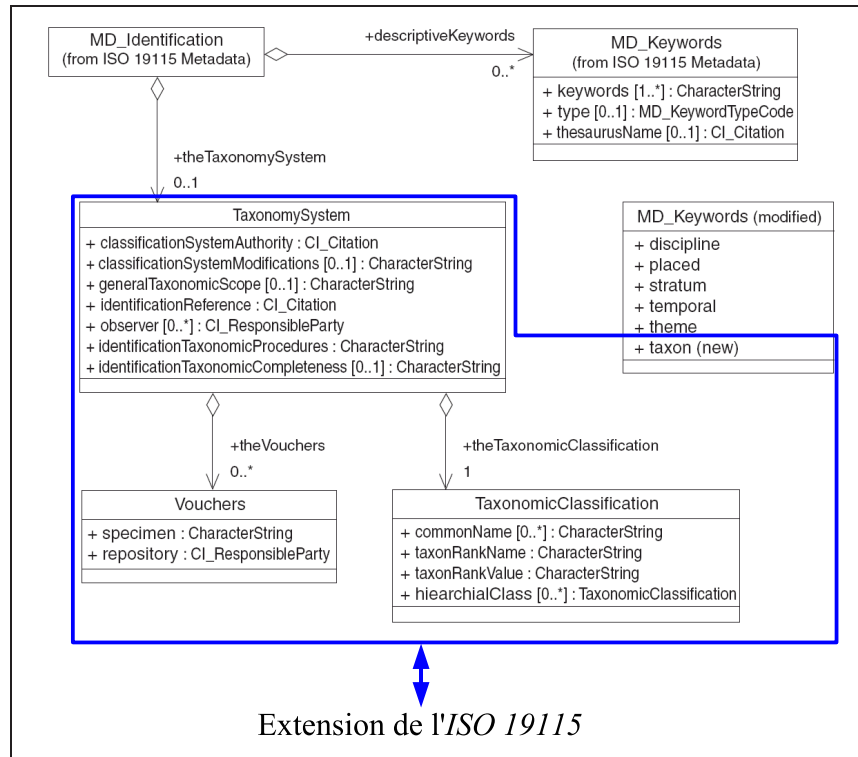


FIG. 5.10 – Exemple d’extension de la norme *ISO 19115* pour la mise en place d’une *taxonomie* pour décrire un nouveau type d’élément *mot clé* [2]

lisation *UML* qui respecte les recommandations de la norme.

L’ontologie permet d’aller plus loin que le glossaire ou le thésaurus dans la description des concepts ainsi que de leurs relations et permet une adaptation plus fine au domaine d’application. L’exploitation des connaissances associées aux concepts est ainsi optimisée, au bénéfice des acteurs comme des machines.

5.1.4 Ontologie

L’*ontologie*, comme la *métadonnée*, est un terme d’origine ancienne, employé plus récemment dans le domaine de l’informatique pour décrire un *outil de contrôle du vocabulaire* plus puissant que le *thésaurus* grâce à l’*augmentation de la richesse sémantique des liens introduits entre les concepts inventoriés* ainsi que la possibilité de leur *attribuer des propriétés* et des *instances*. Grâce à ces caractéristiques, l’*ontologie* permet une gestion plus fine des éléments thématiques et spatiaux.

Ainsi, le potentiel d’une *ontologie* est plus intéressant, en particulier du point de vue de l’*expression* et de la gestion des *connaissances expertes d’un domaine*.

Quelques définitions d’ontologie

Pour introduire le sujet, on peut dire qu’en informatique le terme *ontologie* a été d’abord utilisé dans les domaines de l’*intelligence artificielle* et de la *représentation des connaissances*

[65]. Son champ d'application s'est considérablement élargi avec l'émergence du *Web sémantique* et il fait désormais partie d'une terminologie couramment utilisée, notamment dans le secteur de la *modélisation des systèmes d'information et du XML* [99, 158].

En outre, comme les *métadonnées*, ce terme, dans son sens de structuration de concepts, n'est pas employé seulement en *informatique* mais aussi par tous ceux concernés par l'*indexation de documents* : en particulier les documentalistes et les bibliothécaires. Ce qui rend un peu plus évident pour nous les liens que peut nourrir l'*ontologie* avec les *métadonnées* qui, comme nous l'avons évoqué plus haut, sont issues directement des fiches descriptives de bibliothèques.

- **En philosophie** : Le terme *ontologie* apparaît pour la première fois en *philosophie* dans la *Grèce antique* avec *Aristote* et désigne une *partie abstraite de la métaphysique centrée sur l'étude de l'être en tant qu'être, indépendamment de ses déterminations particulières* [149].
- **En informatique** : Le terme *ontologie*, dénomme un "*système de classification*". On doit comprendre par là que l'*ontologie* s'apparente à la famille des outils qui structurent les concepts, parmi eux : *taxinomie/taxonomie, thésaurus...* Une *ontologie* décrit de manière générique les connaissances propres à un domaine donné et offre de celui-ci une *compréhension consensuelle* [149].

Pour ce qui nous concerne, nous retiendrons la définition de *Gruber* [84] :
une ontologie est une spécification explicite et formelle des termes d'un domaine et des relations qu'ils entretiennent entre eux.

Sous une autre formulation, nous retenons l'objectif suivant pour le travail qui nous intéresse [130] : *une ontologie définit un vocabulaire commun pour des chercheurs qui ont besoin de partager de l'information dans un domaine (la GIZC en LR pour notre part). Elle comprend des définitions des concepts de base du domaine et les relations entre eux sous une forme interprétable par la machine.*

Par ces seules définitions, la frontière entre *thésaurus* et *ontologie* est délicate à percevoir. Nous allons tenter de définir plus concrètement en quoi ces deux outils diffèrent.

Par *spécification des concepts* et des *relations*, il s'agit concrètement de détailler les *relations* entre concepts, les *attributs* ou *propriétés* associées aux concepts et leurs *instanciations*. L'attribution de propriétés aux concepts, due à l'*approche objet*, est une nouveauté par rapport au *thésaurus*. De plus, dans une *ontologie*, la nature des relations entre concepts est également enrichie par rapport au *thésaurus* [97].

C'est dans la nature des relations supplémentaires de l'ontologie que réside son intérêt par rapport au thésaurus. En effet, ce dernier établit des liens sémantiques entre termes mais ne permet pas de relier des termes de catégories thématiques distinctes par le biais de relations spécifiques à un domaine (voir détail en partie 5.1.4).

Pour définir une ontologie donnée, il est préalablement essentiel d'en décrire les objectifs, puisque ce sont eux qui conditionnent sa mise en place.

Objectifs de l'ontologie

Une ontologie peut avoir plusieurs objectifs et leur définition est essentielle à sa mise en œuvre. En effet, elle conditionne la réalisation des différentes étapes de l'élaboration d'une ontologie : l'inventaire des concepts, l'attribution des propriétés et l'introduction des relations entre concepts tiennent directement compte des objectifs visés [130].

Les objectifs liés à la mise en place d'une ontologie, pour la part qui nous intéresse, et par ordre d'importance, peuvent être les suivants :

- *contrôle du vocabulaire* pour l'*indexation* et la *recherche de métadonnées par mots clés*,
- *intégration de la connaissance* experte par instanciation des classes du modèle et des liens possibles entre concepts du domaine (voir notion de *Base de connaissances* en plus loin dans cette partie 5.1.4),
- *expansion de requêtes*,
- *génération dynamique de liens hypertextes*, par exemple, lors de la consultation du sens d'un concept, en restituant, sous forme d'hyperliens, la liste des concepts qui lui sont liés par différentes relations.

En fonction de ces objectifs, pour un même domaine, l'inventaire des *concepts* ainsi que la nature de leurs *relations*, *propriétés* et *instances* peuvent être réévalués.

Les relations standards d'une ontologie

Aux relations standardisées entre concepts définies par les *thésaurus*, les *ontologies* permettent d'ajouter des relations plus riches sémantiquement.

En effet, les types de relations proposées par les *thésaurus* ne sont pas suffisamment fines pour exprimer explicitement certaines *connaissances élémentaires* dans notre domaine. La nature (ou la qualité) des relations entre concepts offerts par un *thésaurus* étant trop grossière pour cela [105].

Il est difficile, par exemple, de traduire à l'aide d'un *thésaurus* la connaissance suivante : *un pêcheur du LR pêche du thon rouge*. Le *thésaurus* permet d'exprimer le lien entre *pêcheur* et *thon rouge* à travers une relation *RT*, mais il n'exprime pas la nature ou la qualité de ce lien, qui est un lien de *pêche*. Nous voudrions exprimer que *RT=pêche*, ie qu'il s'agit bien de *RT* mais d'une sorte particulière : nous voulons *spécialiser RT*.

L'ontologie, quant à elle, permet d'exprimer la ou les significations de n'importe quel verbe et d'en faire autant de *sortes de relations ontologiques* (par exemple à l'aide d'*URI*, voir détail en partie 5.2.1). Elle offre donc une richesse supplémentaire dans la description des liens entre concepts, et autorise concrètement la spécialisation de la relation *RT* qui permet d'exprimer n'importe quel verbe essentiel à notre domaine (voir Figure 5.11). Nous représentons ceci en Figure 5.11 par la spécialisation de la classe *RT* en une classe ***RTS*** (pour *RT Spécialisée*).

En outre, l'ontologie se distingue du *thésaurus* par la possibilité de mettre en place des *propriétés* pour les *concepts* qui peuvent être *héritées* d'un concept plus générique.

Propriétés des concepts d'une ontologie

Un trait distinctif de l'*ontologie* par rapport au *thésaurus* est donc la possibilité d'*attribuer des propriétés aux concepts inventoriés*, de la même manière que, dans l'*approche objet*, une *classe* peut disposer d'*attributs* ou de propriétés. A ce propos, compte tenu des formalisations présentées à l'aide de la notation objet *UML*, nous employons indifféremment dans ce document les notions de *classe* ou de *concept* comme *élément* d'une *ontologie*, comme c'est le cas dans [176].

Cela permet, par exemple, d'exprimer qu'à *tout animal* est associé une *biométrie* et un *biotope*, qu'un *biotope* peut être décrit par un terme qu'on peut représenter cartographiquement (comme une *aire de répartition*) ou qu'à un *acteur* est associé une *adresse physique*, *électronique*, un *numéro de téléphone*, des *compétences*. ...

Lorsqu'on instancie ces propriétés on partage alors les connaissances recensées sur l'existence d'*individus* d'un *domaine*.

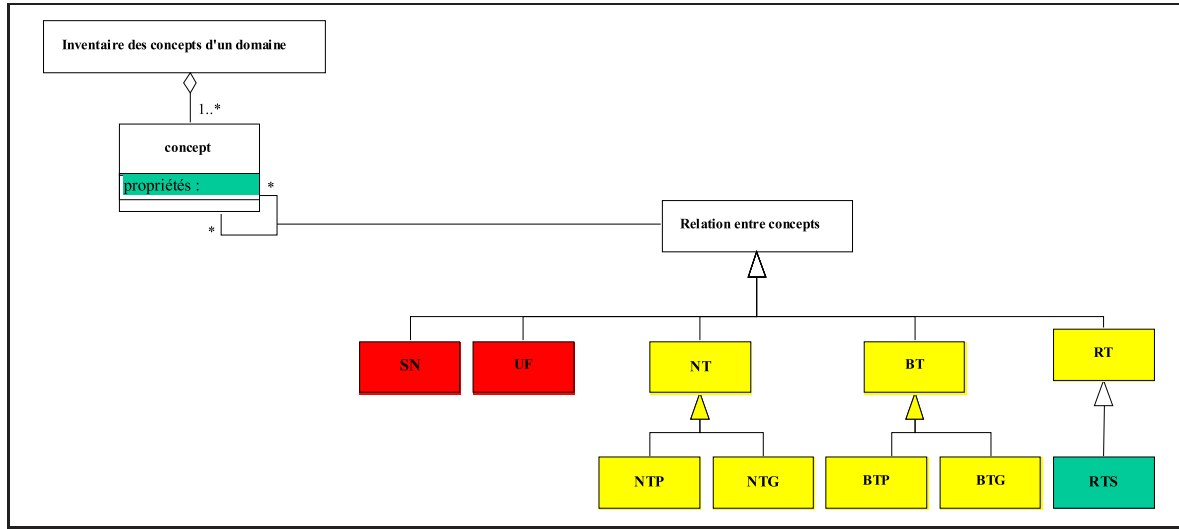


FIG. 5.11 – Les différentes relations standard d’une ontologie

Instanciation des concepts d’une ontologie

Une fois les classes *inventoriées* et pourvues d’attributs¹³, il est possible de les *instancier*. L’*instanciation des classes de l’ontologie* permet son adaptation à un contexte particulier d’implémentation.

Une ontologie du vin sera instanciée différemment selon le vignoble concerné [130]. Une même ontologie du domaine de la GIZC sera donc instanciée différemment selon la zone côtière considérée : la zone côtière du sud-ouest malgache bordée de récifs coralliens ou celle du Golfe du Lion ont peu de caractéristiques (physiques ou humaines) communes.

Par l’*instanciation des classes d’un modèle d’ontologie*, on intègre donc une part de la *connaissance experte* du domaine considéré. Ainsi, [130] parle de constitution de *base de connaissance*¹⁴ par l’*instanciation des classes d’une ontologie*. Cette définition est intéressante à considérer puisqu’un des objectifs du projet *Syscolag* qui souhaite développer une base de connaissances sans toutefois arriver à en définir strictement le sens [18].

A noter que [130] attire l’attention sur la différence, peu évidente lors de la mise en place d’une ontologie, entre *instancier une classe* et la *spécialiser*. Ceci est parfois flou en fonction du cas d’implémentation [130]. Soit, par exemple, la classe *chalutier* qui spécialise la classe *bateau de pêche*. Une instance de *chalutier* sera le *chalutier Toto* avec les propriétés suivantes : longueur 25 mètres, enregistré à Sète... Ce cas ne pose pas de difficultés. Par contre, *poisson* est une classe qui spécialise la classe *animal* et *Loup de mer* spécialise *poisson*. A ce niveau de spécialisation, l’*instanciation* de la classe *loup de mer* n’a pas de sens ou plutôt d’intérêt, à moins que la connaissance de tous les individus de la population de *loup de mer* en *Languedoc Roussillon* ne le permette et le justifie, ce qui n’est pas le cas. On pourrait donc dire d’après [130] que *loup de mer* instancie *poisson*.

¹³ Slots en anglais.

¹⁴ Knowledge Base (KB) en anglais.

Les différentes catégories d'ontologies

Il existe différentes catégories d'ontologie selon leur niveau de généralité ou de spécialisation, *ie* selon l'ouverture de leur domaine d'application. [70] distingue les ontologies suivantes :

- l'*ontologie top-level* (ou *haut niveau*) est une *ontologie* partagée par différentes communautés. A ce niveau, on peut distinguer des types d'ontologies très génériques par exemple dans les *domaines* suivants : l'*occupation du sol*, la *faune et la flore*...
- l'*ontologie de domaine* prend en compte . Cette ontologie est plus spécifique que l'*ontologie top level* mais encore suffisamment générique pour s'appliquer à différents espaces et pas seulement à une aire géographique donnée.
- l'ontologie appliquée prend en compte le cas spécifique de l'aire géographique à laquelle elle se réfère et dans laquelle on peut identifier les concepts plus en détail (par des relevés de terrain par exemple).

En outre il existe différentes approches pour mettre en place une ontologie :

- l'*approche bottom-up* ou *approche ascendante*, qui, récursivement, à partir des classes plus spécifique d'un même niveau, introduit des classes qui les généralisent jusqu'à la racine.
- l'*approche top-down* ou *approche descendante*, qui, à l'inverse, spécialisent des classes générales par des classes spécifiques,
- finalement, en pratique, il est possible d'utiliser une approche qui mixe les deux premières.

Intérêt de l'ontologie dans Syscolag

Compte tenu des attentes du programme *Syscolag* en terme de gestion des connaissances, les avantages de la mise en place d'une *ontologie* sont évidents par rapport à ceux d'un *thésaurus* et donc d'un *glossaire*.

L'objectif principal est d'enrichir la sémantique de la norme ISO 19115 utilisée en contrôlant, d'une part la publication des métadonnées référençant les ressources, et d'autre part en améliorant l'expressivité des requêtes formulables par les différents acteurs, permettant ainsi de cibler ce qui dans l'offre correspond effectivement à la demande [15].

Par des *liens sémantiques plus riches*, par l'*attribution de propriétés aux classes inventoriées* et par les possibilités d'instanciation de ces dernières, l'*ontologie* permet une *modélisation plus fine de la connaissance experte du domaine* qu'un *thésaurus*.

En particulier, il devient possible d'exprimer la nature (sémantique) des interactions entre les *éléments* constitutifs du système côtier et de faire comprendre les processus essentiels de notre domaine aux acteurs mobilisés dans la GIZC, *ie* de partager la connaissance. En cela nous répondons aux souhaits de la partie 2.1.3.

Voyons à présent les initiatives de référence existantes dans ce domaine ou des domaines proches.

Etat de l'art des ontologies dans le domaine de la GIZC

Il n'existe actuellement aucun outil opérationnel de langage commun dédié au domaine de la GIZC, que ce soit un *glossaire*, un *thésaurus*) et encore moins une *ontologie* [112, 141]. Cela constitue un verrou majeur pour le partage d'information et de connaissance. Il existe par contre des briques réutilisables dans des disciplines impliquées dans ce domaine. Cette lacune paraît normale compte tenu du caractère flou de la GIZC (comme nous l'avons mis en avant

dans le chapitre 2 en partie 2.1.2) qui complexifie lourdement la mise en place d'un tel outil, à commencer par l'inventaire des concepts pertinents dans notre domaine.

Comme nous l'avons vu dans l'état de l'art (voir partie 2.3.2), certains programmes de recherche proposent des ontologies sur des domaines associés à la GIZC sur lesquelles nous pouvons nous appuyer. Nous les détaillons à présent.

Exemple de l'ontologie FOS La FAO a fait le constat que les *thésaurus* détaillés en partie 5.1.3 sont des structures distinctes qui donnent des résultats hétérogènes pour la description des propriétés sémantiques d'un même terme [75].

A l'aide de *FOS* (*Fishery Ontology Service*), les thésaurus *Agrovoc*, *ASFA*, *FIGIS* ou *oneFish* ont été retravaillés de manière à les unifier en utilisant des *méthodes formelles d'ontologies* (voir Figure 5.12(a)). Le résultat est déployé sous forme d'une *librairie* nommée *Integrated Fishery Ontology* (constituée de plus de 35000 *classes* et 10000 *individus*) dans les organismes en charge de la gestion des *bases de données* et *répertoires* sur les *pêcheries* et directement servie sur le Web (à l'aide du serveur *FOS*).

Techniquement, on retrouve le souci de *réutiliser les briques existantes dans les domaines concernés* qu'elles soient générales ou spécialisées avec des distinctions entre *ontologies de domaine*, *upper-level*, *core ontology* (voir Figure 5.12(b) et définitions en partie 5.1.4). Les modèles présentés utilisent le formalisme *UML* dans des diagrammes de classes ou d'activités. Un des exemples donné dans la littérature associée à ces projets est celui du terme *aquaculture* dont la signification proposée diffère selon que l'on consulte *Agrovoc*, *ASFA*, *FIGIS* ou *oneFish*. Cet exemple met en évidence, selon le thésaurus considéré, la variété des *définitions* attribuée à un même concept (*polysémie*) et des *relations* qu'il peut entretenir avec d'autres concepts. Les formalisations utilisent le langage *OWL* comme langage d'implémentation du *Web sémantique* pour la mise en place de l'*ontologie* [74]. Cette dernière est d'ailleurs téléchargeable sous la forme d'un fichier *OWL*.

Hormis ces initiatives de la FAO, d'autres programmes de recherche (recensés dans notre état de l'art en partie 2.3.2) partagent l'objectif de mettre en place une *ontologie* dans des domaines concernés par la GIZC.

Ontologies issues des travaux conjugués de THETIS-POSEIDON-ARION *POSEIDON* est un système qui s'appuie sur la création de *métadonnées* et d'*ontologies* pour le traitement de données géophysiques distribuées (voir partie 2.3.2).

La nécessité d'utiliser une ontologie est justifié ainsi : *alors que nous pouvons chercher les métadonnées pour l'information dont nous avons besoin, nous avons besoin en plus d'un vocabulaire commun, ou ontologie, pour nous assurer que nous pouvons identifier les données sans ambiguïtés* [177]. Or, le constat de l'absence de vocabulaire recouvrant tous les aspects de l'océanographie et de la gestion des systèmes océanographiques a encouragé la mise en place d'un outil sur Internet permettant aux experts thématiques du domaine (océanographes, gestionnaires du littoral...) d'enrichir une ontologie du domaine [...]. Si nous pouvons trouver ou créer un tel langage, il pourra être implémenté en tant que vocabulaire contrôlé à l'intérieur du service de métadonnées (pour la valuation des champs des interfaces de saisie et de recherche) et permettra la cohérence des données utilisées [141].

Dans ce cas, l'ontologie est nécessaire en complément des métadonnées pour permettre les *correspondances* entre données issues de diverses sources et leur utilisation au sein des mêmes paramètres du modèle de prévision. Elle permet donc le traitement des ressources informa-

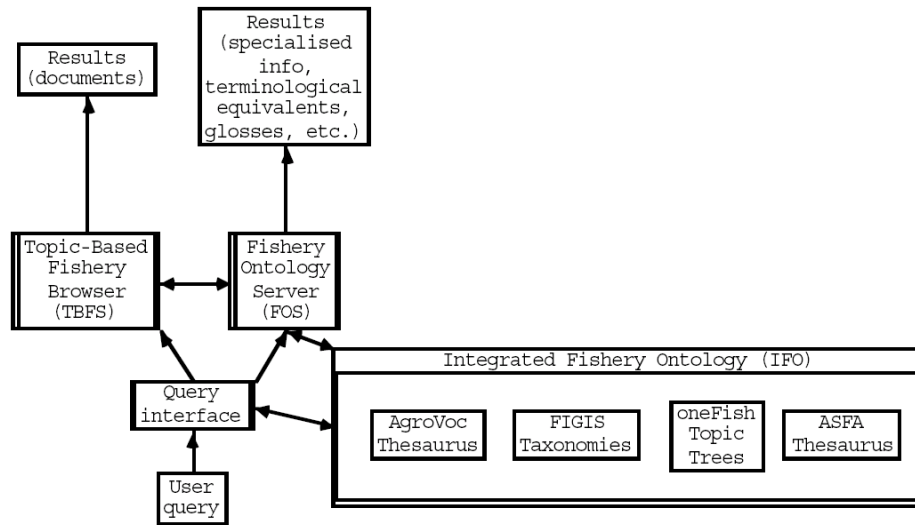
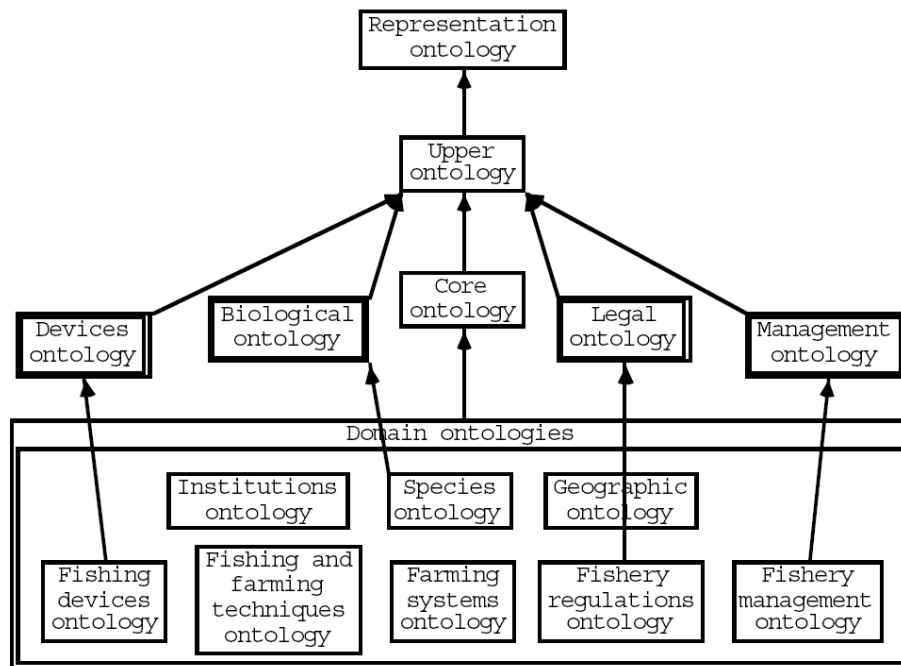
(a) Rôle d'intégration de *FOS* [75](b) Réutilisation d'ontologies de domaines pour *FOS* [75]

FIG. 5.12 – Structure et éléments de l'ontologie [75]

tionnelles en plus de leur localisation. L'ontologie proposée est basée sur la mise en place de catégories générales des domaines de l'*océanographie physique* et *acoustique* qui sont ensuite spécialisées par les utilisateurs eux-mêmes [177] (voir Figure 5.13). Afin de faciliter l'élaboration de l'ontologie des interfaces utilisateurs permettent, directement sur le Web, une construction collaborative plus conviviale [141] (voir Figure 5.14).

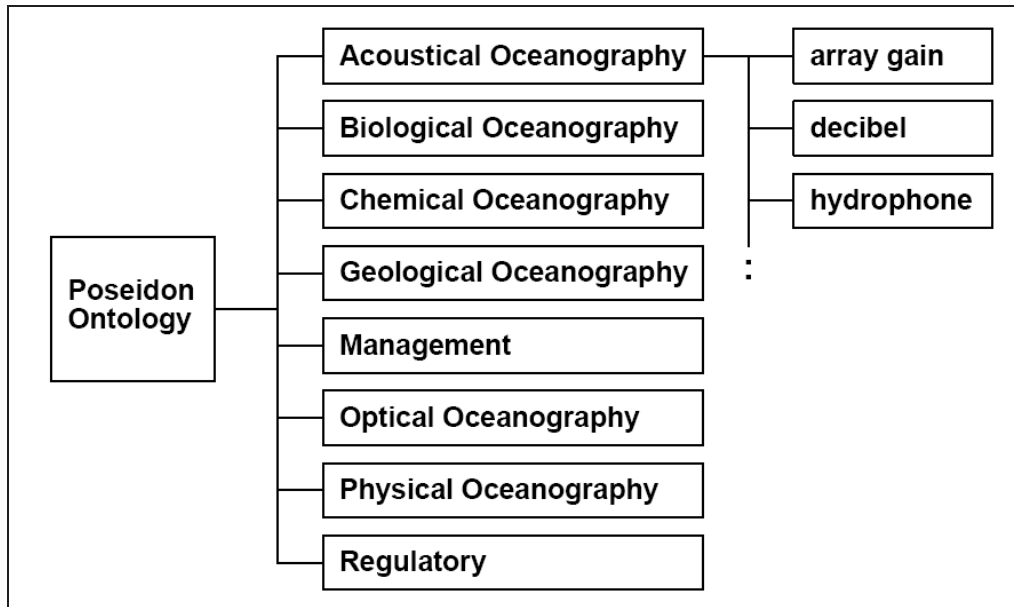


FIG. 5.13 – Catégories générales des domaines de l'océanographie physique et acoustique dans l'ontologie POSEIDON [177]

Cependant, la portée de ces résultats est limitée à des cas d'utilisation trop spécifiques, compte tenu de la variété des ressources informationnelles et des thématiques nécessaires à prendre en compte dans le domaine de la GIZC.

Exemple de l'ontologie ARION [161][89] Comme cela a été décrit dans la partie 2.3.2, le projet *ARION* a pour objectif d'améliorer la recherche et l'accès à des ressources réparties dans des collections scientifiques d'organismes distincts par le biais de d'interfaces Web dédiées. L'infrastructure mise en place s'appuie essentiellement sur des métadonnées et l'utilisation complémentaire d'ontologies.

A travers la mise en place d'une ontologie, le projet *ARION* souhaite [90] :

- améliorer la qualité des requêtes dans le service de métadonnées (pertinence et expansion de requêtes possible par la formalisation de la connaissance),
- améliorer le partage de la connaissance entre les acteurs mobilisés,
- réutiliser des briques existantes pour spécialiser certaines branches de l'ontologie,
- appliquer les standards permettant une utilisation en dehors du système mis en place.

La thématique de l'ontologie mise en place concerne la *houle océanique* (voir Figure 5.15). Les relations entre les concepts sont du type *BTS*, *RTS* (voir définitions en partie 5.1.4) auxquelles s'ajoutent les *propriétés* des *concepts* et leurs *instances*.

Techniquement *ARION* utilise le standard *RDF(S)* aussi bien pour la gestion de métadonnées que pour la mise en place d'une *base de connaissances interprétables par la machine*. D'après leur étude, l'*éditeur d'ontologies* le plus approprié et accessible pour être manipulé par les experts du domaine est *Protégé 2000* [162].

L'état de l'art des ontologies exploitables pour la GIZC est donc assez pauvre, mais il ne faut pas occulter que les *catalogues*, *glossaires* et *thésaurus* sont autant d'outils qui contribuent à la mise en place d'*ontologies* par les inventaires de concepts réalisés, l'attribution de

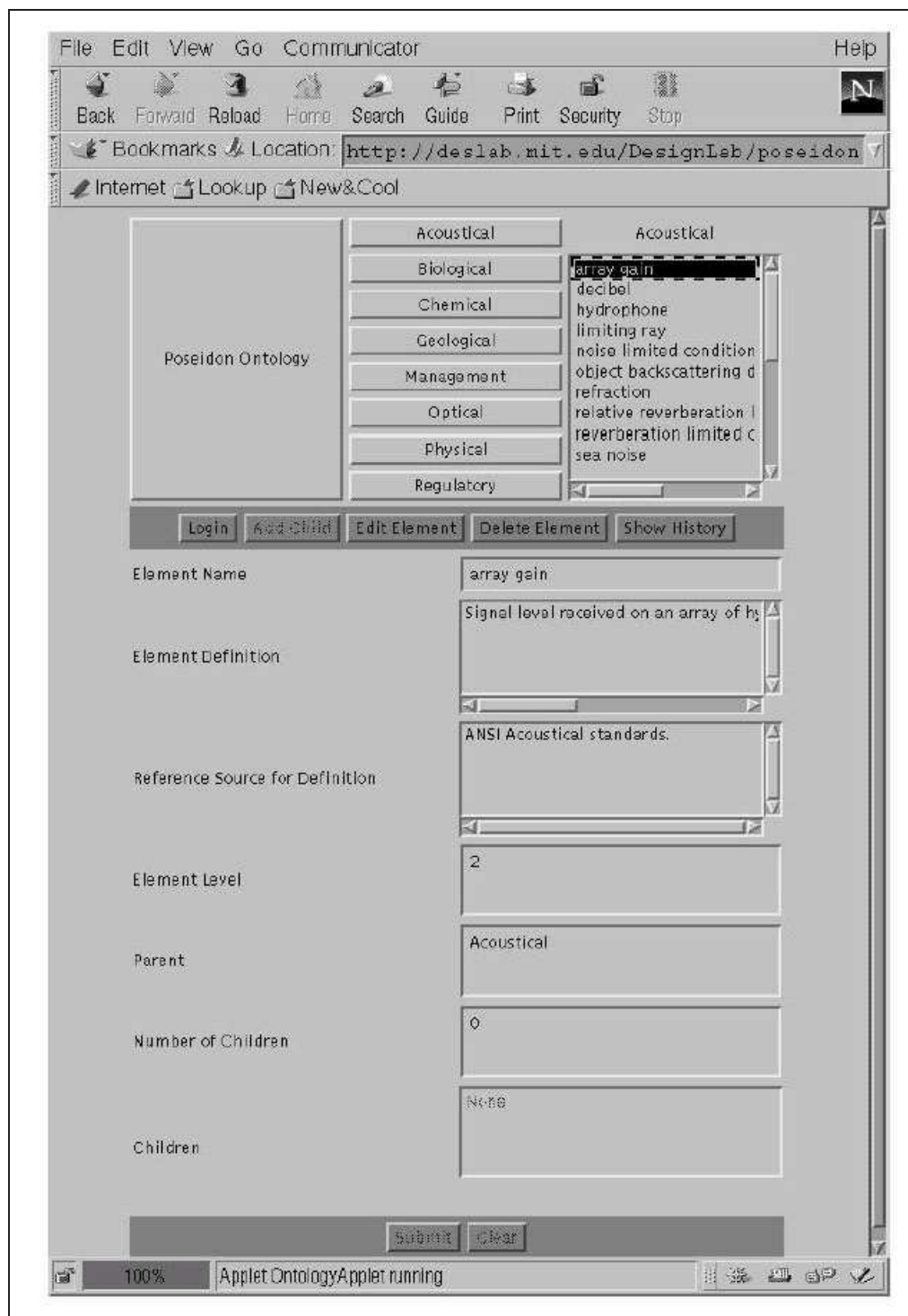
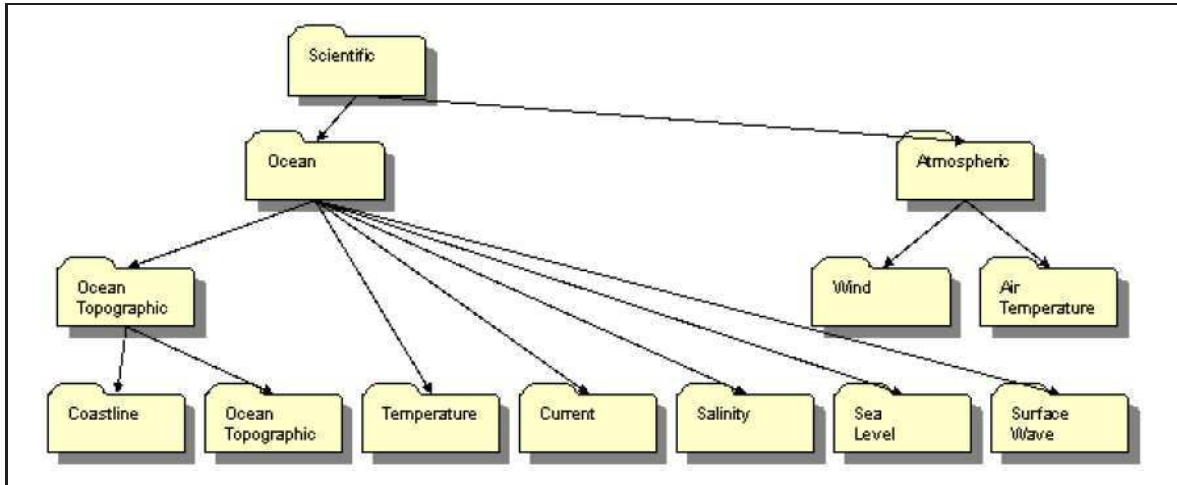


FIG. 5.14 – Exemple d'interface Web pour la construction collaborative de l'ontologie *POSEIDON* [141]

FIG. 5.15 – Exemple d'ontologie *ARION* pour le domaine de la houle océanique [162]

définitions, voire l'introduction de liens entre les concepts. Ils constituent ce que [178] appelle des *efforts ontologiques* et sont réutilisables pour l'élaboration d'une ontologie au sens strict. C'est précisément ce qui se profile à l'avenir avec la mise en place de groupes de recherches pour fusionner ces initiatives de créations d'ontologies marines, côtière ou littorales à travers des initiatives telles que *MarineXML* [126] ou *MOTIVE*, comme nous l'avons détaillé dans la partie 2.3.2. Le scénario idéal étant, comme pour l'*ISO 19115*, que les différentes expertises et expériences d'ontologies de domaine fusionnent en une ontologie plus générale et unique adoptée par l'ensemble des pays mobilisés dans cet effort de recherche.

Pour conclure cette section, nous proposons en Figure 5.16 une synthèse sur les caractéristiques des différents outils possibles pour le contrôle du vocabulaire.

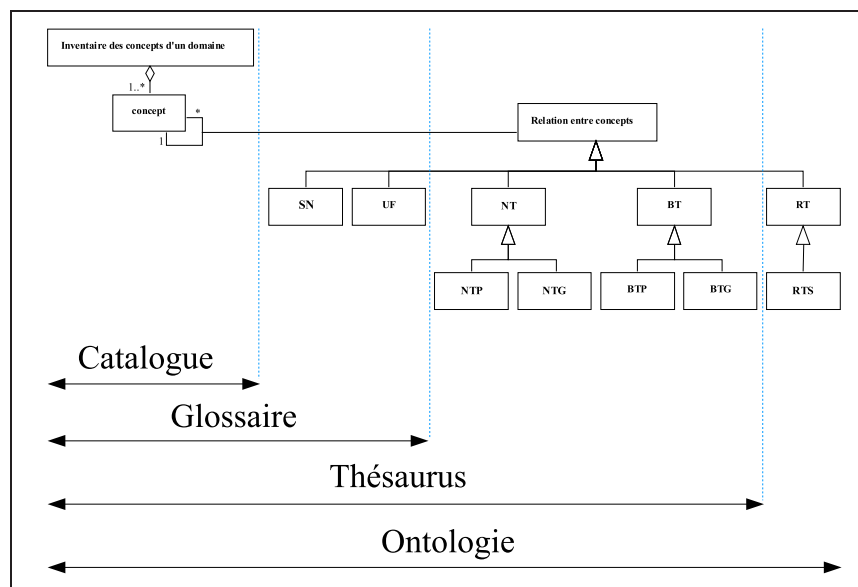


FIG. 5.16 – Complémentarité des différents outils de contrôle du vocabulaire

Synthèse de la section 5.1

- Le *partage des connaissances* associées à la GIZC implique la mise en place d'un *référentiel sémantique* exploitable par les *acteurs* et les *machines*, cette problématique rejoint celle plus générale du *Web sémantique*.
- En particulier, un *référentiel sémantique* permet le *contrôle du vocabulaire* dans un *service de métadonnées* afin d'optimiser la *description des ressources informationnelles* (et donc leur *localisation*) ainsi que l'aide à l'élaboration d'un *langage commun* dans un contexte pluridisciplinaire.
- Il existe différents types d'*outils de contrôle du vocabulaire* : *glossaires*, *thésaurus* ou *ontologies*, dans des domaines proches ou impliqués dans la GIZC.
- Les *ontologies* permettent une gestion plus fine de la *connaissance* associée aux *concepts* qu'elles inventorient. En effet, elles permettent d'attribuer des *propriétés* aux *concepts*, de les *instancier* et d'introduire des *liens sémantiques* plus fins en spécialisant les relations *RT* classiques des *thésaurus*.
- De plus, les *glossaires*, *thésaurus* et *ontologies* existants dans des domaines impliqués par la GIZC sont réutilisables pour la mise en place d'une *ontologie* spécifique à ce domaine. Ces initiatives de référence rassemblent des *éléments* et *interactions* qui permettent de décrire les *processus* essentiels de *sous-domaines* plus spécialisés (*pêcheries*, *océanographie*...). Cependant, elles ne peuvent pas répondre isolément au *contrôle du vocabulaire* dans notre *service de métadonnées* et encore moins à la *gestion des connaissances des partenaires impliqués dans notre programme*.

Nous proposons donc d'agréger ces briques existantes pour construire une *ontologie de la GIZC* en *Languedoc Roussillon* qui respecte le caractère *plurithématique* de ce domaine.

5.2 Proposition d'un modèle d'ontologie spécifique à la GIZC

Nous venons de voir qu'il n'existe pas actuellement d'*ontologie* dédiée au domaine de la *GIZC* mais que des efforts de *contrôle du vocabulaire* dans des *domaines connexes* émergent et peuvent être réutilisés comme briques constitutives d'une ontologie plus générale. En effet, d'un côté, l'hétérogénéité thématique du domaine de la GIZC complexifie la tâche d'élaboration d'une ontologie, de l'autre, elle permet d'envisager la réutilisation de composants communs à d'autres domaines.

Pour intégrer les *éléments* (*concepts* et *relations*) issues de ces différentes briques et les *structurer* dans le respect du fonctionnement du *système de la zone côtière*, nous proposons d'élaborer un *modèle d'ontologie haut niveau* (voir partie 5.1.4) spécifique à la GIZC. De plus, nous souhaitons tenir compte du mode de représentation de la connaissance utilisé par les langages du Web sémantique. Les formalisations présentées utilisent cependant le langage *UML* compte tenu de son intérêt pour les entretiens auprès de nos partenaires et de l'existence de passerelles qui permettent si nécessaire la formalisation avec les langages précédents.

Cette proposition s'appuie sur la consultation des partenaires de notre projet ainsi que sur une *approche écosystémique*, fondamentale en GIZC. En effet, les recommandations pour la mise en œuvre de la GIZC mettent en avant l'intérêt de l'*approche écosystémique*. Or cette dernière partage les objectifs d'*inventaire des éléments* (respectivement *concepts* pour une ontologie)

du *système* (respectivement *domaine*) et de représentation de leurs *interactions* (respectivement *associations* ou *relations* entre *concepts*) et contribue donc largement à l'élaboration d'une *ontologie*.

5.2.1 Démarche collaborative

La *participation des acteurs* mobilisés dans notre projet était essentielle pour valider nos analyses ainsi que l'utilisation d'un *formalisme commun* qui soit à la fois compréhensible côté acteur et interprétable côté système.

Utilisation d'un formalisme commun pour représenter les connaissances : UML

Compte tenu des liens existants entre notre problématique et celle du Web sémantique, nous présentons brièvement comment la connaissance est représentée à l'aide des langages du Web sémantique avant de justifier l'intérêt et la compatibilité d'UML dans le contexte de la GIZC pour formaliser notre modèle.

Les formalisations de la connaissances avec les langages du Web sémantique Il existe plusieurs langages dérivés du langage *XML* pour le *Web sémantique* : *OWL*, *DAML*, *RDFS*, *RDF* [175] et globalement, tous permettent de traduire nos attentes techniques pour l'élaboration d'une ontologie.

Il existe, en effet, une filiation entre ces différents langages (voir Figure 5.17(a)). Tous dérivent du *RDF*, qui est une syntaxe *XML* dont l'objectif est de rendre *interprétable par la machine* certains concepts clés utilisés dans les pages Web. Plus précisément, il s'agit d'ajouter de la logique au Web en lui donnant la possibilité d'utiliser des règles pour faire des *inférences*. Pour cela *RDF* code le sens dans des ensembles de *triplets*¹⁵ (voir Figure 5.17(b)). Chaque élément du *triplet* jouant le rôle du *sujet*, du *verbe* et de l'*objet* dans une phrase élémentaire. Le *sujet* et l'*objet* sont identifiés par des *URI* (*Universal Resource Identifier*) ainsi que les *verbes*, ce qui permet à chacun de définir un nouveau *concept*, un nouveau *verbe*, en définissant simplement une *URI* quelque part sur le Web [22]. Les langages plus récents qui découlent de *RDF* conservent cette structuration de la connaissance sous forme de *triplet*, en particulier *OWL* utilisé par *FOS* (voir exemples de triplets tirés de *FOS* en Figure 5.17(c)).

L'utilisation d'*URI* permet en particulier de résoudre les problèmes de *polysémie* que les personnes tolèrent dans le *langage naturel* mais que ne tolère pas le *traitement automatique* des données.

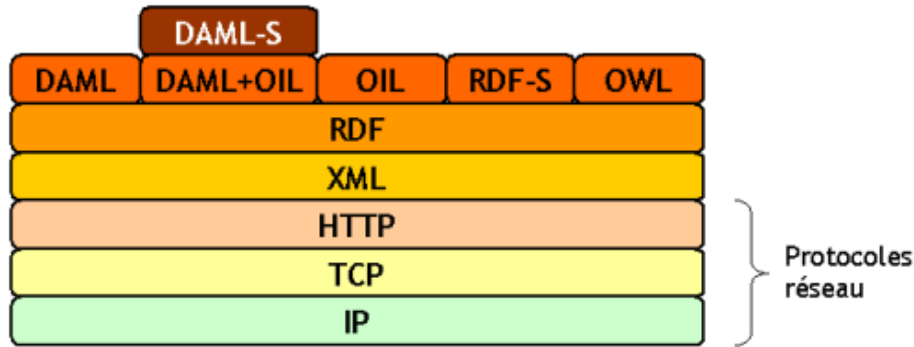
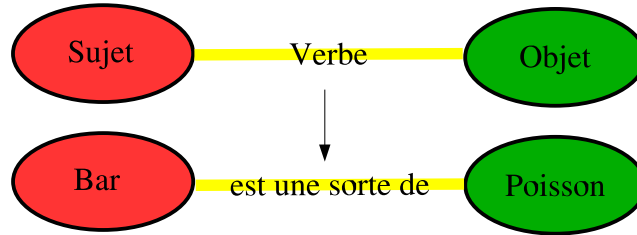
Ainsi, par ces langages, on peut exprimer des connaissances telles que :

- le *véliplanchisme* est pratiqué sur l'étang de Thau,
- le *véliplanchisme* est interdit sur l'étang de Palavas,
- les *pêcheurs du LR* pêchent du thon rouge, du merlu, du rouget....
- le *merlu* est une sorte de poisson.

Les règles d'*inférence* sont encore plus puissantes dans le cadre d'une *ontologie*. Elles permettent par exemple d'exprimer la règle suivante : "*si un espace est exploité par des ostréiculteurs, alors le nautisme y est nécessairement interdit*"¹⁶.

¹⁵On parle de *triple* en anglais.

¹⁶Par contre ce n'est pas forcément le cas pour l'aquaculture en général.

(a) Filiation entre les langages du *Web sémantique*

(b) Formalisation de la connaissance à l'aide de triplets

```

<owl:Class rdf:about="http://www.fao.org/aos/agrovoc/Image_Processing_at_ag">^M
  <rdfs:subClassOf rdf:resource="http://www.fao.org/aos/agrovoc/AGTOP"/>^M
  <rdfs:comment></rdfs:comment>^M
  <rdfs:label>Image_Processing_ag</rdfs:label>^M
</owl:Class>^M
<owl:Class rdf:about="http://www.fao.org/aos/ontologies/WordNet/wnat#GENOTYPE">^M
  <rdfs:comment></rdfs:comment>^M
  <rdfs:label>GENOTYPE</rdfs:label>^M
  <rdfs:subClassOf>^M
    <owl:Class rdf:about="http://www.fao.org/aos/ontologies/WordNet/wnat#BIOLOGICAL_GROUP"/>^M
  </rdfs:subClassOf>^M
</owl:Class>^M
<owl:Class rdf:about="http://www.fao.org/aos/asfa#Aquatic_plants_at_asfa">^M
  <rdfs:subClassOf rdf:resource="http://www.fao.org/aos/asfa#Aquatic_organisms_at_asfa"/>^M
  <rdfs:label>Aquatic_plants_asfa</rdfs:label>^M
  <rdfs:comment></rdfs:comment>^M
</owl:Class>^M

```

(c) Exemples de triplets *OWL* fournis par l'ontologie *FOS* [75]FIG. 5.17 – Les différents langages du *Web sémantique* 5.17(a), la formalisation élémentaire de la connaissance employée 5.17(b) et un extrait de l'ontologie *FOS* au format *OWL* 5.17(c)

Formalisation d'un modèle d'ontologie pour la GIZC avec UML Notre approche de mise en place d'une ontologie pour un service de métadonnées en ligne peut bénéficier des apports du *Web sémantique* et de ces syntaxes XML. Cependant, afin de dialoguer avec nos partenaires, le choix d'*UML* nous a semblé plus approprié sans pour autant compromettre, à terme, la génération d'une ontologie dans un de ces langages. On peut d'ailleurs remarquer la correspondance entre les Figures 5.16 et 5.17(b) où *concept* peut être lu comme *Sujet* ou *Objet* alors que *Relation* peut être lue comme *Verbe*. En effet, les langages du *Web sémantique* ne sont ni assez explicites ni assez intuitifs pour être manipulés par un utilisateur néophyte.

Toutefois, à l'aide d'*éditeurs d'ontologies* (comme *Protégé* ou *Ontolingua* par exemple), on peut passer outre la difficulté syntaxique de ces langages et rendre la génération de tels fichiers transparente pour les utilisateurs à l'aide d'*interfaces graphiques* adaptées [136]. *Protégé* est d'ailleurs souvent recommandé [162] pour sa facilité d'utilisation, relativement intuitive pour un néophyte et possède d'ailleurs un *plugin UML* qui permet de générer du code *RDF* ou *OWL* depuis une formalisation *UML*. Il faut cependant retenir les limites d'importation d'*UML* vers *OWL*, en particulier du point de vue de la traduction des *relations binaires*. Il existe deux approches pour contourner ce problème : on traduit la relation binaire soit comme une *propriété*, soit comme un *classe* [120, 62], ce qui reste finalement gérable compte tenu du fait que les traductions des *classes*, des relations de *généralisation / spécialisation* et des *attributs* sont compatibles.

De plus, le choix de l'implémentation des spécifications en *UML* nous rapproche de nombreux travaux de référence dans des domaines proches [75] et en particulier en accord avec les exemples d'ontologies que nous avons présentés, comme celui de la taxonomie détaillée dans l'*ISO 19115* et formalisé également en *UML* et ceux de *FOS* pourtant implémenté en *OWL*. Ceci donne de la lisibilité à nos modèles en dehors de notre groupe de travail. En outre, nous avons pour le moment choisi de traduire ces spécifications *UML* en *SQL*, dans le même *SGBDR* que notre service de métadonnées, ce qui simplifie la gestion de notre système et n'exclue pas à terme une éventuelle exportation dans une application utilisant une syntaxe *XML*.

Nous présentons désormais les modèles que nous avons pu plus facilement construire par une démarche collaborative grâce à l'utilisation d'*UML*.

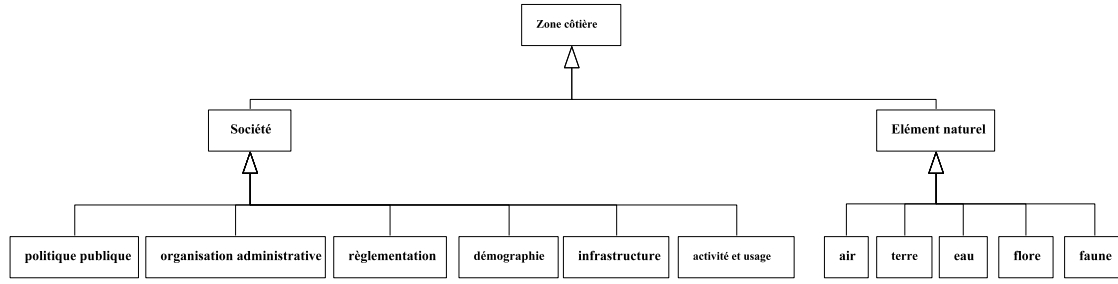
Modèles co-construits

Comme pour la mise en place du *service de métadonnées* proposé, nous avons souhaité impliquer les futurs utilisateurs du système dans cette démarche de mise en place d'un *référentiel sémantique*. Plus encore que lors de la mise en place du *profil* de l'*ISO 19115*, l'*aspect collaboratif* de cette démarche était nécessaire afin d'obtenir l'agrément des experts thématiques du domaine, condition nécessaire pour que l'outil soit utilisé et mis à jour à terme. En effet, pour l'élaboration du *profil Syscolag*, la place laissée à l'expertise du domaine s'est limitée à un avis sur les propositions de la norme, alors que pour la mise en place d'une *ontologie*, les acteurs se sentent plus directement sollicités et impliqués puisqu'il s'agit de retranscrire leur expertise et de compléter les lacunes de la norme en proposant des éléments de métadonnées plus spécifiques à leur domaine.

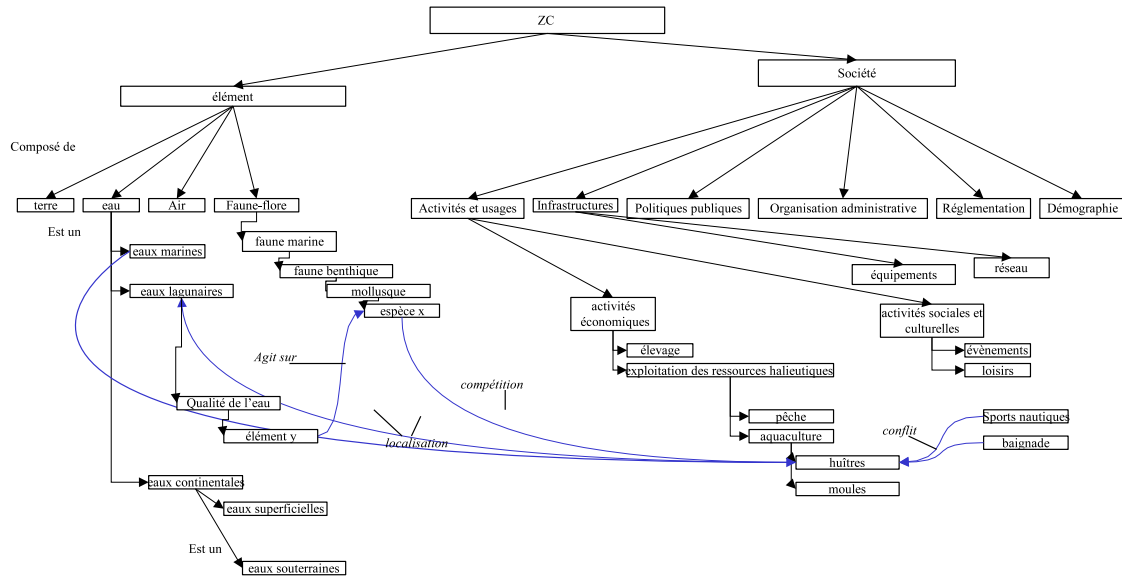
Pour cela, nous avons procédé de *manière itérative*, en faisant s'exprimer ou réagir les experts du domaine sur la base de *formalisations des discours* que nous avons analysés au fur et à mesure des *réunions* et *entretiens*.

Dans un premier temps, nous avons présenté une *catégorisation des composantes principales du domaine* que nous avons souhaité structurer par des diagrammes de classes *UML* en introduisant des liens de *généralisation / spécialisation*, voire d'*agrégation* (voir Figure 5.18(a)).

Cette proposition, bien qu'en accord avec les structurations de la littérature de ce domaine, en particulier les catégories principales mises en avant dans l'*approche écosystémique* (voir partie 5.2.2), a rapidement montré ses limites. En effet, par les seuls liens de *généralisation / spécialisation* ou d'*agrégation*, les experts du programme ont reproché une restitution trop figée d'*éléments de la zone côtière* qui sont dans la réalité en perpétuelle *interaction* [16]. Cette première proposition n'a donc pas obtenu de consensus dans l'enthousiasme mais plutôt par



(a) Catégories principales proposées par les partenaires de Syscolag pour la GIZC [18]



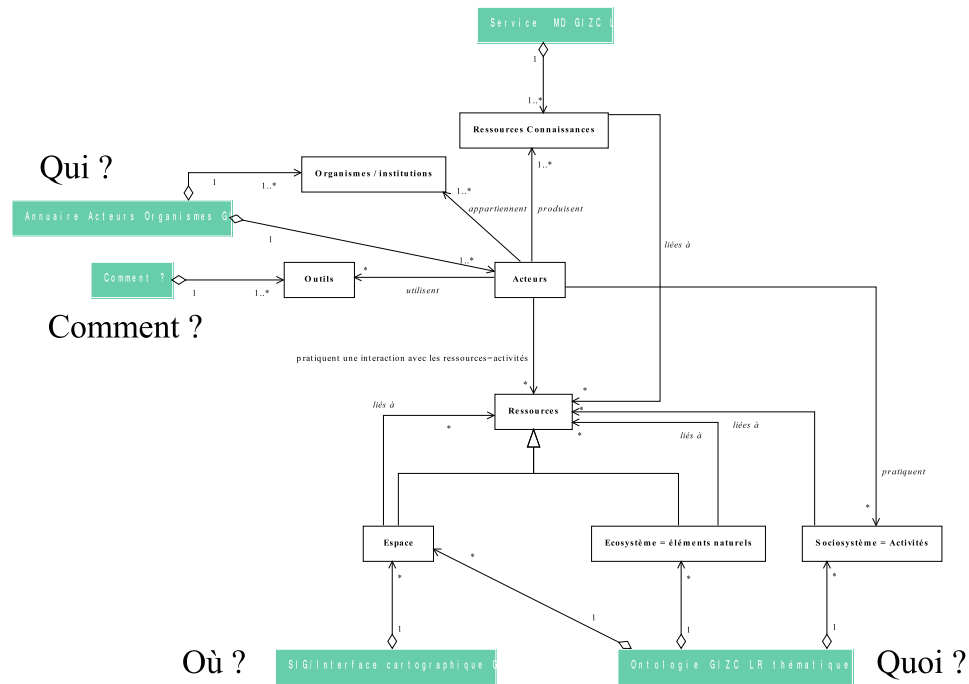
(b) Prise en compte des liens transversaux en complément de la hiérarchisation des concepts [15]

FIG. 5.18 – Premiers travaux de représentations des concepts et de leurs relations dans le domaine de la GIZC

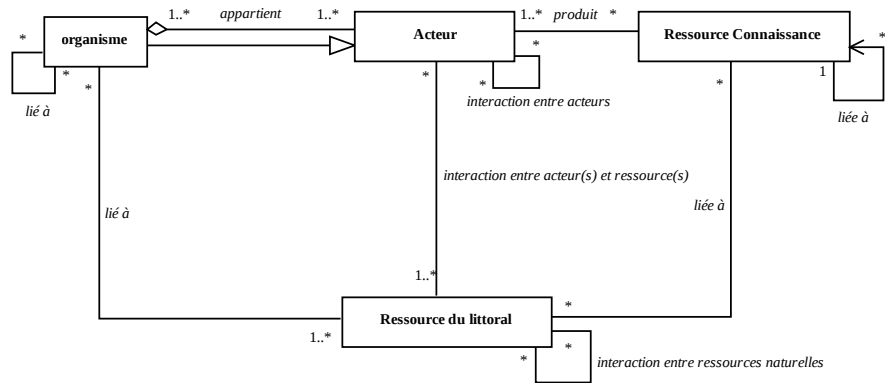
dépit, faute de meilleure proposition. Elle a néanmoins eu le mérite d'amorcer la démarche. Les diagrammes *UML* suivants ont recueilli et suscité beaucoup plus de remarques exploitables pour valider et affiner la modélisation (voir Figures 5.18(b) et 5.19(a)). En outre, il ne sont pas en rupture avec les formalisations précédentes et permettent de mieux gérer les liens qu'entretiennent les catégories générales. On remarque, au fur et à mesure de l'évolution des modèles qui ont été proposés, que ceux-ci sont plus génériques, en particulier la place accordée à la modélisation des interactions entre les éléments de la zone côtière, absente dans le premier modèle (Figure 5.18(a)) apparaît de plus en plus clairement.

Finalement l'évolution chronologique de ces modèles et la prise en compte des approches de référence dans le domaine nous a permis de formaliser un modèle qui intègre les différents points de vue à un niveau plus générique (voir Figure 5.19(b)).

Avant de détailler le modèle d'ontologie que nous proposons pour la GIZC et afin de pouvoir en apprécier le contenu, nous présentons, l'intérêt complémentaire de l'*approche écosystémique*. En effet, nous avons jusqu'à présent discuté de la mise en forme de la connaissance par le choix



(a) Diagramme intermédiaire de synthèse pour la structuration des concepts de notre domaine



(b) Diagramme intermédiaire présenté aux partenaires du projet Syscolag (formalisme UML)

FIG. 5.19 – Evolution des modèles validés avec les partenaires

du formalisme, des relations standardisées et de leurs syntaxes. Il est important de prendre en compte le complément nécessaire à cette approche qui est la nature thématique des classes et des relations que nous allons manipuler, qui sont propres à notre domaine et sur lesquelles va s'appliquer cette *mise en forme* : ie le *fond* de l'ontologie par opposition à la *forme*.

5.2.2 Intérêt de l'approche écosystémique

L'*approche écosystémique* entretient des points communs avec les *ontologies* pour l'inventaire des *éléments* de la zone côtière et de leurs *interactions* : ses applications au domaine de la GIZC sont d'autant plus intéressantes à prendre en compte qu'il n'existe pas véritablement d'*ontologie* dédiée au domaine. Souvent, on la retrouve sous-jacente dans les outils de contrôle du vocabulaire existants comme méthodologie pour le recueil des concepts ainsi que leur structuration.

Pour répondre au souci d'intégration de l'information pertinente, les principes de mise en œuvre du *développement durable* préconisent de s'appuyer sur une *représentation partagée du fonctionnement du système étudié* (voir partie 2.1.1). Les guides méthodologiques pour la GIZC [87, 170, 143], conscients de la nécessité d'une démarche commune et partagée pour l'élaboration de *SI*, prônent l'utilisation de *cadres de référence* ou *référentiels*, par des recommandations qui sont des catégorisations standardisées des composantes principales de la zone côtière, afin de procéder méthodiquement à un *inventaire* des données existantes sur les *plans thématique et spatial* [18].

En particulier, le guide [170] indique en introduction que "*pour gérer et aménager les zones côtières de manière durable, les décideurs doivent comprendre comment s'imbriquent l'environnement naturel et les activités humaines pour former un éco-sociosystème[...]*".

Il s'agit donc, au cas par cas, pour une zone donnée, d'appréhender l'ensemble de ses caractéristiques *physiques* et *humaines* et de les *présenter de façon appropriée* (voir partie 2.1.1 et Figure 5.20(a)) pour mieux appliquer les décisions de gestion. Cependant, si les approches recommandées apportent une méthodologie pour inventorier les *éléments* du système et leurs *interactions*, la représentation de l'information utilisée est inadaptée à une exploitation par les machines.

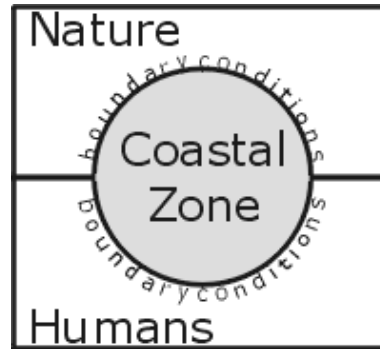
Le schéma de [96] présenté en Figure 5.20(b), célèbre dans la communauté mobilisée dans la GIZC, va dans ce sens dès 1985. Il illustre une zone côtière classique avec un *inventaire non exhaustif d'éléments et d'activités générales* et constitue un premier effort pour partager une *trame* ou une *méthodologie commune* afin de catégoriser l'*espace* et les *activités* du littoral. Avec l'apparition d'outils tels que les *SGBD*, les *SIG*, la *télédétection* et plus généralement les *NTIC*, ces initiatives peuvent désormais bénéficier d'une *mise en forme standardisée* et d'une *accessibilité* simplifiée.

Cependant, si beaucoup de ces analyses prospectives se recourent, il subsiste encore trop de divergences (moins sur le *fond* que sur la *forme*) pour considérer qu'il existe des méthodes communes de recueil d'information et de connaissance nécessaires à une bonne lisibilité de chaque problématique de GIZC présentée.

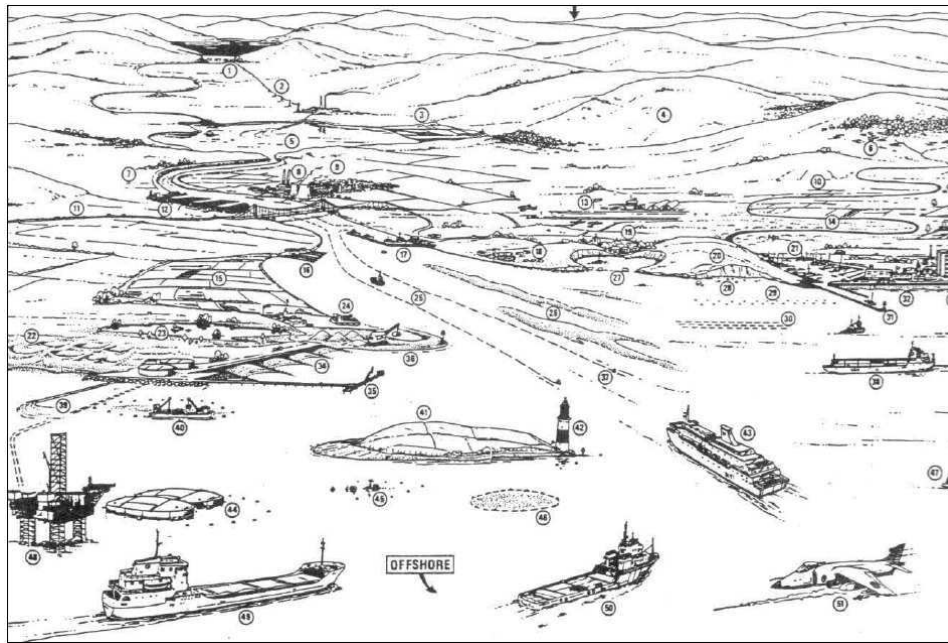
Toutefois, dans la littérature, de nombreux avis convergent pour reconnaître l'intérêt de l'*approche écosystémique* à ces fins [87, 170, 143, 171, 33].

Cette approche transversale est intuitive, on la retrouve de manière sous-jacente dans le discours scientifique, comme celui d'*Anita Conti* présenté en partie 2.1.1. Il s'agit de représenter la zone côtière comme un *système*, milieu complexe où de nombreux *objets* (*éléments* ou encore *items*) physiques (les hommes, l'air, l'eau, la terre, le vivant ...) et non physiques (organisations, lois, etc.) cohabitent et *interagissent*.

Dans la partie que réserve le guide méthodologique [143] à l'introduction à l'approche écosystémique, il est écrit ceci : "[...] *Le système est un outil pratique servant à décrire des situations complexes du monde réel et à analyser des problèmes complexes. Par ailleurs, l'approche systémique est un outil utile pour :*



(a) Approche écosystémique pour la structuration des composantes principales de la zone côtière [33]



(b) Un schéma générique de zone côtière avec des éléments et activités caractéristiques [96]

FIG. 5.20 – La vision binaire de l’approche écosystémique opposant *sphères humaine et naturelle* 5.20(a) et un premier travail de catégorisation et de spatialisation des *éléments et activités* liés à au *développement humain* sur la zone côtière 5.20(b)

- visualiser la structure du monde réel comme première étape dans une présentation SIG,
- schématiser les processus se déroulant dans le monde réel comme base pour simuler le système au moyen de modèles, et
- étudier l’interaction dynamique entre l’homme et la nature.

La théorie des systèmes est basée sur l’idée que **le monde réel peut être décrit par le biais d’un ensemble d’éléments et d’interactions entre ces éléments**. En général, le système ne concerne qu’une petite partie de l’univers, notamment la partie comprise dans les limites

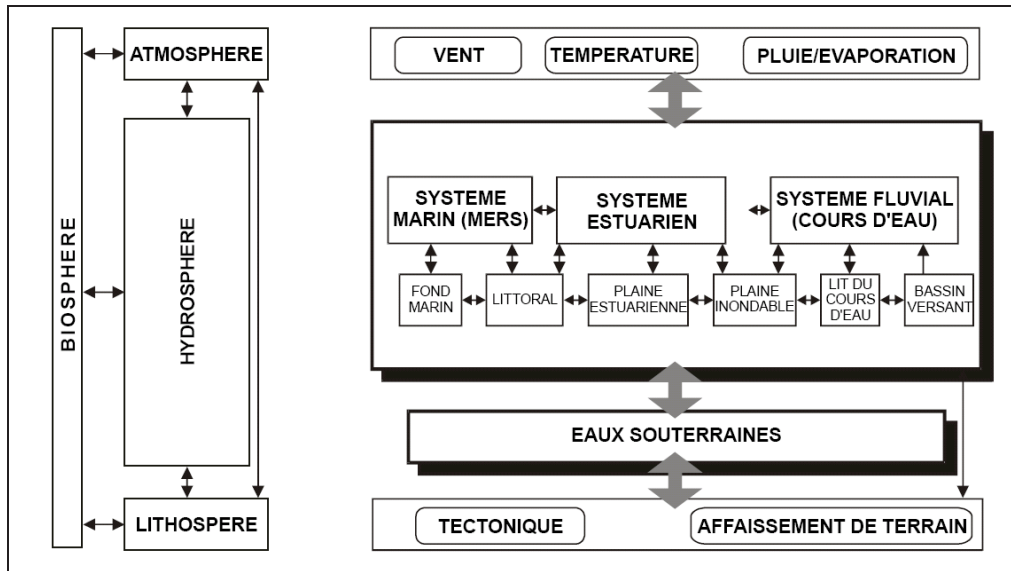


FIG. 5.21 – Approche écosystémique pour la structuration des composantes principales de la zone côtière [143]

du système. La zone au delà de ces limites est appelée *l'environnement du système*. [...]"

Si la figure 5.20(a) propose un schéma qui illustre cette approche au niveau le plus général [96], en distinguant les *sous-systèmes humains et naturels*, la figure 5.21 propose une représentation d'*interactions entre éléments* plus spécialisés (pour les paramètres physiques de la zone côtière). Pour citer [143] : *les flèches de la Figure 5.21 indiquent une diversité d'interactions à petites et moyennes échelles entre les composantes du système*. On retrouve bien dans cette dernière représentation certaines des catégories principales présentes dans le diagramme de synthèse du groupe de travail *Syscolag* (le découpage du système physique en *hydrosphère*, *atmosphère*, *lithosphère*, *biosphère* qui correspondent respectivement à *eau*, *air*, *terre*, *faune* et *flore*). Ceci rend un peu plus évident la complémentarité entre ces deux démarches.

Comme le note [33], un bon programme de GIZC peut être obtenu à travers une *compréhension profonde de ces objets (items) et de leurs relations*.

On peut donc établir un parallèle entre l'*approche écosystémique* qui a pour objectif de catégoriser les composantes essentielles du système de la GIZC et s'appuie sur des *éléments en interactions* (imbriqués, associés dans la réalisation de processus de différentes sortes...) et l'effort d'élaboration d'une *ontologie*. Toutefois, si l'*approche écosystémique* convient parfaitement pour inventorier *éléments* et *interactions* de la zone côtière, l'ontologie apporte le cadre formel, la syntaxe nécessaire à la mise en forme des concepts inventoriés. En effet, les schémas que nous présentons posent encore un réel problème d'interprétation entre les différentes classes et associations représentées. Ces deux approches sont donc complémentaires, l'une apportant le *fond* et l'autre la *mise en forme*, et indispensables pour un partage efficace des connaissances.

Compte tenu, d'une part, des analyses formalisées à la suite d'*entretiens* avec nos partenaires et d'autre part, de l'*approche écosystémique* qui fait référence dans le domaine de la GIZC et qui rejoint par certains de ses objectifs ceux liés à la mise en place d'une *ontologie*, nous proposons un nouveau *modèle* qui combine ces deux approches.

5.2.3 Présentation du modèle d'ontologie

Conformément à ce qui a été présenté auparavant, nous proposons de distinguer quatre principales *classes* (ou *concepts*, voir partie 5.1.4) pour valuer les *mots clés* (voir Figure 5.22). Ce sont :

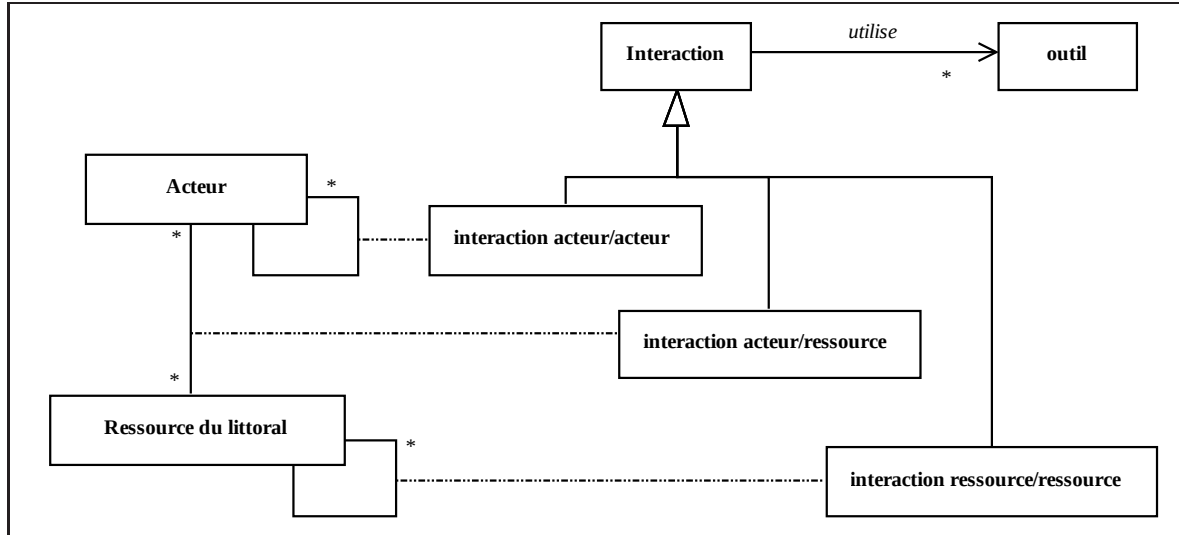


FIG. 5.22 – Diagramme finalisé du modèle de l'ontologie [18]

1. la classe **Acteur** (pêcheur, scientifique...),
2. la classe **Ressource du littoral** (huître, récif artificiel...),
3. la classe **Interaction** (pêche, conflit d'usage...),
4. la classe **Outil** (chalut, SCOT...),

Dans le modèle que nous proposons, ces quatre classes entretiennent des relations qui expriment le fonctionnement général de la zone côtière selon une approche qui tient compte des principes du Développement Durable et de la GIZC. Pour cela, le modèle proposé s'appuie sur l'approche écosystémique :

- La classe *Acteur* représente la *sphère anthropique* (les humains du système voir Figure 5.20(a)) opposée à la *nature / sphère naturelle* dans l'approche écosystémique et représentée ici par la classe *Ressource du littoral* (dans laquelle on distinguera, en pratique, les *ressources naturelles* des *ressources anthropiques* ou *artificielles*). Ces deux classes permettent de décrire les différents *éléments du système* (aspect *statique*).
- La classe *Interaction* représente les *processus* entre les *éléments* de la zone côtière (*acteur* et *ressource du littoral*). Elle permet d'exprimer l'aspect *dynamique* du système à travers la mise en relation de ses *éléments statiques*,
- La classe *outil* exprime le fait que dans le fonctionnement général du système, certains *éléments* sont nécessaires à la réalisation d'*interactions*. Par exemple un *chalut* est un *outil* au service de la réalisation de l'*activité humaine* de *pêche en mer* (sorte d'*interaction*).

La Figure 5.22 formalise ces analyses et exprime la nature des associations entre ces classes.

Une lecture générale de ce diagramme pourrait être : "*Dans le système de la zone côtière, des **Acteurs interagissent** entre eux ou avec des **Ressources du littoral** (elles-mêmes en interaction) en employant éventuellement des **outils***".

Ce modèle est donc *ouvert* pour décrire des *connaissances élémentaires* sur la GIZC et suffisamment *générique* pour laisser la place à la *représentation de connaissances plurithématiques* comme cela est nécessaire dans ce domaine. C'est en spécialisant ces classes les plus générales que nous allons retrouver les classes retenues par les partenaires du programme *Syscolag* (voir Figures 5.18 et 5.19).

On peut constater que la *classe centrale de ce modèle* est la classe *Interaction* qui est à la croisée des associations entre concepts. Nous discutons par la suite des perspectives de représentation des connaissances par le biais de cette unique classe puisque la réalisation d'une *interaction* suppose une dynamique entre *deux éléments* (voire *trois* lorsqu'un *outil* intervient dans l'*interaction*). Ainsi, *en décrivant l'ensemble des interactions de la zone côtière, on est amené à décrire l'ensemble des éléments qui agissent sur le système*.

Ceci laisse supposer que l'essentiel de la connaissance sur la GIZC est implicitement représenté dans cette classe. En outre, *interaction*, en tant que *classe d'association UML*, permet de qualifier la nature /sémantique de toutes sortes d'associations entre *éléments*. En d'autres termes, la classe *interaction* permet d'instancier une relation **RTS** (voir partie 5.1.4, Figure 5.11).

Remarques :

- Le modèle d'ontologie présenté est la version la plus générique de l'ontologie. A ce niveau, il n'est plus propre au domaine thématique de la GIZC mais peut tout autant s'appliquer à des problèmes de gestion de n'importe quel autre "*écosystème anthropisé*". Nous pouvons donc lui attribuer le qualificatif d'*ontologie top-level* ou *haut niveau* [70] (voir partie 5.1.4).
- Ce modèle n'est pas en rupture avec le mode de formalisation des connaissances utilisés par les langages du *Web sémantique*. En effet, il y a bien dans le modèle proposé une formalisation sur le mode de *triplet* (*sujet, verbe, objet*) où le *sujet* et l'*objet* peuvent être un *acteur*, une *ressource* ou une *interaction* alors que le *verbe* correspond forcément à une *interaction*.

Pour terminer, en réponse à l'importance de l'*espace* dans la GIZC, nous devons positionner ce *concept* dans le *modèle d'ontologie* proposé.

5.2.4 Spatialisation des concepts de l'ontologie

Compte tenu de l'importance de l'*espace* dans la GIZC (voir parties 2.1.2 et 3.1.4) nous proposons d'introduire brièvement la place de l'*espace* dans le *modèle d'ontologie* proposé. Nous reviendrons largement sur ce point dans le chapitre 6 puisqu'il lui est entièrement consacré.

Nous avons vu que l'*ISO 19115* propose des descripteurs élaborés pour décrire l'*étendue géographique concernée par les ressources informationnelles* qui font l'objet d'une *fiche de métadonnées*. Nous avons choisi de rendre cette *description spatiale des ressources informationnelles* obligatoire (voir partie 4.3.1). Comme une *ressource informationnelle*, les *éléments* et *interactions* de la zone côtière peuvent être décrits spatialement.

En effet, dans le *modèle d'ontologie* proposé en partie 5.2.3, on peut tout aussi bien décrire l'*étendue géographique concernée par* :

- un *Acteur* (*zone d'étude, zone d'activité, zone de loisirs...*),
- une *Ressource* (une *lagune*, une *forêt...*),

- une *Interaction* (un courant marin, un espace alloué à une activité ou un espace associé à l'écologie d'une espèce animale ou végétale...),
- *outil* (une station d'observation, un SCOT...).

Ceci revient finalement à dire que les *concepts* de notre *modèle d'ontologie* sont *spatialisables* ou encore qu'il est possible de leur attribuer des *propriétés* valuées par un *type de donnée spatial*. Dans le modèle d'ontologie proposé en Figure 5.22, la présence de l'espace est donc implicite. Les *descripteurs spatiaux* sont potentiellement dispersés dans chacun des *concepts* de l'*ontologie*. Il est, par exemple, possible de formaliser ceci de la manière présentée en Figure 5.23 en *réifiant* une *classe espace* et en indiquant qu'elle correspond à un *type de donnée* (*Data Type* dans l'*ISO 19115*) associé à différentes *propriétés* potentiellement présentes dans chaque *concept* et nommées *Description spatiale*.

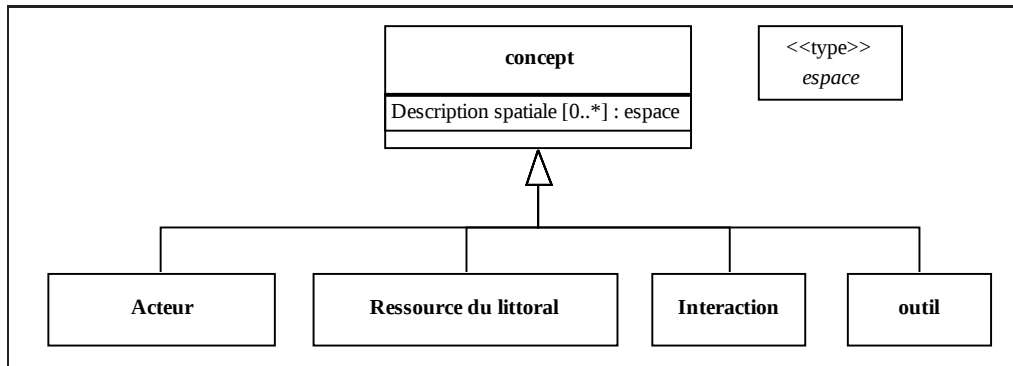


FIG. 5.23 – Spatialisation des concepts du modèle d'ontologie

De cette manière, la *propriété* de *description spatiale* est alors *héritée* dans toutes les sortes de *concepts*. Nous verrons par la suite que nous utiliserons ces *descriptions spatiales* pour proposer la mise en place d'un *référentiel spatial*, voir chapitre 6.

Synthèse de la section 5.2

- Les objectifs de *contrôle du vocabulaire* et de *partage des connaissances* sont optimisés par l'usage d'une *ontologie* et de langages adaptés. Nous proposons le formalisme *UML* pour la GIZC, compte tenu de la qualité du dialogue qu'il permet avec les *acteurs* et de l'existence de passerelles avec les langages du *Web sémantique*.
- Afin de proposer un *modèle d'ontologie* adapté spécifiquement à la GIZC, nous nous appuyons sur une démarche collaborative avec nos partenaires pour analyser et formaliser leurs attentes. De manière itérative, les formalisations successives leur ont été présentées puis ont été affinées et validées de manière consensuelle.
- La prise en compte de l'*approche écosystémique*, qui fait référence pour la GIZC, est très pertinente pour la mise en place d'une *ontologie*. En effet, cette *approche* est particulièrement intéressante pour *inventorier* et *structurer* les *éléments* et *interactions* du *système* de la zone côtière.

- La synthèse de ces deux approches nous a finalement permis de proposer un *modèle générique d'ontologie* qui répond aux besoins spécifiques de la GIZC et qui est compatible avec le mode de formalisation de la connaissance utilisé par les langages du Web sémantique.
- Le modèle proposé permet de représenter la nature des *interactions* essentielles dans la GIZC entre les *éléments* principaux de la zone côtière que sont les *acteurs* et les *ressources* du littoral. En outre, nous permettons de *spatialiser* ces *éléments*.

Pour la réalisation concrète de notre *ontologie*, il nous faut à présent *inventorier les différents concepts* appartenant à ces quatre catégories générales, *les structurer par l'introduction des différents types de relations ontologiques* présentés en Figure 5.11, leur *attribuer des propriétés* (en particulier des *descriptions spatiales*) et finalement *les instancier*.

C'est par la *spécialisation* (au sens *UML*) de ces classes que notre modèle devient vraiment *spécifique* à notre domaine thématique et géographique d'application : "*la gestion intégrée du littoral de la région Languedoc Roussillon*". Ceci est l'objet de la section 5.3 dans laquelle nous exposons quelques exemples de *typologies* obtenues par la *spécialisation* des *principales classes* du modèle ainsi que l'attribution de propriétés. Cette étape est nécessaire pour pouvoir ultérieurement instancier ce modèle par la représentation de liens transversaux (*RTS*) exprimant les *interactions* entre ces classes. Pour l'instant nous nous limitons à une structuration simple des concepts afin de permettre de comprendre l'utilisation d'un tel modèle. Ceci permettra, à terme, l'intégration de connaissances plus complexes du domaine.

5.3 Typologies –spécialisations– des classes du modèle d'ontologie

A présent, nous allons *spécialiser les classes de notre modèle d'ontologie* de la section 5.2.3 en fonction de notre domaine d'application : la *Gestion Intégrée de la Zone Côtière en Languedoc Roussillon*. La validation de la pertinence thématique des schémas qui suivent n'a pas été faite auprès de nos partenaires. Nous nous appuyons cependant sur les *typologies des sources de références du domaine* et sur les résultats des échanges avec la communauté *Syscolag*. De plus, nous avons vu que les niveaux les plus élevés (les plus importants dans le contexte pluridisciplinaire de la GIZC) ont déjà été validés. Il s'agit donc plus, dans cette section, de décrire la démarche suivie et son intérêt pour résoudre les objectifs visés (contrôle du vocabulaire et intégration de la connaissance à l'aide d'une méthode pour alimenter l'ontologie) que d'évaluer la qualité thématique de son contenu (cette tâche revenant, à terme, aux experts du domaine).

5.3.1 Méthodologie générale

Il s'agit de *spécialiser les classes les plus génériques* sous forme de *typologies* (ou *arborescences*) conduisant jusqu'aux éventuelles *instances* des *classes filles* (ie *feuilles*).

Les classes plus spécifiques que nous présentons proviennent de *références bibliographiques* ou d'*entretiens auprès d'acteurs scientifiques du domaine* au cours desquels des catégories ont émergé dans le discours. Nous les avons formalisées par la suite en *UML* pour permettre l'*implémentation de notre ontologie en tant que complément du service de métadonnées* ou

comme base de connaissance¹⁷.

Il est important de rappeler que, dans le processus de co-construction d'un *vocabulaire commun* dans un *contexte pluridisciplinaire*, la phase principale du point de vue de la *collaboration* est la définition des catégories les plus générales, conformément à la démarche générale prônée par Fonseca [70] et adoptée, par exemple, pour l'élaboration du thésaurus *GEMET* ou de l'ontologie *FOS*. En effet, à mesure qu'on spécialise les classes, les connaissances thématiques nécessaires s'affinent et le débat devient alors réservé aux initiés.

Cela pose un problème de définition des limites de la GIZC en terme de finesse nécessaire mais suffisante pour exprimer et échanger les connaissances transversales. Il est clair que l'objectif de la GIZC n'est pas le partage des connaissances les plus fines dans chaque discipline mais bien la prise en compte de certains éléments, plus généraux et plus compréhensibles, considérés comme fondamentaux du fait de l'importance de leurs *interactions* dans le système côtier. A ce titre, le nombre de niveaux maximum pour la qualité du modèle est encore inconnu et il est probable que seul l'usage apportera une réponse concrète à cette question.

Ainsi, il nous a semblé opportun de nous entendre prioritairement et consensuellement sur les premiers niveaux (les plus *hauts*) puis, en procédant selon une *approche top-down*, [16] d'affiner ces classes générales par des entretiens avec les spécialistes du domaine [15, 91] ou par la formalisation en *UML* des sources de références dans les domaines concernés, ceci devant conduire à l'élaboration d'*ontologies de domaine* [71] (voir Figure 5.24).

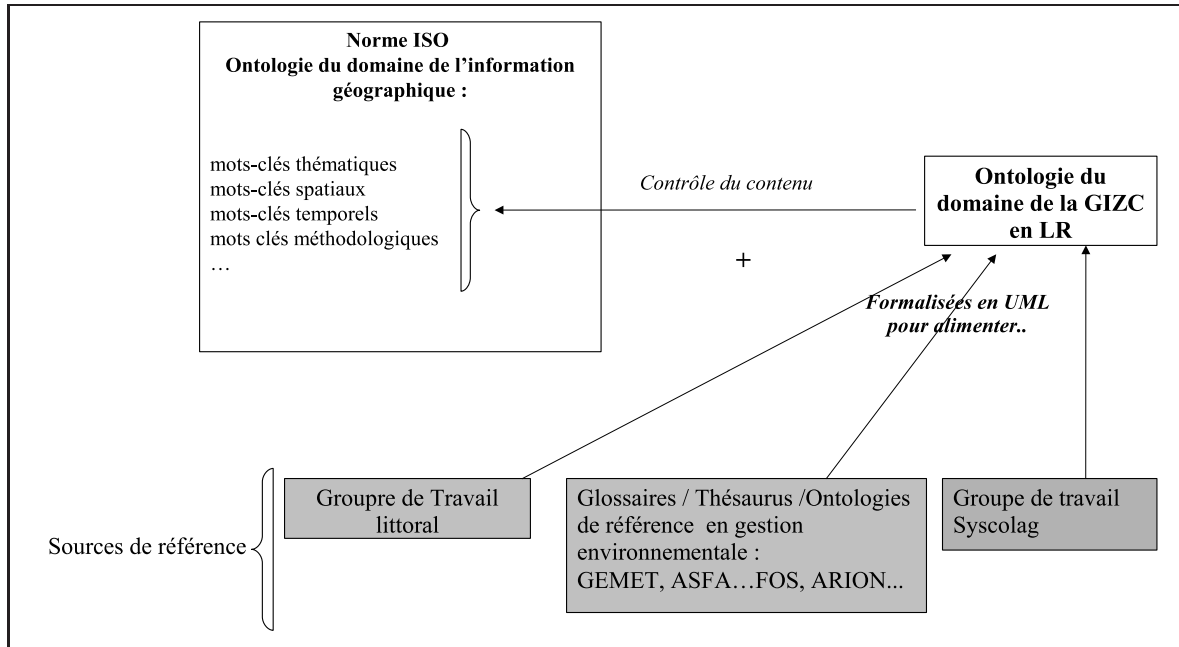


FIG. 5.24 – Complémentarité entre normes ISO et vocabulaires contrôlés [15]

En outre, priorité a été donnée à la structuration des concepts par des *liens de généralisation / spécialisation* et d'*agrégation* (respectivement *BTG*, *NTG*, *BTP* et *NTP*) pour les raisons suivantes : ces types de liens sont les plus employés par les utilisateurs, souvent de manière intuitive. De plus, un des objectifs principaux de notre ontologie est de permettre le

¹⁷Le terme *base de connaissance* est entendu ici au sens ontologique voir partie 5.1.4.

contrôle du vocabulaire par une navigation aisée dans l'inventaire réalisé, ce qui se fera par un système d'*arborescence hiérarchique* (*générique* ou *hiérarchie agrégative*) basé sur l'exploitation de ces types de liens.

Pour ce qui concerne les définitions et synonymes possibles pour chaque concept (respectivement *SN* et *UF*), nous préconisons de nous référer aux sources de références existantes (voir Figure 5.24) en y choisissant celles et ceux qui représentent au mieux la perception des acteurs mobilisés dans la GIZC. Il est possible pour cela d'utiliser l'état de l'art de ce chapitre sur les glossaires, thésaurus et ontologies de référence et d'intérêt pour notre application. Cette étape, par son ampleur et sa complexité, reste à l'appréciation des experts du domaine et devrait se réaliser à l'usage par la manipulation des applications qui présentent en ligne le contenu du *référentiel sémantique* (voir chapitre 7), à l'image de l'approche préconisée par [141, 177].

L'identification des relations de type *RTS* pourra alors s'appuyer sur ces inventaires en se limitant à des sous-domaines d'application d'intérêt prioritaire pour la GIZC en LR et formalisés dans des scénarios. Un scénario pourrait, par exemple, porter sur *la contamination d'une lagune*.

Nous avons vu également que les *ontologies* permettent l'attribution de *propriétés* aux *concepts* inventoriés qui sont évaluées lors de leur *instanciation* (voir parties 5.1.4 et 5.1.4). Nous proposons, sans chercher à être exhaustif, quelques exemples d'attributs possibles (suggérés par nos analyses) pour les classes que nous présentons par la suite et montrons que certains des *concepts* du *modèle d'ontologie* proposé peuvent partager des *propriétés* avec les classes proposées par l'*ISO 19115* (*acteur*, *Objet géographique...*) alors que d'autres restent plus spécifiques de notre domaine.

5.3.2 Les Acteurs de la GIZC en LR

La connaissance de la diversité des *acteurs* mobilisés dans la GIZC est un point clé pour la GIZC (voir partie 3.1.1). Elle permet de restituer les différentes classes d'*acteurs* intervenants sur une zone côtière donnée. L'instanciation des classes les plus spécifiques permet la constitution d'un *annuaire électronique* sur les *individus* et *organismes* (qui peuvent être associés à certaines *activités* et *usages* sur la zone côtière conformément à notre *modèle d'ontologie*).

Typologie des Acteurs

En gestion intégrée, le concept *acteur* ne distingue pas *acteurs personnes physiques* ou *individus* (*gestionnaire*, *scientifiques...*) ou les *acteurs personnes morales* (*entreprise*, *association*, *collectivité territoriale*, *centre de recherche*, *services de l'Etat...*). Nous nous conformons donc à cette représentation.

Typologie des Acteurs Individus Nous nous sommes appuyés sur les *typologies de référence* pour la GIZC, comme celles présentées en partie 2.1.2 et en Figure 5.25(a). On peut remarquer, comme cela avait déjà été soulevé, les limites de cette figure en terme de formalisme et d'interprétation par les utilisateurs (la nature des liens entre les différentes classes étant équivoque). A l'inverse, la Figure 5.25(b) présente une formalisation *UML* qui tente de faciliter la lisibilité de cet inventaire en distinguant deux sortes d'*acteurs* : les *individus* (personnes physiques) et les *organismes* (personnes morales : entreprise, association, collectivité territoriale, centre de recherche, services de l'Etat...). Les *organismes* y sont présentés comme des *agréats* d'*acteurs* de types non homogènes.

En ce qui concerne les *individus*, nous proposons des distinctions sommaires visant à différencier les principales catégories. Il est toutefois souhaitable, à terme, de *spécialiser ces typologies* de manière plus fine en intégrant d'autres *sources de référence*. Par exemple, pour les *acteurs scientifiques*, il peut être intéressant de faire des recherches sur des catégories *disciplinaires* : *géologue, physicien, botaniste...*

Nous détaillons en outre la typologie de la classe *organisme* pour l'intérêt qu'elle présente en complément de la classe *acteur*, ainsi que son importance en vue de la constitution d'un *annuaire électronique* pour la GIZC.

Typologie des Organismes En général, un *organisme* contient plusieurs *acteurs* (voir Figure 5.25(b)). Cette classe est donc *a priori* plus aisément maîtrisable que la classe *acteur* pour en contrôler l'exhaustivité lors de l'*instanciation*, puisque l'effectif des *organismes* est plus faible et moins fluctuant, en particulier à l'échelle régionale. En outre, le nombre d'organismes impliqués au niveau régional dans notre domaine est *a priori* suffisamment faible pour ne pas trop spécialiser cette typologie.

Nous pouvons là encore nous appuyer sur la Figure 5.25(a) pour proposer des classes pertinentes. La Figure 5.25(c) propose une *typologie* de classes générales d'*organismes*.

A chacune des *classes* inventoriées il est également possible d'attribuer des *propriétés*, pour cela nous pouvons nous appuyer en particulier sur les recommandations l'*ISO 19115*.

Propriétés de la classe Acteur

Pour décrire les **attributs** de notre classe *Acteur*, l'*ISO 19115* propose cinq classes liées qui établissent un inventaire déjà conséquent de *propriétés* pour décrire des *acteurs* (voir Figure 5.26); il est toutefois possible d'en ajouter de nouvelles jugées pertinentes dans le cadre de *Syscolag*.

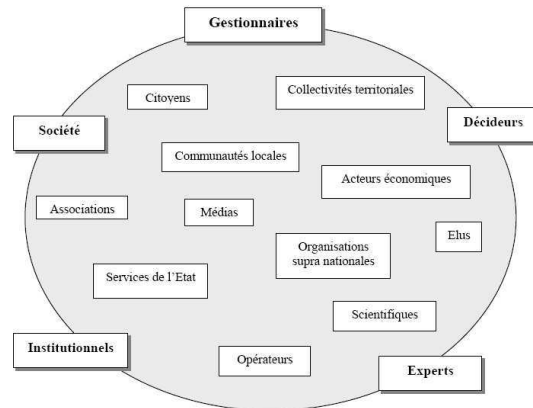
Propriétés de la classe Acteur Individu Il est par exemple possible d'attribuer à un *acteur* physique une *propriété* qui lui associe un *organisme* (éventuellement plusieurs). Ceci permet de gérer le lien d'*agrégation* proposé en Figure 5.25(b). Les attributs des classes de la Figure 5.26(a) peuvent être répartis entre les classes *Individu* et *Organisme*. Les adresses physiques des *acteurs* étant alors en partie déduites des *organismes* dans lesquels ils travaillent, elles n'ont donc plus à figurer explicitement dans la classe *acteur*.

Il paraît également pertinent d'ajouter à la classe *Acteur* deux *propriétés* qui caractérisent ses *compétences thématiques* par des *concepts* issus de l'*ontologie* ainsi que ses *zones d'études* (descriptions spatiales représentées par des *objets géographiques* gérés par le biais de l'*ontologie spatiale* (voir partie 6.1.2).

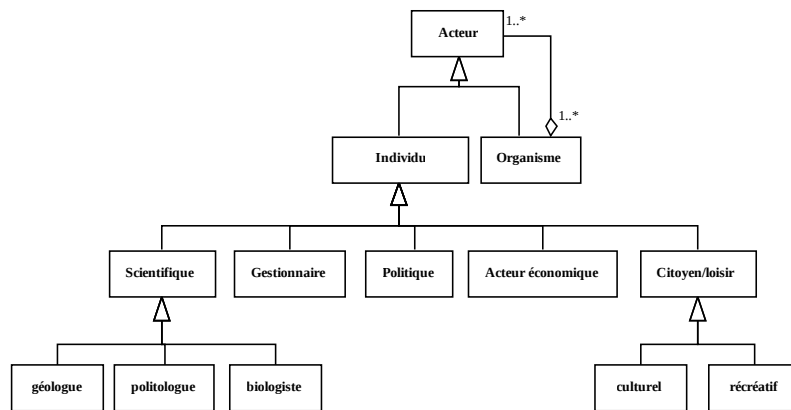
Propriétés de la classe Organisme Là encore, une partie des attributs possibles pour cette classe provient des *Data Type Responsible Party* et *Contact* de l'*ISO 19115* présentés en Figure 5.26(a). Une partie des *propriétés* est apportée par l'*ISO 19115*, l'autre est à apporter par le biais des *sources de référence* et du projet *Syscolag*.

D'après ces recommandations, nous proposons donc d'utiliser ces *éléments de métadonnées* comme *attributs* de la classe *Acteur*, tel que cela est présenté en Figures 5.26(b) et 5.26(c).

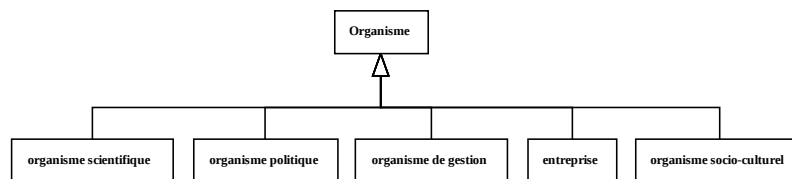
Finalement, pour restituer le réseau d'*acteurs* local, il s'agit concrètement d'*instancier* les classes qui spécialisent *Individu* et *Organisme*. Ces instances gérées au sein d'une application



(a) "La ronde des acteurs" dans la GIZC d'après [87]



(b) Diagramme des classes relatives aux *acteurs* dans l'ISO 19115



(c) Détail sur une typologie possible pour les organismes

FIG. 5.25 – Traduction des recommandations de la GIZC 5.25(a) en typologies d'acteurs (individus ou organismes) utilisant le formalisme UML : 5.25(b) et 5.25(c)

indépendante correspondent concrètement au contenu de l'*annuaire électronique* que nous avons présenté en parties 3.1.1 et 4.3.1 (voir détail sur l'utilisation en partie 7.2.1).

5.3.3 Les Ressources du littoral

Cette classe est importante pour décrire tous *éléments* de la zone côtière (autres que les *acteurs*). Ces ressources sont donc de deux sortes : des *éléments naturels* ou des *éléments artificiels* (ie d'origine *anthropique*). Cependant, il n'existe pas de terminologie de référence propre à la GIZC. Néanmoins, il est possible de réutiliser les terminologies de référence dans

CL_ResponsibleParty nom de la personne/individualName [0..1] : string nom de l'organisme/organisationName [0..1] : string position/poste du contact/positionName [0..1] : string informations sur le contact/contactInfo [0..1] : CL_Contact rôle du contact/role : Code	CL_Address adresse/deliveryPoint [0..*] : string ville/city [0..1] : string code postale/PostalCode [0..1] : string pays/country [0..1] : string adresse électronique/electronicMailAddress [0..*] : string aire administrative US/administrativeArea [0..1] : string
CL_Contact téléphone/phone [0..1] : CL_Telephone adresse/address [0..1] : CL_Address ressource en ligne/onlineResource [0..1] : CL_OnlineResource instructions pour le contact/contactInstructions [0..1] : string heures de service/hoursOfService [0..1] : string	CL_OnlineResource lien/linkage : undefined protocole/protocol [0..1] : string profile application/applicationProfile [0..1] : string nom/name [0..1] : string description [0..1] : string fonction/function [0..1] : Codelist
CL_Telephone vocale/voice [0..*] : string fax/facsimile [0..*] : string	

(a) Classes de l'ISO 19115 relative à la description des organismes et acteurs

Acteur			Organisme_laboratoire		
<u>id individu</u>	<u>LVA50</u>	<u><M></u>	<u>organisationName</u>	<u>TX150</u>	<u><M></u>
IndividualName	LVA50		facsimile	LVA20	
PositionName	LVA100		deliveryPoint	LVA100	
Voice	LVA20		postalCode	LVA15	
Facsimile	LVA20		city	LVA30	
ElectronicMailAddress	LVA50		country	LVA30	
Linkage	LVA100		linkage	LVA100	
Adresse photo identité	LVA100		logo	LVA100	
Descriptif	LVA300		Voice_org	LVA20	
			ElectronicMailAddress_org	LVA50	

(b) Table acteur proposée (c) Table sur les organismes d'après les recommandations implémentée d'après les de l'ISO 19115 pour la mise recommandations de l'ISO en place d'un annuaire 19115

FIG. 5.26 – Utilisation des recommandations de l'ISO 19115 5.26(a) pour l'attribution de propriétés aux concepts acteur 5.26(b) et organisme 5.26(c) de notre ontologie

des domaines impliqués dans la GIZC (voir détail en section 5.1).

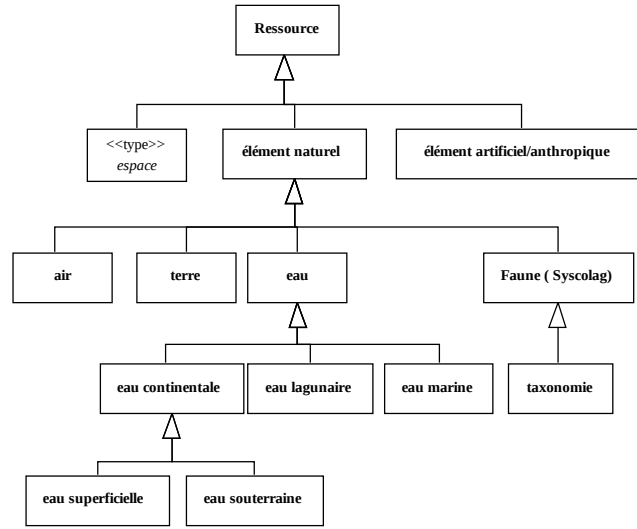
Typologie des Ressources

Nous allons retrouver ici les premiers niveaux de détail sur lesquels nous nous sommes déjà entendus puis nous allons voir comment nous pouvons conjuguer les analyses précédentes avec cette nouvelle approche.

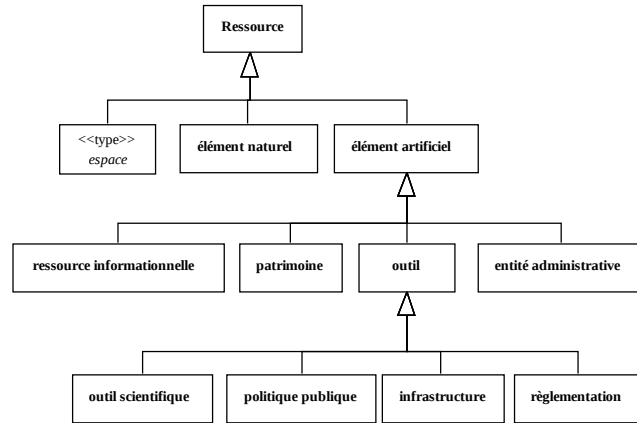
Dans la Figure 5.18(a) présentée en partie 5.2.1, les partenaires de *Syscolag* avaient plébiscité *trois* niveaux dont le dernier consiste en *onze* catégories générales qui spécialisent les concepts *ressource* et *société*.

Dans cette nouvelle approche, nous proposons de spécialiser la classe *ressource* en différentes classes qu'on peut également spécialiser :

- *élément naturel*,



(a) Diagramme des classes de la typologie des ressources naturelles, exemples de spécialisation



(b) Diagramme des classes de la typologie des ressources artificielles, exemples de spécialisation

FIG. 5.27 – Exemples extraits de typologies de ressources naturelles 5.27(a) ou artificielles 5.27(b) d'intérêt pour la GIZC

- *élément artificiel*,
- *espace*.

Nous discutons à présent de la compatibilité de ces approches.

En ce qui concerne les *éléments naturels* (*eau*, *air*, *terre*, *faune*, *flore*), il n'y a pas de changements majeurs par rapport au modèle proposé par nos partenaires si on compare les Figures 5.18(a) et 5.27(a). Ces figures illustrent d'ailleurs une représentation de l'environnement assez commune des *éléments naturels* [145]. De plus, il existe des *vocabulaires contrôlés* de référence pour chacune de ces classes : la *taxonomie*, par exemple, est largement décrite dans la littérature [157, 75]. Nous proposons donc pour spécialiser les *classes* correspondantes de notre ontologie de réutiliser les sources et outils existants. Nous proposons en perspectives,

une illustration de démarche possible pour réaliser cet inventaire sur la base de source de référence (voir partie 8.2.3).

Concernant les classes qui spécialisent le concept *société* dans la Figure 5.18(a) et celles qui spécialisent le concept *Ressource artificielle*, plusieurs remarques peuvent être formulées :

- une *entité administrative* peut être spécialisée en *région*, *département*, *cantons*, *commune*, *communes littorales*...et s'intègre donc bien dans le modèle proposé.
- une *infrastructure* peut être spécialisée en *infrastructure de transports*, *infrastructure d'assainissement*, *infrastructure de télécommunications*...et s'intègre donc également sans problème particulier dans le modèle proposé.
- la *démographie* est l'étude des caractéristiques des *individus* composant une population (humaine ou animale) déterminée [149]. Elle n'est donc pas, à proprement parler, une *ressource artificielle* de la GIZC (son importance dans l'analyse de la GIZC peut toutefois justifier de la *réifier* pour l'étudier en détail). Nous proposons plutôt de gérer ce *concept* sous la forme de *propriétés* de la classe *Individu* (humain) qui généralise le concept *Acteur* (voir Figure 5.25(b)), mais aussi, plus généralement, comme *propriétés* de toute *espèce animale*.
- les classes *politique publique* et *réglementation* relèvent d'après le *modèle d'ontologie* proposé plutôt des *outils* qui permettent la réalisation des *activités de gestion* (voir typologie d'outils en partie 5.3.5). Ceci n'est cependant pas contradictoire, étant donné qu'il est possible de représenter la classe *outil* comme une sorte de *ressource artificielle* du littoral.
- nous proposons, en complément, de spécialiser *Ressource artificielle* par *ressources informationnelles*. En effet, ces dernières sont produites par des *acteurs* et interviennent directement dans la réalisation de certaines *interactions* (acteur/acteur ou acteur/ressource : par exemple la *prise de décision pour l'aménagement*, la *gestion*, les *études scientifiques*...). Nous avons présenté en Figure 4.10 dans la partie 4.3.2, une typologie de *ressources informationnelles* associée aux *gabarits* possibles pour décrire les *fiches de métadonnées*.
- la classe *activité et usage* va désormais être représentée comme une *interaction* dans le modèle d'ontologie proposé (voir partie 5.3.4).

La classe *espace* est indiquée ici en tant que ressource du littoral conformément au sens qui lui est attribué dans la littérature où l'espace est explicitement présenté comme une ressource du littoral, on trouve ainsi des expressions du type *activité consommatrice d'espace*.

Propriétés des Ressources

Globalement, contrairement à la classe *Acteur*, les *propriétés* des concepts du type *ressource du littoral* ne peuvent provenir de l'*ISO 19115* et sont à définir par les *experts du domaine*. Par contre, cette *norme spécialisée* pour la description d'*information géographique* nous sert à décrire les *ressources informationnelles* dans notre *service de métadonnées*.

Cas général des Ressources du littoral Pour les *ressources* du littoral (*artificielles* ou *naturelles* : *eau*, *air*, *terre*, *faune*, *flore*, *infrastructures*, ...), l'*ISO 19115* ne propose aucun descripteur spécifique puisque ces notions sont propres à notre contexte d'application et donc, tous les *attributs* de ces *classes* doivent être proposés dans le cadre d'une *extension de la norme*, propre à notre programme. L'exemple de *taxonomie* sur la *Flore* donné par l'*ISO 19115* (voir détail en partie 5.1.3) illustre bien les nécessaires extensions de cette norme en complément de

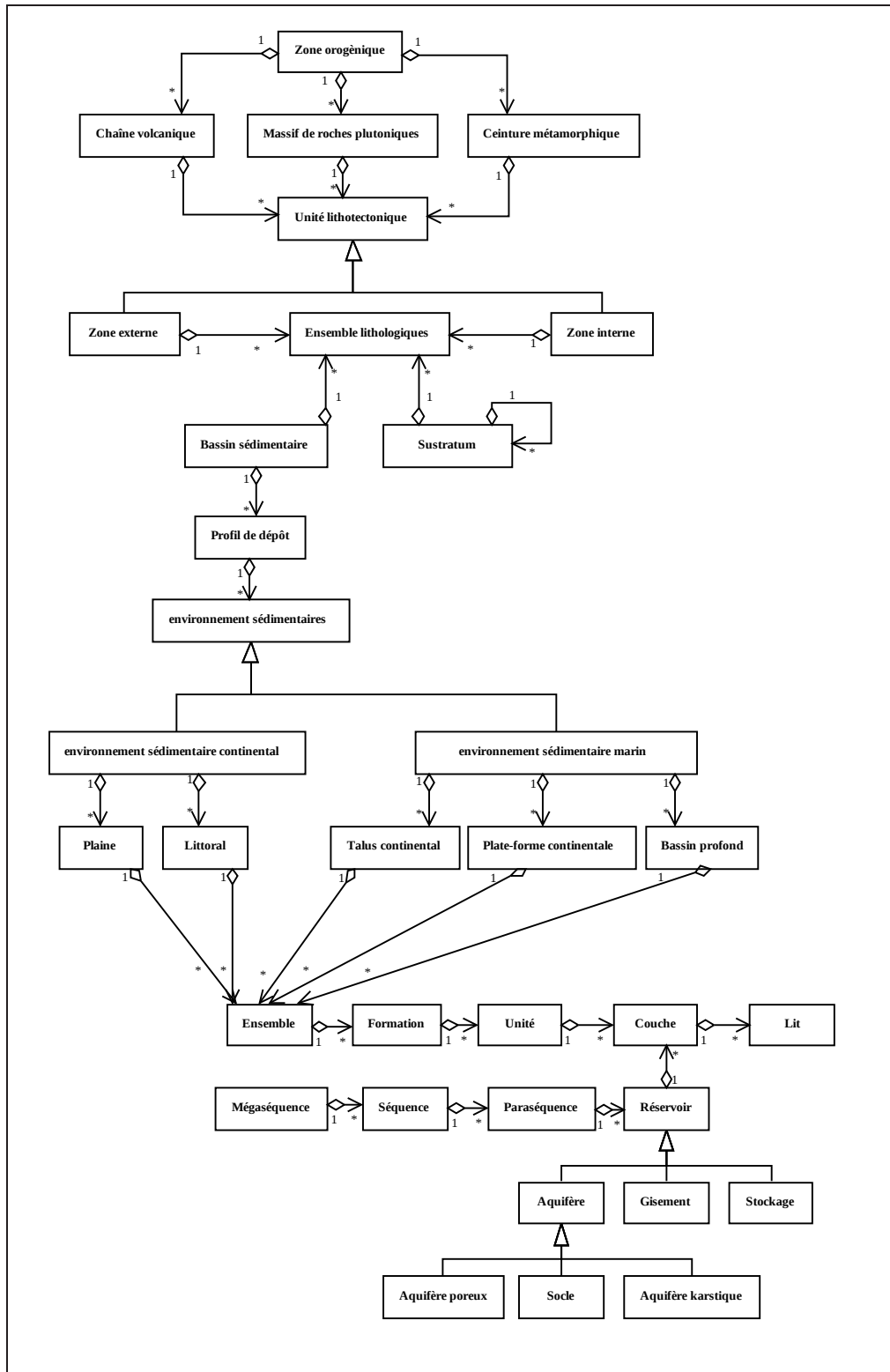


FIG. 5.28 – Diagramme de classes traduisant le point de vue des *géologues* (variété de concepts dérivant du concept *terre*) [91]

ses recommandations générales pour s'adapter aux particularités d'un domaine. Compte tenu de la diversité des *ressources* à prendre en compte, cette tâche ne pouvait être assumée par ce travail de thèse et doit faire l'objet d'un travail à part entière qui s'appuie sur l'expertise des *acteurs* de la GIZC. Nous donnons en perspectives (voir partie 8.2.3 et Figure 8.2) un exemple d'attribution de *propriétés* au concept *Faune* à partir des *sources de référence* du domaine de la GIZC.

Cependant, nous avons considéré dans le modèle proposé que les *ressources informationnelles* sont une sorte de *ressource artificielle* du littoral. Pour ce type particulier de *Ressource*, l'*ISO 19115* sert à en définir les propriétés.

Le cas particulier des Ressources informationnelles Le travail qui consiste à détailler les *propriétés* pertinentes pour décrire les *ressources informationnelles* détenues par les *acteurs* mobilisés dans la GIZC en *Languedoc Roussillon* correspond concrètement à la réalisation du *profil Syscolag* détaillée en partie 4.3.1. Les *classes* présentes dans cette *typologie* possèdent donc des *propriétés* qui correspondent aux *éléments de métadonnées* qui composent les *gabarits* qui leur sont associés (voir partie 4.3.2).

5.3.4 Les Interactions de la GIZC en LR

Nous avons vu le rôle central de la classe *interaction* dans le *modèle d'ontologie* proposé en tant que *classe d'association* entre les classes *acteur* et *ressource* (voir Figure 5.22 en partie 5.2.3).

Typologie des Interactions

Par *interaction* on entend les *processus* (*humains* ou *naturels*) qui mettent en relation différentes *classes d'acteurs* et / ou de *ressources* et dont la réalisation peut (selon la nature de l'interaction) impliquer un ou plusieurs *outils*.

Globalement, les interactions à prendre en compte peuvent être déclinées en :

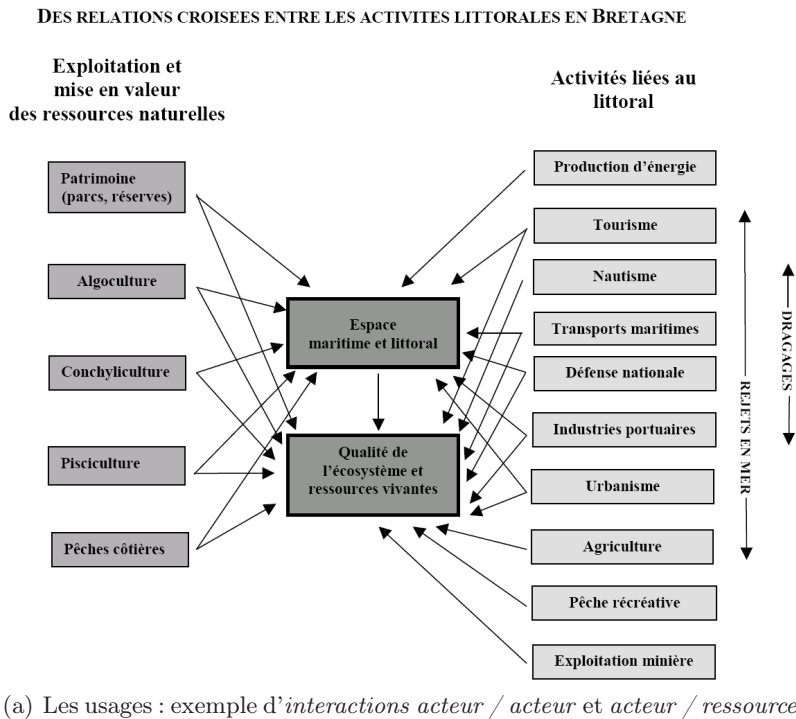
- *Interaction acteur / acteur* : par exemple un *conflit d'usage* ou des *activités socio-culturelles*,
- *Interaction acteur / ressource* : par exemple la *pêche*, l'*aquaculture*,
- *Interaction ressource / ressource* : par exemple la *prédation* d'une espèce sur une autre, ou l'*action érosive d'un cours d'eau* sur un *élément géomorphologique*.

Si les deux premières sont plus faciles à inventorier puisqu'elles impliquent des *acteurs*, les dernières sont naturellement plus complexes et variées. En outre, il est difficile de savoir lesquelles sont pertinentes à intégrer dans une *typologie* propre à la GIZC. Il s'agit, en effet, de définir une limite de finesse dans la nature des *interactions* à prendre en compte, en relation avec l'échelle de la zone d'étude (voir détail en partie 5.3.1).

Pour réaliser cet inventaire d'*interactions* nécessaire pour la description de *processus* ou *scénarios* de GIZC, nous nous sommes appuyés sur les *connaissances* acquises auprès de nos partenaires mais aussi sur des *sources de référence* comme celles associées à l'*approche écosystémique* (voir partie 5.2.2, Figures 5.20(b) et 5.21).

Typologie d'Interactions acteur / acteur et acteur / ressource Les termes *activités* et *usages* correspondent à *sortes d'interactions* entre *acteurs* ou entre *acteurs* et *ressources naturelles* qui peuvent être spécialisées. Par exemple la *pêche*, l'*aquaculture*... sont des *sortes*

d'interactions d'acteurs sur les ressources naturelles. A ce sujet, il existe de nombreux travaux de référence, comme le travail de [166, 167] sur la représentation logique des activités humaines en zone côtière (voir Figure 5.30(a)).



SE DIVERTIR ET PARTAGER L'ESPACE				
Activité de loisirs		Concurrence avec :	Origine des difficultés	Que faire ?
Pêche	Pêche récréative (à pied)	Conchyliculture	- Risque de pillage de la ressource (les touristes se rapprochent trop des sites des professionnels) - Gêne dans les numéros des professionnels	- Bon balisage des parcs
		Protection du milieu naturel	- Bruit occasionné par l'activité et la fréquentation d'un site occasionnant un dérangement et une fuite de certaines espèces (oies bernaches dans le Golfe du Morbihan)	
	Pêche sportive	Véloplanchistes	- Lorsque les pêcheurs sportifs organisent des compétitions amicales, ils balisent la zone neutralisée. Certains véloplanchistes ignorent trop souvent ce balisage : vives tensions	
		Protection du milieu naturel	- Dérangement occasionné par l'activité et la fréquentation d'un site occasionnant un dérangement et une fuite de certaines espèces (oies bernaches dans le Golfe du Morbihan)	
Sports de plein air		Autres activités de plein air de tous types	- Cohabitation dans l'espace	- Mettre en place un PDESI (Plan Départemental des Espaces, Sites et Itinéraires)
		Toutes activités consommatrices d'espaces	- Construction et extension de très nombreux ports de plaisance pour faire face à la demande : au « détriment » d'autres activités à la recherche d'espaces (sites de loisirs : extension Océanopolis) - Construction de locaux d'accueil, vestiaires... - Risque de privilégier certaines activités au détriment des autres (conchyliculteurs, pêcheurs...) - Problèmes de stationnement. Le fort développement du kayak de mer n'est pas non plus sans poser problème, essentiellement en ce qui concerne l'embarquement et le débarquement du matériel qui se fait par remorques (problèmes de stationnement (stockage) des remorques, donc de partage de l'espace).	- Développer la construction de ports de plaisance à sec - Encourager la location de voiliers de plaisance pour limiter la construction de nouveaux ports de plaisance
		Autres sports nautiques (plongée...)	- Dangers de collisions	- Encadrer davantage l'activité « kayak » et encourager l'adhésion à un club pour réguler les pratiques
		Baignade	- Risque de collisions avec baigneurs, surtout lorsque des embarcations quittent des plages non aménagées de chemaux - Certains baigneurs utilisent les cales comme plongeurs	- Aménager des chemaux sur les plages

(b) Les conflits d'usage comme source pour l'inventaire d'interactions acteur / acteur et acteur / ressource

FIG. 5.29 – Utilisation des sources de référence pour l'inventaires des activités et usages sur la zone côtière

En ce qui concerne les interactions acteur / acteur, une catégorie fondamentale dans la GIZC concerne les conflits d'usages. De nombreux travaux sont dédiés à la résolution ou à la

caractérisation des *conflits d'usages*. La caractérisation de ces *conflits* passe par l'identification des *activités* et *usages* en opposition autour d'une *ressource* (par exemple l'*espace*). Il est donc possible d'intégrer les typologies d'*activités et usages* qui y sont implicitement présentées dans le modèle d'ontologie proposé. La Figure 5.29 montre un exemple de source potentielle pour construire les typologies souhaitées. La Figure 5.29(a) illustre les *interactions* entre *activités* de la zone côtière en Bretagne (mais qui sont transposables à d'autres zones côtière), la Figure 5.29(b) est un extrait de l'inventaire des *conflits d'usages* répertoriés entre différentes sortes d'*activités* dans [28].

En dehors des relations qui concernent les acteurs de la zone côtière, il convient de prendre en compte celles qui concernent les ressources du littoral.

Typologie d'Interaction ressource / ressource Cette *sorte d'interaction* n'implique pas les *acteurs* mais uniquement les *éléments* de la zone côtière. Par exemple : les processus climatiques (*tempête...*) ou biologiques (*prédation, reproduction...*), l'hydrodynamique (*houle, vague, inondation...*)...

Remarque – On peut noter que des *interactions* peuvent être contenues implicitement (voire exprimées explicitement) dans les éléments d'indexation (*mots clés*) des fiches de métadonnées. En effet, des mots clés juxtaposés dans l'indexation impliquent généralement un lien dans le fonctionnement de l'écosystème qu'on peut réutiliser comme source pour insérer de la connaissance dans l'*ontologie*. Les métadonnées peuvent donc aider à alimenter le modèle de l'ontologie, comme le montre l'exemple d'indexation suivant : *conflit d'usage, plaisance, conchyliculture, étang de Thau...*

Propriétés des Interactions

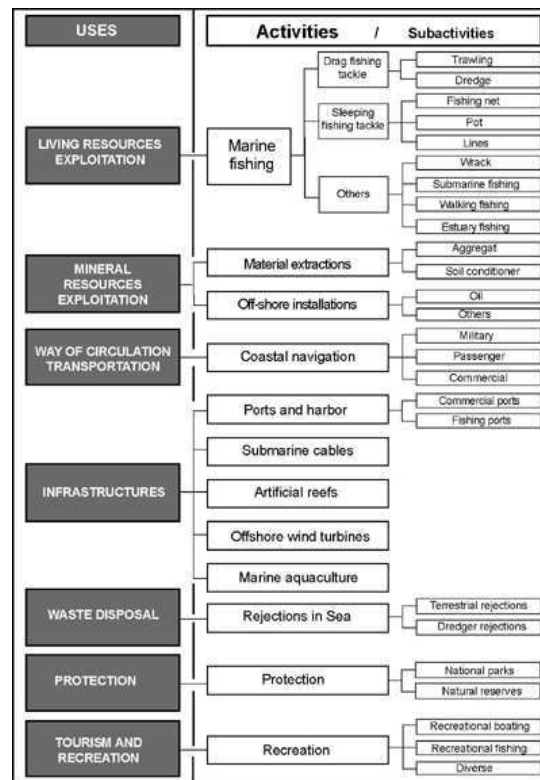
Les propriétés des *interactions*, comme celles des *ressources*, sont entièrement à définir, selon la nature des *acteurs* et des *ressources* considérés. C'est un travail spécifique à chacune des classes qu'il convient d'effectuer avec les experts du domaine en fonction des *interactions* pertinentes pour la GIZC. Nous attirons l'attention sur la possibilité de *spatialiser* les *interactions* : *spatialisation des activités et des usages*, *spatialisation des conflits d'usages*, *spatialisation de phénomènes climatiques ou écologiques...*

5.3.5 Les outils de la GIZC en LR

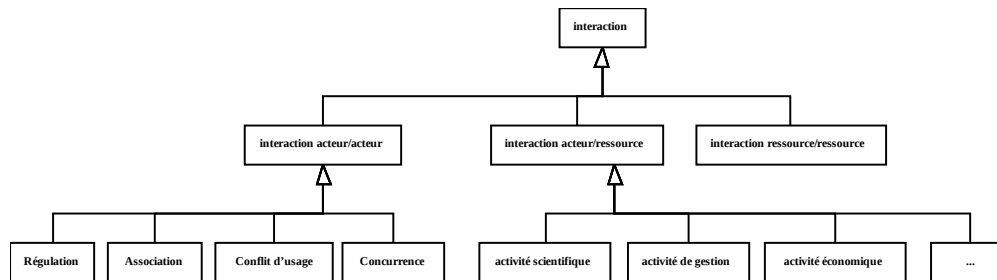
Par *outil* de la GIZC, on entend les *instruments, matériels, méthodes scientifiques, économiques, politiques, réglementaires...* qui permettent la *réalisation* de diverses *interactions* essentiellement humaines.

Typologie des outils

Le concept d'*outil* est très présent dans le domaine de la GIZC (voir partie 2.1.3). Les *outils* sont obligatoirement associés à une *interaction* d'origine humaine (voir Figure 5.22). Ils sont conçus pour la réalisation d'une *activité* (de gestion, économique, de loisirs...). Par exemple l'*activité* de *pêche* nécessite des *outils* ou des *engins de pêche* du type : *palangres, chaluts...*, alors que l'*activité* de *gestion* va nécessiter des *outils* du type : *Natura 2000, SCOT, PLU, SMVM...* De la même manière la *production scientifique* utilise des *outils* du type : *SIG, télédétection, analyses statistiques, relevés, forages...*



(a) Typologie d'activités humaines en zone côtière [167]



(b) Diagramme des classes de la typologie des interactions [18]

FIG. 5.30 – Utilisation de sources de référence 5.30(a) pour alimenter la typologie d'interactions de notre ontologie 5.30(b)

Cette classe va donc comprendre des *outils* de natures variées, comme le suggère la Figure 5.31(a) qui présente la *boîte à outil* nécessaire de la GIZC [147], ie *outils scientifiques*, de *politiques publiques*. ... Nous en proposons une formalisation UML en Figure 5.31(b) que nous pouvons utiliser comme trame pour agréger d'autres *concepts* puisés dans d'autres *sources de référence*.

Les apports issus des différentes *sources de référence* peuvent donc, là encore, alimenter la *typologie des outils* nécessaire à la formalisation des problèmes de GIZC. Nous proposons un exemple de premier niveau de typologie dans la Figure 5.32.

Propriétés des outils

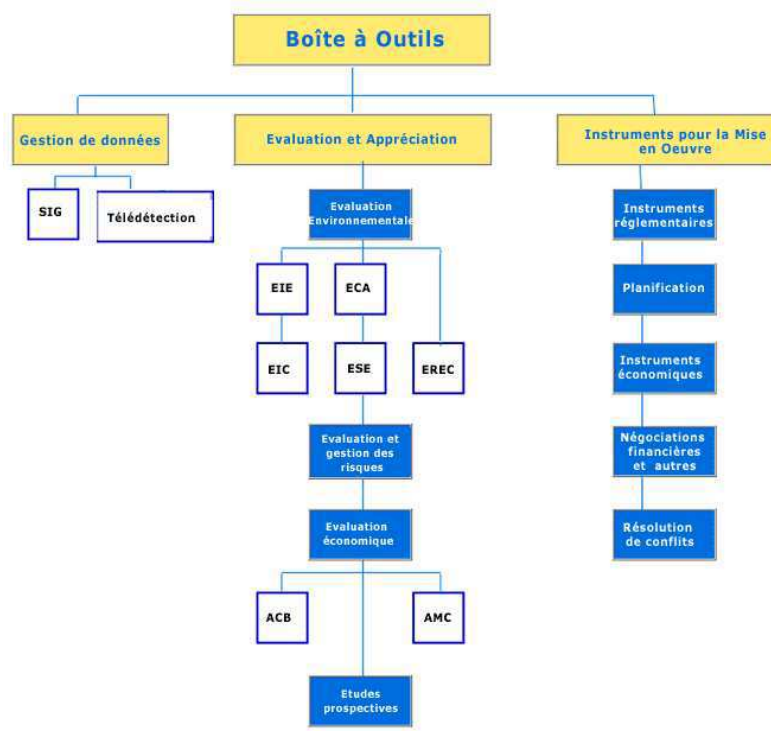
Nous avons vu qu'un *outil* est un élément ou une *ressource artificielle* du littoral au service de la réalisation d'une *interaction* qui implique des acteurs. Comme pour les *ressources*, l'*ISO 19115* ne permet pas de définir de *propriétés* spécifiques de la classe *outil* sauf bien sûr, si on considère que certaines *ressources informationnelles* sont des *outils*.

Remarque – un *outil* n'existe que par rapport à la réalisation d'une *interaction*. Or une *interaction* n'existe que par les relations entre deux *éléments* de la zone côtière (*acteur(s)* et/ou *ressource(s)*, voir partie 5.2.3 et Figure 5.22). Le concept d'*outil* véhicule donc implicitement *trois concepts* : deux *éléments* en *interaction*.

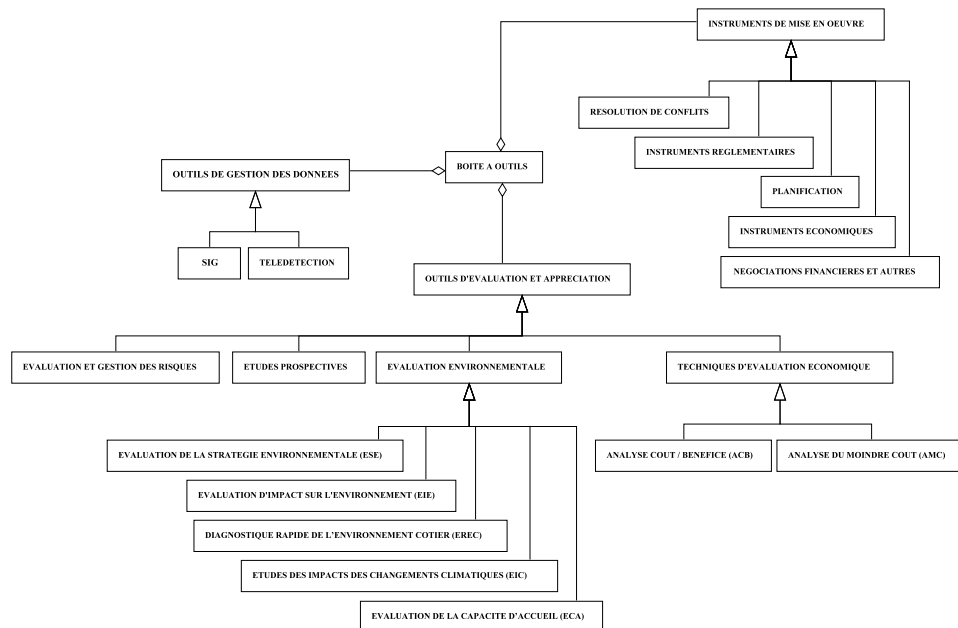
Cela permet de répondre à des questions du type : à quelles espèces de poissons sont associés les *outils (engins) de pêche x* et *y* ? Quels sont les *outils de gestion* associés à l'*eutrophisation* ? ... En plus de cela, il est bien sûr possible selon la nature des catégories d'outils, de leur attribuer des propriétés plus particulières.

Synthèse de la section 5.3

- Afin d'implémenter le *modèle* d'ontologie que nous proposons pour la GIZC, nous réalisons l'*inventaire des concepts* pertinents appartenant à chacune des *quatre principales classes* et nous proposons un premier niveau de *structuration* par l'utilisation de hiérarchies (*topologies*) basées sur l'utilisation de liens de *généralisation* / *spécialisation* et d'*agrégation* (liens *BT* et *NT*), plus intuitifs pour les utilisateurs.
- Nous intégrons les *typologies* de référence du domaine pour nourrir ce modèle ainsi que les travaux existants qui peuvent supporter l'élaboration d'ontologies, en particulier pour l'attribution de *propriétés* (en plus des apports de l'*ISO 19115*...).
- Les sources de référence des domaines correspondants aux concepts retenus peuvent être utilisées, dans un deuxième temps, pour définir leurs propriétés sémantiques (les *définitions*, liens *SN*, les *synonymes* de référence, lien *UF*...).
- Ces typologies peuvent servir, à terme, de support pour représenter d'autres types de *relations* entre les *concepts inventoriés*, notamment les relations de types *RTS* qui expriment l'*aspect dynamique du système* et sont spécifiques à la GIZC.
- Plus généralement, la mise en ligne de ces typologies doit permettre un travail collaboratif d'alimentation, d'évaluation de la pertinence du modèle d'ontologie proposé et, éventuellement, sa validation.



(a) "Boîte à outils" pour la GIZC



(b) Formalisation UML de la Figure 5.31(a)

FIG. 5.31 – Utilisation de sources de référence 5.30(a) pour alimenter la typologie d'interactions de notre ontologie 5.30(b)

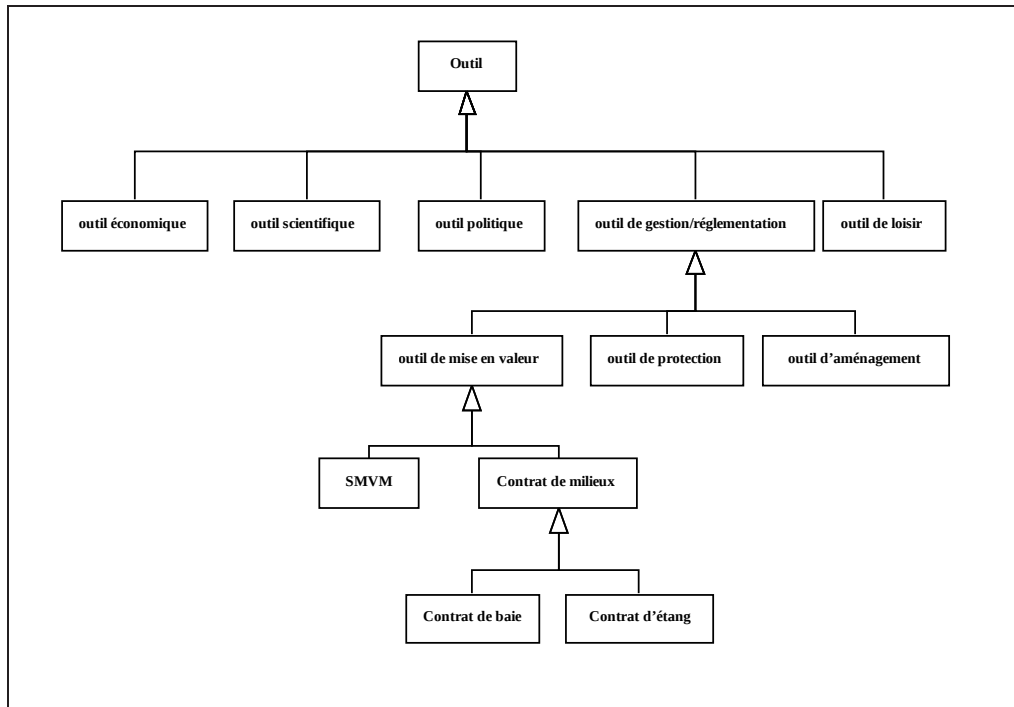


FIG. 5.32 – Diagramme des classes de la typologie des outils impliqués dans la GIZC

5.4 Implémentation du modèle d'ontologie proposé

Après avoir synthétisé les analyses que nous avons présentées dans ce chapitre, nous proposons un exemple pratique d'utilisation du *modèle d'ontologie* proposé pour la formalisation de *connaissances* plus complexes.

5.4.1 Utilisation "raisonnée" du modèle d'ontologie

L'utilisation de notre *modèle* ne présente pas de difficultés majeures. Il s'agit de l'utiliser comme *patron* pour permettre la *mise en forme de connaissances* en répondant aux règles du modèle. Il permet, par exemple, la formalisation de *connaissances élémentaires* qui correspondent à des phrases du type :

- Le *loup de mer* a pour nom scientifique *Dicentrarchus labrax*,
- Le *loup de mer* a pour synonymes *loup* ou *loup blanc* en Méditerranée, *loubarrou* en Roussillon, *loubassou* en Provence, *ragnola* en Corse...
- Le *loup de mer* habite les *fonds rocheux*...
- Le *loup de mer* chasse les *Poissons*(lançons, sardines), *Céphalopodes*(encornets, seiches), *Crustacés*(crabes, crevettes), *Vers marins*(arénicoles, gravettes, siponcles), *Coquillages*...
- Le *loup de mer* est pêché à l'aide de *chaluts* et de *palangres* dans le *Golfe du Lion*.
- *Chaluts*, *palangres*...sont des sortes d'*outils* du type *engins de pêche*...

L'exploitation optimale de l'*ontologie* que nous proposons sous-entend au préalable l'*inventaire* aussi exhaustif que possible des *concepts* nécessaires à l'*indexation* des ressources et au *partage des connaissances* entre les *acteurs* mobilisés dans la GIZC, puis l'introduction de *relations*

entre ces *concepts* qui obéissent à celles proposées par notre diagramme de la Figure 5.34(a) et enfin l'attribution de *propriétés* et d'*instances*.

Actuellement, compte tenu de l'ampleur de la tâche imposée par la diversité des thématiques impliquées, nous proposons d'exploiter le *modèle d'ontologie* proposé sur la base de *scénarios* de GIZC tels que, par exemple, l'*eutrophisation dans l'étang de Thau*. En cela, nous adhérons aux démarches d'autres programmes de recherche tels que *COASTBASE* [132] (voir présentation en partie 2.3.2).

Ainsi (grâce à l'existence de *typologies* détaillées en conséquence) le *scénario* de la *pêche au merlu en Languedoc Roussillon* peut, en suivant les règles de notre *modèle*, être *instancié* tel que cela est présenté en Figure 5.33.

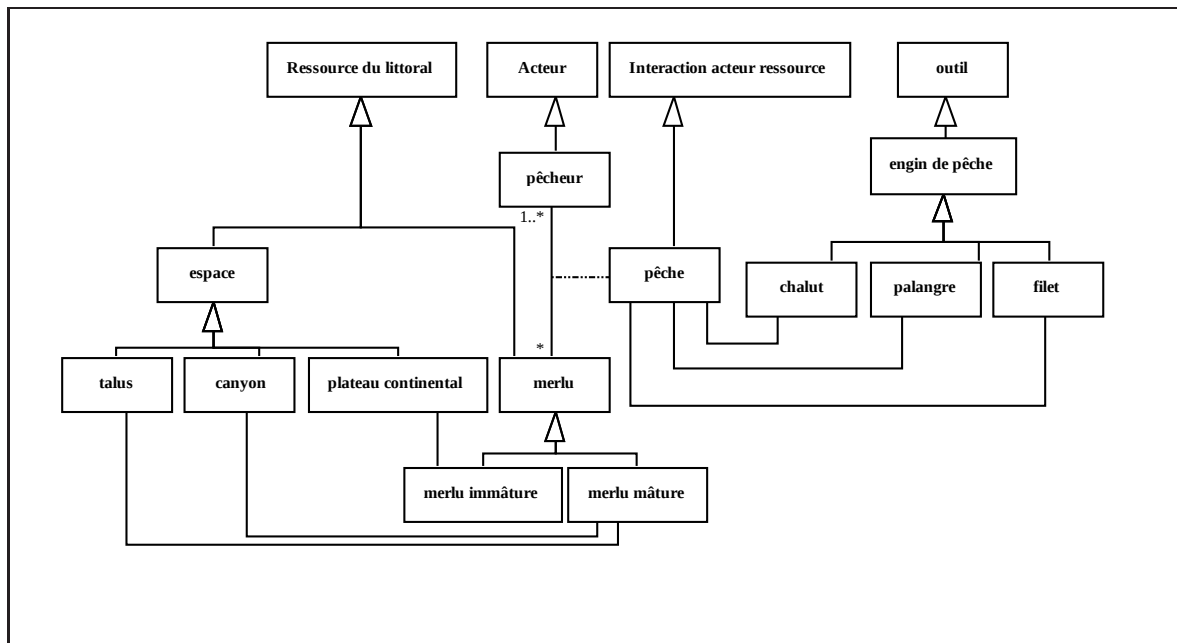


FIG. 5.33 – Exemple de scénario sur la "pêche au merlu" formalisé à l'aide de notre *modèle d'ontologie*

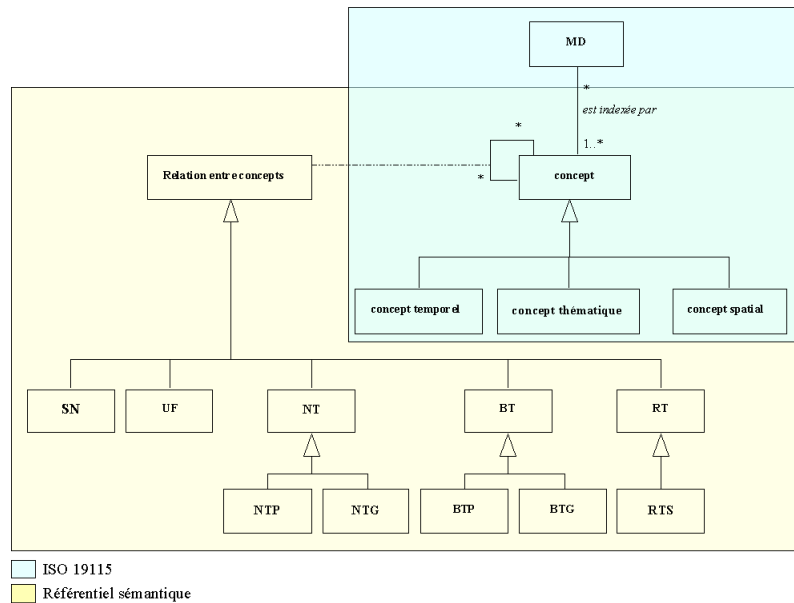
Dans la partie suivante nous formalisons le lien qui permet la *connexion* du *référentiel sémantique* au *service de métadonnées* en vue de son utilisation comme *outil de contrôle du vocabulaire*.

5.4.2 Connexion entre métadonnées et connaissances

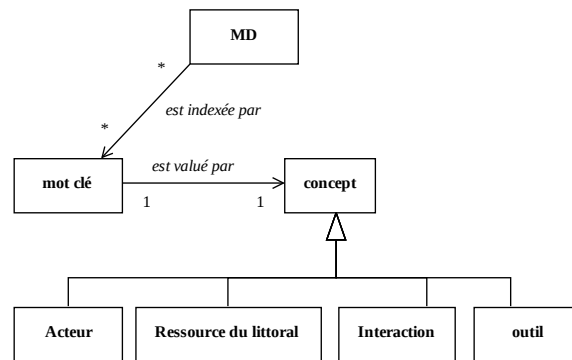
La *connaissance* contenue dans le *référentiel sémantique* peut, à présent, être exploitée dans le *service de métadonnées* pour assister le *contrôle* de la *saisie* de l'*élément de métadonnée mot clé*, mais aussi, à terme, pour affiner les *requêtes de consultation*.

Valuation de mots clés à l'aide du référentiel sémantique

Conformément à l'exemple d'*extension* de la norme *ISO 19115*, donné par la norme elle-même [2] (voir partie 5.1.3 Figure 5.10), nous proposons l'*adaptation* et l'*extension* de la norme par une *ontologie* tel que cela est illustré en Figure 5.34(a).



(a) Nature du lien entre un référentiel sémantique et un service de méta-données [14]



(b) Adaptation de la nature des concepts du référentiel sémantique au domaine de la GIZC

FIG. 5.34 – Synthèse sur le lien du modèle conceptuel du référentiel sémantique avec le service de métadonnées 5.34(a) et la nature des concepts qui l'alimentent pour le domaine de la GIZC 5.34(b)

Lors de l'indexation d'une fiche de métadonnée pour la GIZC, on prête une attention particulière aux *éléments de description thématiques* et *spatiaux*. Les valeurs qu'on attribue aux *mots clés* (lorsqu'on les instancie) sont tirés de *concepts* gérés au sein d'une *ontologie* qu'on peut considérer comme des éléments ajoutées à l'*ISO 19115* pour contrôler la valuation des mots clés. Mais on va voir que c'est aussi une *ontologie spatiale* qui permet de gérer le géoréférencement de *toponymes* ou *mots clés spatiaux* (voir partie 6.1.2).

En dehors de la simple *valuation* de l'élément de métadonnées *mot clé*, ce lien permet d'exploiter la connaissance associée aux concepts gérés dans le référentiel sémantique lors de la *consultation* du *service de métadonnées*.

Exploitation du référentiel sémantique pour l'amélioration des consultations du service de métadonnées

Les exemples que nous proposons ici illustrent les possibilités offertes par la mise en place du *modèle d'ontologie* pour des *expansions de requêtes* au sein du *service de métadonnées*. Sur la base des exemples que nous avons détaillés en partie 5.1.1, il est à présent possible de permettre d'enrichir les réponses apportés par le système lors d'une *consultation* de métadonnées en s'appuyant sur les *relations ontologiques* gérées dans le *référentiel sémantique*. Ainsi, à la question : *requête 1* = "Trouver les fiches de métadonnées associés au concept loup de mer", le système peut, par exemple, étendre la requête aux cas suivants :

- *requête 2* = *requête 1* + "Trouver les fiches de métadonnées associés au concept loup de mer ou à tous ses synonymes (relations *UF*)",
- *requête 3* = *requête 2* + "Trouver les fiches de métadonnées associés aux concepts qui généralisent loup de mer" ie à toutes les espèces du même ordre biologique (relations *BTG rang+1*) ou à toutes les espèces du même genre biologique (relations *BTG rang+2*)...
- *requête 4* = *requête 3* + "Trouver les fiches de métadonnées associés aux concepts associés au concept loup de mer (relations *RTS*)"
- *requête n* =

L'implémentation concrète du *prototype* permettra l'opérationnalisation de ces *expansions de requêtes*.

En pratique, la réalisation du choix d'un *concept* pour la *valuation des descripteurs thématiques et spatiaux* s'appuie sur un *modèle physique de données*.

5.4.3 Réalisation : modèle physique de données

Le travail de traduction des spécifications *UML* en un *modèle physique de données (MPD)* géré dans *Postgres* a fait l'objet d'un travail indépendant [107]. Le *référentiel sémantique* peut ainsi être sollicité dans le *service de métadonnées* ou séparément afin de restituer la *connaissance* liée aux *concepts* qui y ont été intégrés. Pour cela le *MPD* doit traduire le lien entre *métadonnées* et *référentiel sémantique* exprimé en Figure 5.34(a) et ceux entre les *concepts* dans le *modèle d'ontologie spécifique au domaine de la GIZC* présenté en (Figure 5.34(b)). Le *MPD* est présenté en Figure 5.35.

En outre, afin de permettre d'éventuelles *évolutions* du *modèle d'ontologie*, ou des *déclinaisons disciplinaires* de ce modèle, le *MPD* réalisé par [107] est générique : ie indépendant du *modèle d'ontologie*. Nous présenterons dans le chapitre 7 l'exploitation de son contenu pour la gestion du *service de métadonnées* (voir partie 7.2) ou pour la constitution d'une *base terminologique* au service des *acteurs* (voir partie 7.3).

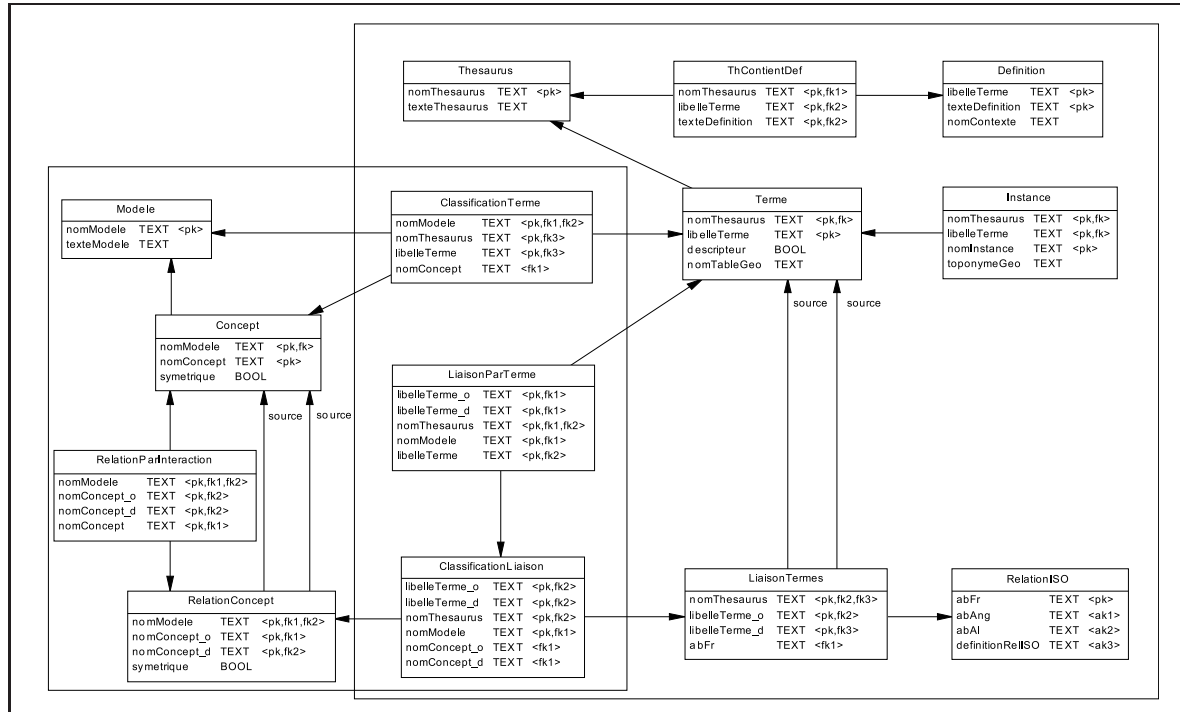


FIG. 5.35 – Modèle physique de données du référentiel sémantique implémenté dans *Postgres* [107]

Synthèse de la section 5.4

- Le modèle proposé ouvre des perspectives intéressantes pour la *représentation de connaissances* simples et complexes (scénarios) indispensables au partage des connaissances de la GIZC.
- Le *référentiel sémantique* peut être représenté comme une *extension d'un service de métadonnées* qui implémente l'*ISO 19115*.
- Il permet le *contrôle de la valuation* de l'*élément de métadonnée mot clé* et rend possible l'*expansion de requêtes* afin d'améliorer les possibilités de consultation des fiches de métadonnées existantes.
- Un *modèle physique de données* générique permet la réalisation du *référentiel sémantique* indépendamment du *modèle d'ontologie* proposé. Il est utilisable dans le *service de métadonnées* ou pour être consulté séparément.

Conclusion du chapitre 5

La *consultation* des partenaires de notre projet pour la mise en place d'un *service de métadonnées* basé sur l'*ISO 19115* a révélé l'importance des *descripteurs spatiaux et thématiques* et la nécessité de les contrôler pour résoudre les problèmes de *sémantique* associés à la valuation des métadonnées en général et pour la GIZC en particulier. Nous souhaitons donc proposer un outil qui permet d'améliorer l'usage de ces éléments dans le *service de métadonnées* ainsi que d'intégrer une part de la *connaissance experte* des *acteurs* mobilisés dans la GIZC. Cet outil constitue ce que nous appelons un *référentiel sémantique* dont l'objectif est d'être interprétable par les machines comme par les acteurs.

Parmi les *outils de contrôle du vocabulaire* existants, les *ontologies* permettent d'intégrer une part plus fine de la connaissance associée aux concepts qu'elles inventorier par des relations sémantiques plus riches que celles des *glossaires* ou des *thésaurus*. Toutefois ces outils ne sont pas incompatibles et peuvent s'enrichir les uns les autres par l'*inventaire des concepts* et la nature de certaines *relations* que leur élaboration implique nécessairement. Nous préconisons donc de réutiliser les outils de contrôle du vocabulaire existants (regroupés dans un état de l'art) pour élaborer une *ontologie du domaine de la GIZC*.

Compte tenu de notre problématique et de celles émergentes du *Web sémantique*, nous proposons la mise en place d'une *ontologie* formalisée en langage *UML* afin de pouvoir interagir avec les acteurs du domaine en prenant en compte la formalisation de la connaissance employée par les langages du *Web sémantique*. En outre, nous complétons nos analyses en nous appuyant sur l'*approche écosystémique*, pertinente pour inventorier et structurer les principaux éléments du système de la zone côtière.

Le *modèle* proposé est un modèle *générique* pour la *gestion d'écosystèmes anthropisés* dont la GIZC n'est a priori qu'une des déclinaisons possibles. Sur la base des *entretiens* et réunions réalisés auprès de nos partenaires, nous avons proposé dans ce chapitre une illustration d'utilisation de ce modèle (composé de quatre classes principales) pour constituer une ontologie spécifique à la GIZC, encore inexistante d'après notre état de l'art. Nous proposons des exemples d'utilisation du modèle que nous proposons basés sur la mise en place de *typologies* qui s'appuient sur les *sources de référence* du domaine.

Pour conclure nous proposons un modèle qui présente le *référentiel sémantique* proposé sous la forme d'une *extension de la norme ISO 19115*. Ce modèle permet notamment d'utiliser les *connaissances* intégrées dans le référentiel pour optimiser la gestion du service de métadonnées (*contrôle de la valuation* de l'élément *mot clé*) mais il ouvre également des perspectives pour mettre en œuvre des *expansions de requêtes* lors de sa consultation ainsi que pour la représentation de *connaissances plus complexes*.

Nous avons vu dans la partie 4.3.1 que nous avons rendu les *descriptions thématique et spatiale* obligatoires dans le *profil Syscolag*. En outre, nous souhaitons pouvoir utiliser le *support cartographique* pour assister la résolution des problèmes d'*interopérabilité sémantique*. Le *référentiel sémantique* que nous venons de décrire doit donc être complété pour supporter le mode de description spatiale basé sur la *topologie*. A ces fins, il est nécessaire d'explicitier le lien existant entre *descripteurs thématiques et spatiaux* afin de réifier le lien implicite qui existe entre ces deux catégories. Cela nous permettra de compléter les *relations ontologiques* déjà riches entre *concepts* par l'adjonction de *relations spatiales* pour les *mots clés spatiaux*, permettant par exemple la gestion des *relations topologiques*. Ainsi, nous allons pouvoir proposer l'utilisation du *support cartographique* comme complément ou alternative à la description textuelle. Nous allons voir que le souci de rester compatible avec les *standards* de l'*OGC* en vi-

gueur pour décrire l'*information géographique* nous permet de proposer des fonctionnalités qui facilitent la *saisie* ou la *recherche* des *métadonnées* sur des critères spatiaux. Elles s'appuient sur des *outils de visualisation conviviaux et standardisés* qui allègent la lourdeur potentielle de l'*ISO 19115* et améliorent la *restitution de la connaissance*.

Synthèse du Chapitre 5

- Le partage des *connaissances* associées à la GIZC implique la mise en place d'un *référentiel sémantique* exploitable par les *acteurs* et les *machines*.
- En particulier, ce *référentiel sémantique* permet le *contrôle du vocabulaire* dans un *service de métadonnées* afin d'optimiser la *localisation de l'information* ainsi que l'aide à l'élaboration d'un *langage commun* dans un contexte pluridisciplinaire.
- Parmi les différents types d'*outils de contrôle du vocabulaire* existants, les *ontologies* permettent une gestion plus fine de la *connaissance* associée aux *concepts* qu'elles inventorient (attribution de *propriétés* aux *concepts*, *instanciation* et *liens sémantiques* plus fins pour les *concepts* inventoriés).
- Les *glossaires*, *thésaurus* et *ontologies* existants dans des domaines impliqués par la GIZC sont réutilisables et peuvent servir de *briques* pour la mise en place d'une *ontologie* spécifique à ce domaine.
- Nous proposons le formalisme *UML* pour mettre en œuvre une *ontologie* du domaine de la GIZC, compte tenu de la qualité du dialogue qu'il permet avec les *acteurs* et de l'existence de passerelles avec la formalisation de la connaissance utilisée par les langages du *Web sémantique*.
- Afin de proposer un *modèle d'ontologie* adapté spécifiquement à la GIZC, nous nous appuyons sur une démarche collaborative avec nos partenaires ainsi que sur la prise en compte de l'*approche écosystémique*. En effet, cette *approche* est particulièrement intéressante pour *inventorier* et *structurer* les *éléments* et *interactions* du *système* de la zone côtière.
- Le *modèle générique d'ontologie* qui en résulte répond aux besoins spécifiques de la GIZC. Il permet de représenter la nature des *interactions* essentielles dans la GIZC entre les éléments principaux de la zone côtière que sont les *acteurs* et les *ressources* du littoral. En outre, il est compatible avec le mode de formalisation de la connaissance utilisé par les langages du *Web sémantique*.
- La dimension spatiale est implicite dans les *concepts* de ce modèle qui sont tous, potentiellement, *spatialisables* par des *descriptions spatiales*.
- Afin d'alimenter le *modèle d'ontologie* que nous proposons, nous amorçons un *inventaire des concepts* et de leurs propriétés pour chacune des *quatre principales classes* sous la forme de *typologies*. Pour cela, nous préconisons de nous appuyer sur les sources de référence du domaine.
- Ce modèle ouvre des perspectives intéressantes pour la représentation de connaissances simples et complexes (*scénarios*).
- Nous proposons un modèle conceptuel qui permet de gérer le *référentiel sémantique* comme extension d'un *service de métadonnées* basé sur la *norme ISO 19115*.

Chapitre 6

Le référentiel spatial

Les méthodes présentées dans ce chapitre sont *génériques*¹, nous présentons ici leur application à la thématique de la GIZC. Nous avons vu dans les chapitres précédents que l'objectif principal du programme *Syscolag* est la construction d'outils scientifiques dédiés à une gestion durable de sa zone côtière et qu'en outre, la région *Languedoc Roussillon* souhaite constituer à moyen terme un *observatoire* pour soutenir la GIZC régionale. En particulier, d'un point de vue informatique, il s'agit d'optimiser la *mutualisation*, la *gestion* et l'*accès des informations et connaissances* associées aux *ressources informationnelles* nécessaires pour améliorer, à terme, le processus de GIZC lui-même [47].

Face à l'hétérogénéité des acteurs et organismes impliqués dans la GIZC ainsi que des modes de gestion des ressources informationnelles produites (voir partie 3.1.2), nous avons opté pour une *architecture de système d'information* qui permet de *fédérer* ces sources d'information *hétérogènes* et *réparties* grâce à un *service mutualisé de métadonnées* [16] (voir partie 3.1.2). Ainsi, nous facilitons et mutualisons les descriptions des ressources existantes et permettons leur localisation. Ce service est accessible via un *portail Web*, dont [101] souligne l'intérêt particulier pour la GIZC. Dans notre cas, ce choix a particulièrement été motivé par le fait que l'outil à réaliser est destiné non seulement à la *communauté scientifique* mais surtout aux *gestionnaires* et au *grand public*.

Or, lors de nos analyses, il est apparu que l'*indexation des problématiques de GIZC* s'effectue avec des *mots clés* et s'articule essentiellement autour de *concepts thématiques et spatiaux* (voir partie 4.3.1). Pour assurer la qualité de l'*indexation*, nous avons proposé dans le chapitre 5 la mise en place d'une *ontologie* comme *référentiel sémantique* pour contrôler la description des différents types de *mots clés* de notre domaine, cependant les exemples donnés concernaient surtout les *concepts thématiques*. A présent, nous allons approfondir la question des *descripteurs spatiaux* au sein du *référentiel sémantique*.

L'omniprésence *explicite* ou *implicite* de la dimension spatiale dans le domaine de la GIZC, qui est un *mode de gestion territorialisé*, nous a encouragé à privilégier cette dimension dans le *service de métadonnées* développé et à l'intégrer dans notre *référentiel sémantique*. En outre, de l'avis de nos partenaires, le *support cartographique* est perçu comme un possible affranchissement des problèmes de terminologies interdisciplinaires qui peut donc assister l'*interopérabilité sémantique*. L'importance de l'*espace* dans la GIZC a été évoquée (voir partie 2.1.2), nous allons à présent proposer une réponse spécifique sur laquelle pourra s'appuyer la *médiation*

¹Ces méthodes ont été validées sur la thématique du programme *ROSELT* à l'aide du logiciel *MDWeb*, voir partie 4.2.3.

spatiale (conformément à l’objectif annoncé en partie 3.1.3) en utilisant les possibilités de *représentation cartographique* associées à ce type d’information.

Dans ce chapitre, nous souhaitons, montrer comment nous proposons d’exploiter les *propriétés géométriques* des *descripteurs spatiaux* pour donner une place centrale à l’espace dans la *description* comme dans la *consultation de métadonnées* et de *connaissances*. Pour cela nous proposons l’élaboration d’un *référentiel spatial* qui regroupe ces descripteurs et s’intègre dans le *référentiel sémantique*. En effet, le *référentiel spatial* n’est pas indépendant du *référentiel sémantique* et nous proposons un *modèle conceptuel* qui explicite la nature des liens qu’ils entretiennent.

Cette fois encore, pour mettre ce modèle en œuvre et afin d’assurer la *pérennité* de ce *système d’information* à l’échelle régionale, ainsi que son *interopérabilité*, à terme, avec d’autres systèmes au plan national ou à l’étranger, nous avons privilégié l’utilisation de *standards* pour les différentes composantes du système mis en place (recommandations de l’*Open GIS Consortium* en particulier pour les spécifications sur les *Web Services* permettant la gestion d’*IG* mais aussi pour les sources de référence pour l’inventaire et la *structuration* des *concepts spatiaux* de référence et *objets géographiques* associés). Le choix de l’*ISO 19115* pour la GIZC, justifié dans le chapitre 4 est, là encore, pertinent en ce qui concerne les éléments de métadonnées proposés pour la description spatiale des ressources informationnelles, aussi bien en mode textuel que cartographique.

Ce chapitre est organisé comme suit. Dans la section 6.1, nous revenons sur l’intérêt d’un développement conçu en fonction des standards internationaux (*ISO* et *OGC*) et des spécificités du domaine de la GIZC ; en particulier, comme nous le souhaitons, l’*ISO 19115* permet de décrire la *portée spatiale d’une ressource* en mode *textuel* ou *cartographique*. Cependant cette norme est permissive sur l’usage de *mots clés spatiaux* et laisse donc l’opportunité de préciser des modalités de description spécifiques à notre contexte. Nous explicitons les liens qu’entretiennent *concepts thématiques* et *spatiaux* et nous proposons un nouveau *modèle conceptuel de données* (*MCD*) qui intègre ces nouvelles *analyses* et *formalisations*. Dans la section 6.2, nous détaillons l’exploitation du *référentiel spatial* dans le *service de métadonnées* à travers la description du *modèle physique de données* (*MPD*) générique que nous préconisons. Nous donnons des exemples de *requêtes spatiales* permises par l’utilisation du *référentiel spatial* pour assister la *saisie* ou la *consultation* des *fiches de métadonnées*. Nous détaillons ensuite l’implémentation de ces recommandations dans le *SGBDR Postgres* avec sa cartouche spatiale *Postgis*. Nous concluons cette section en insistant sur l’apport des spécifications de l’*OGC* pour les *Web Services* qui gèrent de l’*IG*. Dans la section 6.3, nous expliquons comment l’*infrastructure* décrite dans la première partie peut être alimentée par un *inventaire* pertinent des *concepts spatiaux* et *objets géographiques* de référence pour la GIZC en *Languedoc Roussillon*. Nous nous appuyons pour cela à nouveau sur un *état de l’art des approches et travaux de référence* en cours dans ce domaine ainsi que sur des *entretiens* auprès de nos partenaires. Nous proposons enfin l’intégration des *concepts spatiaux* et *objets géographiques* inventoriés ainsi que celles des *relations* qu’ils entretiennent pour alimenter l’*ontologie* proposée dans le chapitre 5.

6.1 Intérêt d’une description spatiale contrôlée

Nous avons vu que le *support cartographique* peut aider à la résolution des *problèmes terminologiques* (voir partie 3.1.4). Cependant, la *description spatiale*, comme la *description thématique*, peut poser, entre autres, des problèmes d’interprétation (*polysémie* ou *synony-*

mie) et par conséquent, des problèmes de *mutualisation* des *ressources informationnelles* et de la *connaissance* auxquelles elle est associée. Le *contrôle de la description spatiale* est donc nécessaire en plus du *contrôle de la valuation des mots clés* (voir partie 5.1) pour optimiser l'utilisation du *support cartographique*. Toutefois, les modes de gestion dédiés à l'*IG géoréférencée* sont spécifiques et doivent de préférence s'appuyer sur des standards pour proposer des *descriptions spatiales contrôlées*.

En partie 6.1.1, nous montrons que la norme *ISO 19115*, en tant que *norme spécialisée*, apporte un cadre élaboré pour la description des caractéristiques spatiales des *ressources informationnelles* de la GIZC. En particulier nous ajoutons au noyau du *profil Syscolag* les trois sortes d'*éléments de description de l'étendue géographique de la ressource* qui permettent, comme nous le souhaitons, des *descriptions spatiales* en *modes textuel* et *graphique* (respectueuse des standards de l'*OGC*) et présentent donc beaucoup d'intérêt pour mettre en avant la dimension spatiale prépondérante dans le contexte de la GIZC. Cependant l'emploi des *mots clés spatiaux* reste ambiguë dans l'*ISO 19115* et nous proposons, en partie 6.1.2 une clarification de leur utilisation par la formalisation des liens existant entre *concepts thématiques*, *concepts spatiaux* et *objets géographiques* dans notre domaine. Nous intégrons ces analyses dans les *modèles conceptuels et physiques* de notre *base de données spatiales* et illustrons comment cette approche permet les requêtes spatiales de saisie et de consultation du *service de métadonnées*² (*SMD*). Nous présentons finalement les choix techniques effectués pour l'implémentation, en particulier les standards utilisés pour la gestion et la diffusion sur le Web du contenu du *référentiel spatial* que nous allons stocké dans cette *SDI*.

6.1.1 Intérêt de la norme ISO 19115 pour la description de l'étendue géographique des ressources

Les *métadonnées* présentent de nombreux avantages pour l'échange d'information (gestion des *contraintes d'accès* en particulier, richesse et pertinence de la description, voir partie 4.1). La norme *Dublin Core* [41] s'est imposée comme norme de découverte pour l'échange de données numériques (voir partie 4.1.6) mais la composante spatiale fortement marquée dans le cas de la GIZC nous a finalement incité à retenir une norme adaptée à des descriptions plus fines et standardisées de l'*étendue géographique des ressources informationnelles* que nous souhaitons partager.

Nous avons retenu la norme *ISO 19115* (Geographic Information - Metadata) [2] qui est une synthèse des standards pré-existants de *métadonnées pour l'échange d'information géographique* (voir partie 4.2). En particulier, le choix de cette norme est pertinent pour la GIZC car elle offre des descripteurs élaborés de l'*étendue géographique concernée par la ressource* conformes à nos attentes.

Description de l'information géographique dans la norme ISO 19115

La norme *ISO 19115* [2] est constituée de onze *rubriques* dont trois portent explicitement sur la *dimension spatiale de la ressource* (voir Figure 6.1 : rubriques colorées *Reference system*, *Spatial representation* et *Content*) et ne concernent que des ressources de type *information géographique* (en mode *vecteur* ou *raster* : par exemple des *cartes*, *images de télédétection*,...). Elles sont logiquement *optionnelles*.

Par contre, on trouve dans la rubrique *obligatoire Identification*, trois éléments *optionnels* de

²Plus largement cela est possible en pratique pour tout *géocatalogue*.

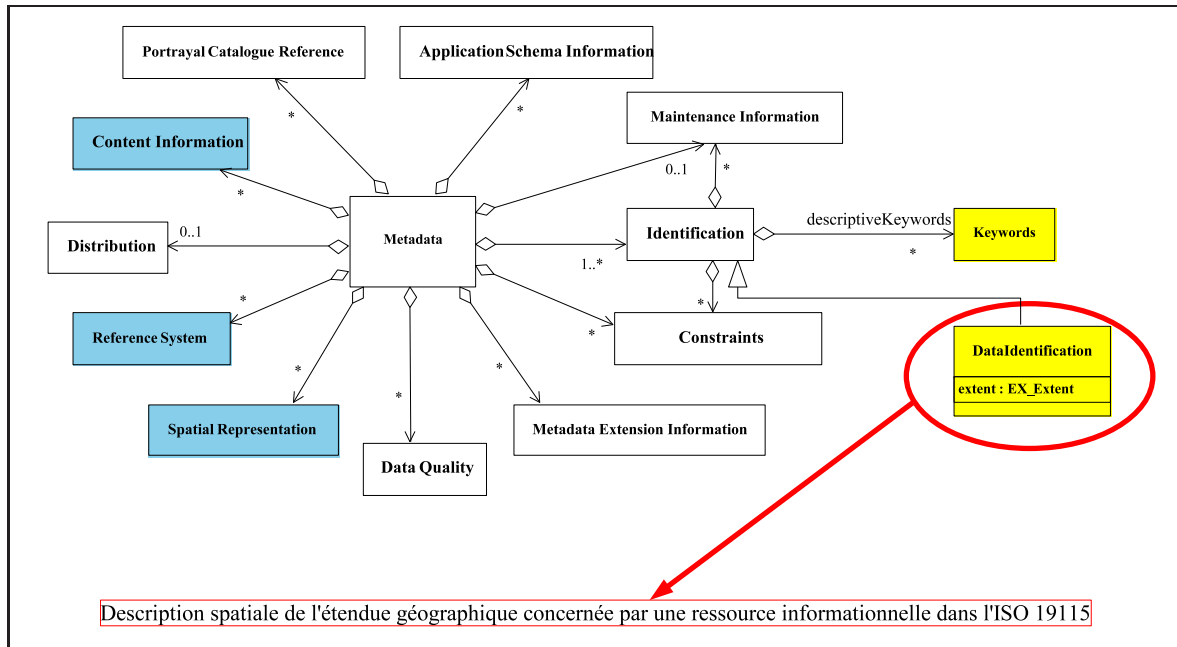


FIG. 6.1 – Les différentes rubriques de l'ISO 19115 d'après [2]

description de l'*étendue géographique concernée par la ressource*. Ces éléments de métadonnées sont présents dans le *DataType EX_Extent* (c'est à dire le type associé à l'élément *extent* de la classe *MD_DataIdentification*, voir Figure 6.1). Bien qu'*optionnels*, nous avons choisi de les rendre tous les trois *obligatoires* dans le profil *Syscolag* (voir partie 4.3.1), et ce, donc, quel que soit le *type des ressources* décrites (*ie* quel que soit le *gabarit*, voir partie 4.3.1). Nous considérons donc ici que la description spatiale concerne l'*information géographique hétérogène*³ selon le sens défini par [119]. Ceci va nous permettre d'effectuer des *requêtes spatiales*, y-compris sur des *fiches de métadonnées* qui ne décrivent pas d'*IG* puisqu'elles contiennent toutes, en plus d'autres descripteurs alphanumériques, ces descripteurs spatiaux désormais obligatoires. Ceci constitue un apport essentiel de notre proposition.

Dans la partie suivante, à travers une description détaillée de ces trois éléments, nous allons voir qu'ils constituent finalement trois manières différentes de décrire un même *objet géographique*.

Les trois descripteurs particuliers de l'étendue géographique dans l'ISO 19115

L'*ISO 19115* permet deux modes de *description de l'étendue géographique de la ressource informationnelle décrite* dans le *service de métadonnées (SMD)* (voir Figure 6.2) :

- un *mode textuel*, par le biais de *toponymes* (ou *mot clés spatiaux*) : *descripteur III* ;

³[...] l'information géographique ne peut pas être uniquement constituée de structures de données manipulables par les SIG. L'information géographique recouvre aussi les analyses, les synthèses et autres interprétations que peuvent faire les utilisateurs tels que décideurs en aménagement, géographes, sociologues ou économistes. L'information géographique est donc contenue dans des ensembles de données factuelles géoréférencées, mais aussi dans des documents structurés, textuels, tabulaires, cartographiques, photographiques, ou composites formés par ces divers éléments cités [...] [119].

- un *mode (carto)graphique*, par le biais de *descripteurs géométriques*⁴ de deux sortes : une *limite géométrique complexe* ou son approximation sous la forme d'un *rectangle englobant* : respectivement *descripteurs I et II*.

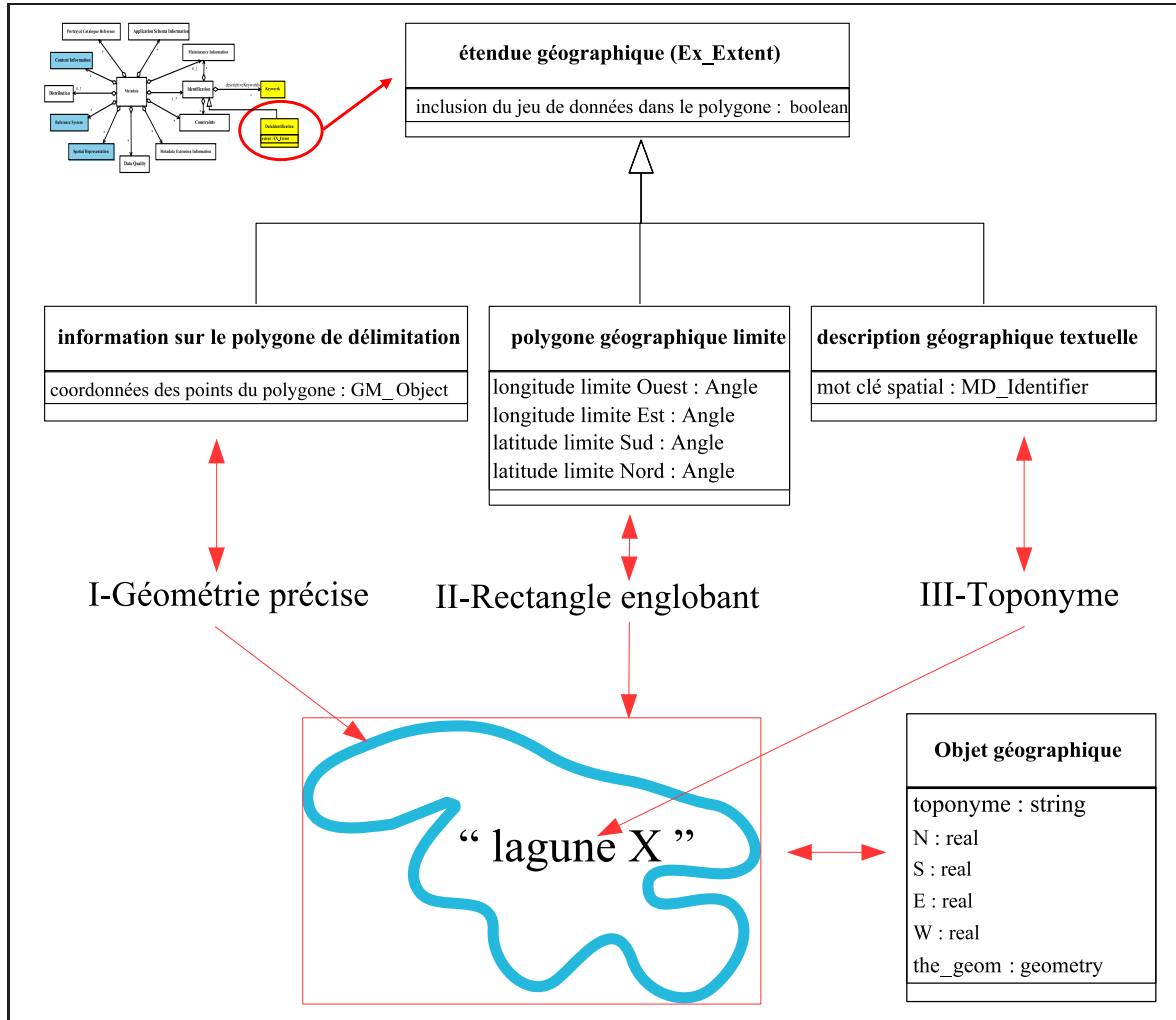


FIG. 6.2 – Illustration des trois types de descripteurs géographiques de l'ISO 19115

De plus, pour ces trois sortes de descripteurs, il convient de prendre en compte les propriétés suivantes :

- *I*, *II* et *III* ne sont pas indépendants. La connaissance de *I* implique la connaissance de *III* et réciproquement (en fait presque étant donné qu'on peut connaître un *toponyme* sans avoir de *descripteurs géométriques (information vectorielle)* associés), et comme *II* se déduit de *I*, alors la connaissance de *I* contient implicitement celle de *II* et *III*. De même, au bémol que nous venons d'évoquer près, la connaissance de *III* contient implicitement celle de *II* et *I*. Enfin, si on utilise *II* comme descripteur, cela sous entend soit que seul *II* convient pour décrire ce type de ressource, soit qu'il n'existe ni de *III*

⁴En pratique ces descripteurs sont géoréférencés par la description d'un *SRID (ou identifiant du système de référence pour la projection)* accolé à la géométrie.

ni de *I* pertinents (qui seraient sinon employés pour la description). Dans ce dernier cas, on peut demander à l'utilisateur, pour combler les lacunes dans les propositions de descripteurs, de proposer un toponyme (*III*) associé et éventuellement une limite plus précise (*I*).

- Bien que le *rectangle englobant* (*II*) puisse être théoriquement considéré comme une approximation d'une géométrie plus précise (*I*), représentation cartographique supposée exacte d'un *toponyme* (*III*), son intérêt n'est pas à remettre en cause. En effet, la pertinence respective de ces trois sortes d'éléments est fonction des types de ressources qu'ils décrivent. Pour certaines (une *carte d'un département* par exemple), le descripteur idéal sera un *toponyme* et/ou la *géométrie précise* associée (le *département* en question). Pour d'autres, le *rectangle englobant* convient mieux (typiquement l'*imagerie satellitale* n'est pas associée à des *géométries précises* mais bien à une zone dépourvue de nom (ie sans *toponyme*) d'emprise au sol déterminée par les caractéristiques de la prise de vue).

Ces propriétés justifient notre interprétation des recommandations de la norme *ISO 19115* qui consiste à choisir de rendre ces trois descripteurs spatiaux *obligatoires* en considérant qu'ils correspondent à *trois différentes représentations d'une même entité géographique* que nous appelons *Objet géographique* dans la Figure 6.2. Par la suite, lorsque nous employons le terme *Objet géographique*, nous faisons donc référence à cette représentation au sein de diagramme (*UML*) sous forme de classe présentant les trois types de *descripteurs spatiaux* de l'*ISO 19115* comme attributs et sur laquelle nous pouvons nous appuyer, en pratique, pour décrire l'*étendue géographique de la ressource* de manière normalisée.

Retenir ces éléments est un choix que nous revendiquons mais leur exploitation opérationnelle est trop contraignante si l'on n'assiste pas l'utilisateur dans ses démarches de saisie et/ou de recherche. En effet, il est difficile de demander à un utilisateur de saisir l'étendue géographique dans le *mode cartographique* sous la forme de stockage demandée par l'*ISO 19115*. Il doit en effet saisir la *géométrie en coordonnées géographiques en degrés décimaux* sous un format texte identique à celui utilisé dans les *bases de données spatiales*⁵ (une suite de points délimitant le *rectangle englobant* ou la *limite précise*, voir illustration en Figure 6.14). Exemple de géométrie de la *lagune de Leucate* :

```
SRID=102582; MULTIPOLYGON( ( ( 657459.884610776 1755250.4630343 , 657534.735897292 1755306.60149919 , 657565.923933341 1755250.4630343 , 657678.200863115 1755200.56217662 , 657572.16154055 1755025.90917475 , 657503.547861244 1755019.67156754 [...] 657459.884610776 1755250.4630343 ) ) ).
```

Aussi avons nous opté, côté utilisateur, pour l'emploi d'une *interface cartographique* plus conviviale qui, insérée dans notre *SMD*, facilitera la *saisie* de ces descripteurs, et à terme, la *recherche de métadonnées* (nous détaillons cela en partie 7.1.2).

Nous avons clarifié notre interprétation des recommandations de l'*ISO 19115*, concernant l'usage des descripteurs de l'*étendue géographique concernée par la ressource*, il subsiste néanmoins une ambiguïté dans cette norme que nous proposons de résoudre en formalisant les liens entre les *concepts thématiques* et *spatiaux*.

⁵En 1997, l'*Open GIS Consortium* a publié un document intitulé *OpenGIS (R) Simple Features Specifications SFS for SQL* : ce document propose plusieurs concepts pour ajouter le support des données spatiales aux serveurs de base de données *SQL*. En particulier, l'*OGC* définit les standards pour les types d'objets contenus dans les *SIG* et les fonctions requises pour les manipuler. Une illustration de la forme du stockage de la *géométrie* d'un *objet géographique* sous forme de caractères alphanumériques est donnée par la suite en Figure 6.14.

6.1.2 Analyse et formalisation du lien entre concepts thématiques et spatiaux

Dans un premier temps, nous illustrons par un exemple l'ambiguïté de l'*ISO 19115* qui tolère l'emploi de *mots clés spatiaux* identiques pour valuer deux éléments de métadonnées distincts. Nous proposons en réponse un *modèle conceptuel* qui intègre les différents *descripteurs thématiques et spatiaux* et explicite la nature des *relations* qu'ils entretiennent. Il est ainsi possible de proposer un contrôle supplémentaire de la valuation.

Ambiguïté de l'ISO 19115 pour l'emploi de mots clés spatiaux

Nos analyses sur les *indexations des ressources informationnelles* de notre domaine (grâce à l'étude des *mots clés* d'un *jeu de références bibliographiques* fourni par nos partenaires) ou des *entretiens* auprès de nos partenaires montrent l'emploi systématique de *mots clés* du type : "*activités conchyliques, printemps, pollution, nitrate, étang de Thau*". Or nous avons vu, en partie 5.1.1, qu'il existe différentes sortes de *mots clés*⁶ que nous avons représentées en Figure 5.1 par des spécialisations de la classe *concept*. Les valeurs des *mots clés* cités en exemple appartiennent à ces différentes catégories : *activités conchyliques, pollution* et *nitrate* sont des *mots clés thématiques* alors que *printemps* est un *mot clé temporel* et qu'*étang de Thau* est un *mot clé spatial* (un des *individus* représentant le *concept spatial étang*) qui peut être représenté *cartographiquement* en tant qu'*objet géographique*.

Or, nous avons vu en partie 6.1.1 que, en accord avec l'*ISO 19115* et avec le *profil Syscolag* (présenté en partie 4.3.1), chaque *fiche de métadonnées* est indexée d'une part avec des *mots clés* (dans la classe *Keywords*) et d'autre part avec différents *descripteurs de l'étendue géographique* (dans la classe *DataIdentification*), en particulier des *toponymes* (descripteur *III*) (voir Figures 6.1 et 6.2).

Dans le cas de notre exemple, il est alors intéressant de considérer la place particulière occupée par les *mots clés* du type *étang de Thau* qu'un utilisateur peut à la fois renseigner en tant que *mot clé spatial* (*concept* au niveau individuel, *individu* représentant le *concept étang*) mais aussi comme *descripteur textuel* de l'*étendue géographique*, ie le *toponyme (III)*, associé à un *individu* qui représente un *objet géographique*.

Afin d'empêcher cette confusion et la redondance d'information possible dans l'*ISO 19115*, nous avons décidé de réserver l'emploi de *toponymes* comme élément de *description textuelle* de l'*étendue géographique* de la ressource et non pas au niveau de l'élément *mot clé* en tant que *mot clé spatial*.

Cependant cela ne signifie pas que l'usage de *concepts spatiaux* est impossible pour renseigner l'élément *mot clé*. C'est ce que nous expliquons dans la partie suivante dans laquelle nous clarifions la nature du *lien* entre *concept thématique* et *concept spatial* ainsi qu'entre *concept spatial* et *objet géographique*.

Le lien entre concepts thématiques et spatiaux

Si on prend l'exemple des *SIG outils*, l'usage veut qu'un *objet géographique* (par exemple celui qui correspond à la *Commune littorale de Gruissan*) soit géré au sein d'une *couche* (ou *Layer*) portant le nom du *concept* qui exprime la *nature thématique de l'ensemble des*

⁶L'*ISO 19115*, sans aller dans le détail, permet de distinguer différents types de *mots clés* grâce à l'élément *type* (de la classe *MD_Keywords*) valué par les éléments contrôlés du *CodeList MD_KeywordTypeCode* dans la rubrique *Identification Information*.

objets similaires qu'elle regroupe sous forme cartographique (dans notre exemple : le *Layer Communes littorales*, voir illustration en Figure 6.3) complété par des données thématiques sous la forme d'un *enregistrement associé dans une table attributaire*.

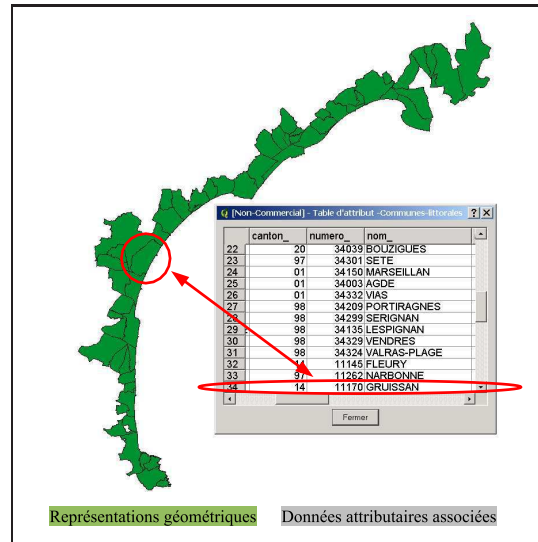


FIG. 6.3 – Structure graphique et tabulaire de l'information géographique dans un SIG-outil

En généralisant ces analyses, cela signifie que les *objets géographiques* sont des *instances* de *concepts thématiques* particuliers qui peuvent donner leur nom à un *Layer*. Nous appelons *concept spatial* cette sorte de *concept thématique* (par exemple : *étang*, *département*, *commune*...).

Une nouvelle relation est introduite entre la classe *concept spatial* et ses représentations graphiques possibles représentées par la classe *Objet géographique*. Ce lien traduit, dans notre diagramme, le passage du *concept (spatial)* à l'*objet (géographique)* : soit de l'*intension*, (par exemple, le concept spatial *étang*), à l'*extension*, (le toponyme *étang de Thau*).

Ces analyses sont formalisées dans la Figure 6.4 qui spécifie la Figure 5.1 et annonce la structure de notre *référentiel spatial*.

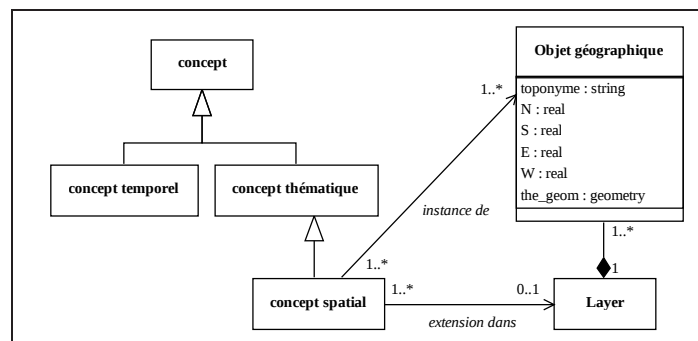


FIG. 6.4 – Lien entre concepts thématiques et spatiaux

Comme nous l'avons dit, l'idée (exprimée par la Figure 6.4) est que :

- à chaque *concept spatial* correspond un et un seul *Layer*, c'est un cas idéal et la réalité peut être différente (il peut ne pas y en avoir, ce qui met en évidence un besoin d'*IG*),
- à un *Layer* correspond un ou plusieurs *concepts spatiaux*, en tenant compte de la *synonymie* des *concepts spatiaux* (par exemple, *lagune* et *étang* sont deux synonymes auxquels correspond un *Layer* unique),
- un *Layer* est une couche d'*IG* composée d'un ou plusieurs *Objets géographiques* de même nature thématique,
- à un *objet géographique* peut correspondre un ou plusieurs *concepts spatiaux* (ceci est dû à la *relation sémantique* de *polysémie* existante entre *concepts thématiques*) et l'ensemble des *objets géographiques* d'un *Layer* constituent / représentent l'*extension* d'un ou plusieurs *concepts spatiaux*,
- à un *Objet géographique* ne peut correspondre qu'un seul *Layer*.

L'usage des *descripteurs spatiaux obligatoires* dans le profil *Syscolag* est désormais clarifié par la formalisation de nouveaux liens entre les *concepts thématiques* et *spatiaux*, essentiels pour l'*indexation des ressources informationnelles* et le *partage des connaissances* dans notre domaine. Nous proposons donc de revenir sur le *MCD* présenté en partie 5.4.2 afin d'y intégrer ces nouvelles *analyses* et *formalisations*.

Modèle conceptuel adapté aux contraintes du domaine

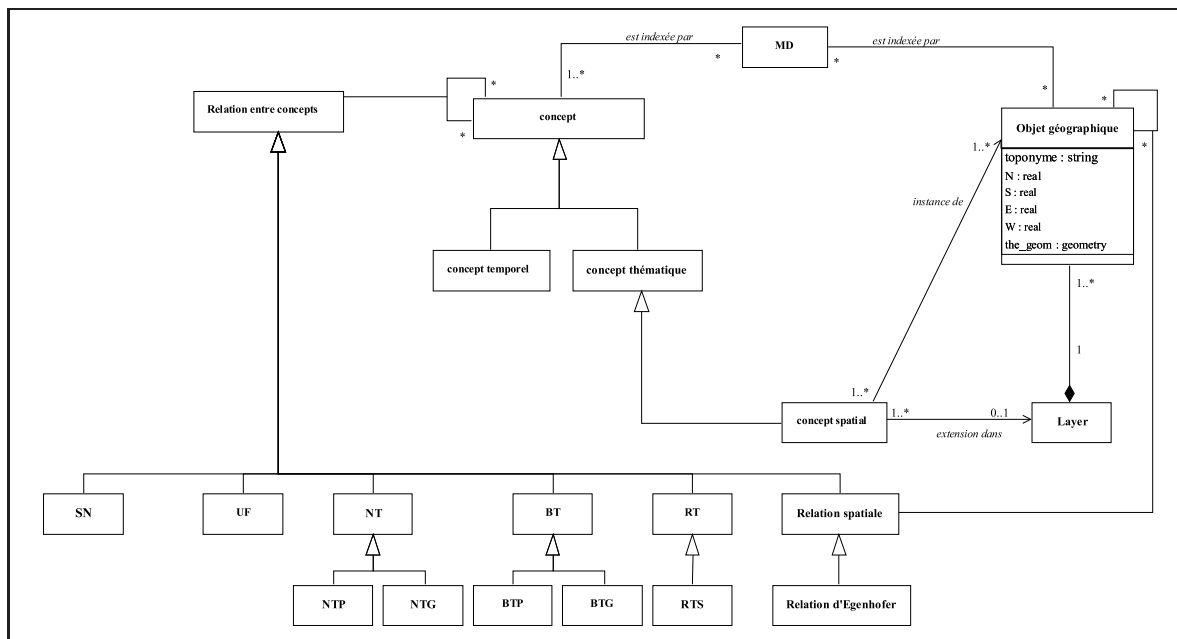


FIG. 6.5 – Modèle conceptuel des référentiels thématiques et spatiaux

Pour assurer la cohérence des *descriptions* de ressources et la qualité des *recherches* dans les *fiches de métadonnées*, il est indispensable de proposer à l'utilisateur un *référentiel sémantique* qui permet de clarifier et contrôler l'emploi des *concepts thématiques* et *spatiaux* (et *objets géographiques* associés). La Figure 6.5 montre comment nous avons modélisé conceptuellement ce référentiel au sein même de notre *base de métadonnées*.

Les *analyses* et *formalisations* précédentes sont intégrées dans ce modèle qui précise le modèle présenté en Figure 5.34(a). On retrouve les classes *MD* (qui indique le lien avec la norme *ISO 19115*) et *concept* ainsi que la classe supplémentaire *Objet géographique*. Ce modèle de *référentiel sémantique* est toutefois indépendant du jeu d'*éléments de métadonnées* choisi pour implémenter la *base de métadonnées*. Son utilisation en tant qu'application indépendante permet la mise en place d'un *langage commun* interprétable par le système et par les acteurs (nous détaillons ceci en partie 5.1).

Afin de mieux adapter le *référentiel sémantique* à la nature des *concepts spatiaux* qu'il intègre désormais, nous proposons d'enrichir la nature des *liens sémantiques* entre concepts, en ajoutant aux *relations ontologiques* (au sens de Gruber [84], voir détail en partie 5.1.4), des *relations spatiales* qui existent entre les *concepts spatiaux* par les *objets géographiques* qui leur sont associés [14]. En particulier les *relations spatiales* utilisées sont celles standardisées par l'*Open Gis Consortium* et plus généralement celles d'*Egenhofer* [148] (voir Figure 6.6).

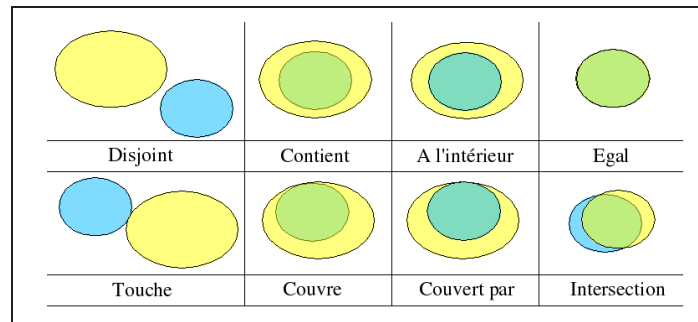


FIG. 6.6 – Ensemble minimal de relations topologiques entre deux objets d'après [118]

- topologiques (disjoints, tangents, intersection, contenus...)
- directionnelles (nord, sud, sud-ouest, sud-est, ...),
- métriques (distance entre les centres de gravité de deux objets).

En guise de synthèse sur ces aspects d'*inventaire* et de *structuration* des *concepts* que nous avons plus largement abordés en partie 5.1.4, nous proposons, en complément, une discussion sur la notion d'*ontologie spatiale*.

Ontologie spatiale

L'*inventaire* des *concepts spatiaux* nécessaires à la formalisation des problèmes de GIZC ainsi que leur *structuration* permet simultanément d'alimenter le *référentiel sémantique* et de l'enrichir de nouveaux *concepts* et de *relations*. Le sous-ensemble qui exploite les *concepts spatiaux*, *Layers*, *objets géographiques*, les *relations ontologiques* et les *relations spatiales* constitue ce que nous appelons le *référentiel spatial* (voir Figure 6.7).

L'*ontologie* que nous avons mise en place sur la base de ce *référentiel sémantique* est donc également enrichie et nous pouvons désormais la qualifier d'*ontologie spatiale* (selon le sens donné par [109, 15]) dédiée au *domaine des objets géographiques de référence pour la GIZC* (voir Figure 6.8).

En effet, les *concepts spatiaux* entretiennent à la fois des *relations spatiales* et des *relations ontologiques* héritées de la classe générale *concept* (voir détail dans le chapitre 5).

Il convient, en outre, de relever que l'introduction de *relations spatiales*, autres que celles

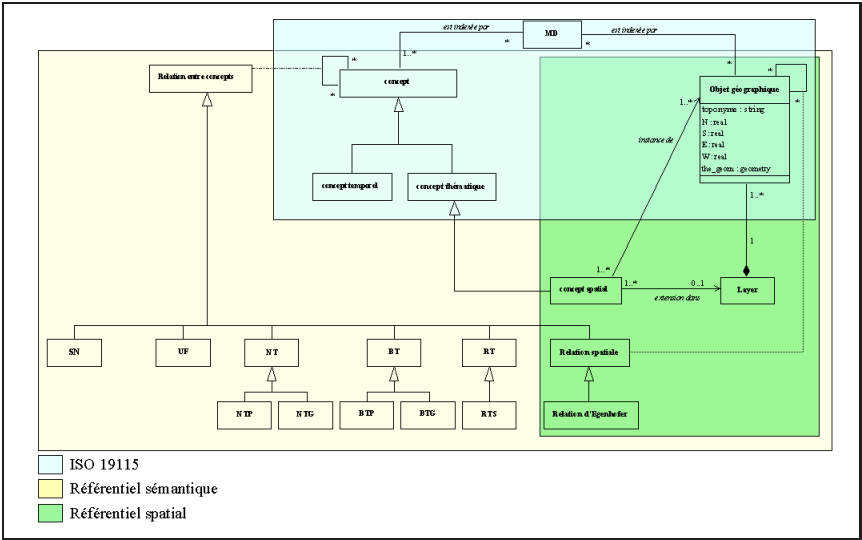


FIG. 6.7 – Synthèse des relations entre l'ISO 19115 et les référentiels sémantique et spatial

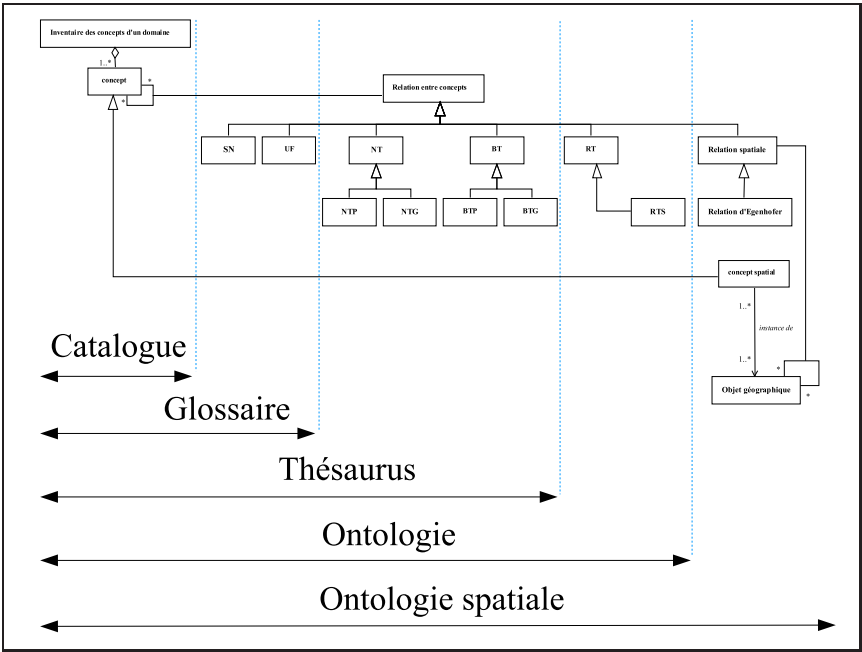


FIG. 6.8 – Evolution de l'ontologie en *ontologie spatiale*

qui utilisent le *géoréférencement* (basée sur une approche euclidienne) : à gauche de ou à droite de par exemple, offre une perspective d'optimisation des traitements d'images conjugués à l'utilisation d'une *approche orientée objet* [72] qui commence à arriver dans les logiciels commerciaux.

Concrètement, cela signifie qu'il est par exemple possible d'utiliser les *relations topologiques* entre *concepts spatiaux* pour résoudre une recherche de *métadonnées*. Par exemple, pour une requête définie sur le critère "*lagune*", il est possible, en plus des *expansions de requêtes* qui exploitent les liens classiques d'*ontologie*, d'utiliser celles basées sur des *critères topologiques* du *Layer lagune* associé (*extension du concept lagune*) pour prendre en compte les fiches de métadonnées dont les descripteurs géométriques sont inclus dans les *lagunes* comme, par exemple, les *concessions conchylicoles*.

Synthèse section 6.1

- La maîtrise sémantique de la *description spatiale* en modes *textuel* et *géométrique* des *ressources informationnelles* de la GIZC est essentielle pour le *partage d'information et de connaissance* dans ce domaine.
- La norme *ISO 19115* propose des *descripteurs spatiaux* pertinents et standardisées (conformes aux recommandations de l'*OGC*). En rendant ces descripteurs *obligatoires* dans notre *profil*, nous améliorons la *qualité des descriptions et des recherches de métadonnées*, mais le mode de valuation demandé par l'*ISO 19115* est trop complexe pour les utilisateurs.
- Nous explicitons les liens entre *concepts thématiques* et *spatiaux* ainsi qu'entre *concepts spatiaux* et *Objets géographiques* afin de lever une ambiguïté de l'*ISO 19115*. Nous intégrons ces nouvelles relations dans notre *référentiel sémantique*.
- Le *MCD* intègre désormais la possibilité de considérer les *relations spatiales* existantes entre les *concepts spatiaux* et *objets géographiques* associés et constitue un *référentiel spatial*.
- L'*ontologie* qui exploite le contenu du *référentiel sémantique* peut donc être qualifiée d'*ontologie spatiale*.
- Nous pouvons utiliser ce modèle pour mettre en place un *référentiel spatial* exploitable pour guider la gestion du *service de métadonnées* et par les *acteurs* de la GIZC.

Nous allons à présent décrire l'*implémentation* que nous proposons d'associer à ces conceptualisations. Nous présentons le *modèle physique de données* qui en découle et décrivons les *standards* utilisés.

6.2 Lien entre référentiel spatial et le service de métadonnées

Nous avons montré en partie 4.3.3 qu'il est souhaitable de gérer les *descripteurs spatiaux* sur l'*étendue géographique de la ressource* dans des *tables* spécifiques. Nous décrivons à présent dans cette section la traduction des spécifications *UML* du *MCD* présenté en partie 6.1.2 en vue d'une *implémentation* en *SQL*. Le *modèle physique de données* (*MPD*) qui en découle et son potentiel pour permettre des *requêtes spatiales* sur les *fiches de métadonnées* sont commentés

ainsi que les choix techniques réalisés pour l’implémentation. En particulier nous justifions le choix, parmi ceux possibles, du *SGBDR Postgres/Postgis* et son intérêt compte tenu de nos objectifs (adaptation au *MPD*, respect des spécifications de l’*OGC* et, en particulier, application dans un *environnement de développement libre*).

6.2.1 Modèle physique de données associé au référentiel spatial

Le *MPD* tel que nous le présentons dans cette partie est générique (il est indépendant du *SGBDR* choisi dès lors que celui-ci permet la gestion de données spatiales).

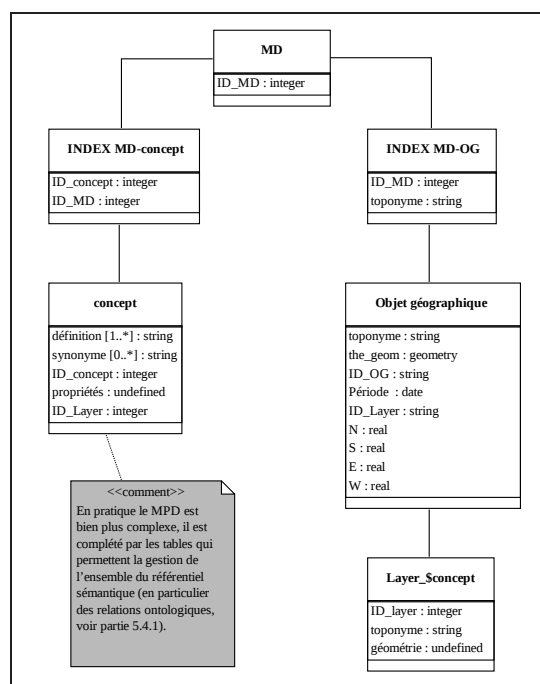


FIG. 6.9 – Modèle physique simplifié de données issu du modèle conceptuel

Conformément au *MCD* présenté en Figure 6.5, la Figure 6.9 matérialise le fait qu’à une *fiche de métadonnées* est décrite par des *concepts spatiaux* (associés au *concepts thématiques* issus de l’*ontologie* que nous proposons en partie 5.2) et par des *Objets géographiques*. Ceci a été détaillé en section 6.1. La *description spatiale* dans le *service de métadonnées* fait donc également intervenir le *MPD* du *référentiel sémantique* présenté en fin de chapitre 5 en Figure 5.35. Nous nous contentons donc de décrire, dans les parties suivantes, les spécificités du *référentiel spatial*, sous-ensemble du *référentiel sémantique*.

Réalisation du schéma

L’interprétation du modèle *UML* présenté en Figure 6.5 implique la mise en œuvre des tables suivantes :

- la table *MD*, qui correspond à la classe *MD*, contient les identifiants des fiches de *MD* dans notre *SMD*,
- la table *concept* correspond à la classe *concept*. Elle contient les *descripteurs textuels* qui ont été utilisés pour indexer les ressources (les *concepts* en général et *concepts spatiaux*

- en particuliers avec lesquels l'élément *mot clé* de l'*ISO 19115* a été renseigné) *inventoriés* et *structurés* dans le *référentiel sémantique* (ou *ontologie*),
- la table *Objet géographique* correspond à la classe *Objet géographique*. Elle contient les *descripteurs spatiaux* qui ont été utilisés pour indexer les ressources (qui correspondent à ceux recommandés par l'*ISO 19115* : *descripteurs géométriques* et *toponymes*, respectivement *I*, *II* et *III*, voir partie 6.1.1). En pratique cette table n'est pas obligatoire puisqu'elle introduit une redondance d'information. Nous justifions son existence et décrivons comment nous remplissons cette table par la suite,
- la table *INDEX MD-concept* permet de gérer la *relation n-n* entre la table *MD* et la table *concept*, avec les identifiants *ID_MD* et *ID_concept*,
- la table *INDEX MD-OG* permet de gérer la *relation n-n* entre la table *MD* et la table *Objet géographique*, avec les identifiants *ID_MD* et *ID_OG*,

Une particularité du schéma vient du fait que les *Objets géographiques* sont issus de *couches* importées depuis des *SIG-outils* qui, généralement, gèrent séparément les attributs textuels et graphiques tel que cela a été présenté en Figure 6.3.

Lors de l'importation d'un *Layer*, celui-ci donne naissance à une *table spatiale* unique portant le nom du *concept spatial* correspondant (voir par exemple le *Layer Communes littorales* dans la Figure 6.10).

Une *méta-table* enregistre tout nouvel import sous la forme d'un enregistrement qui référence le nom du *concept spatial* associé à la *table spatiale* générée.

A présent nous revenons sur la table *Objet géographique* proprement dite, évoquée précédemment qui ne comporte que les enregistrements en provenance des *tables Layers-concepts* utilisées pour l'indexation des ressources.

Intérêt de la table *Objet géographique*

Malgré quelques contraintes de remplissage, la table *Objet géographique* se justifie par une amélioration des *requêtes spatiales*.

Justification de la table *Objet géographique* Conformément au *MCD*, les *descripteurs spatiaux* de l'*ISO 19115* sont regroupés sous forme d'*attributs* de la *table spatiale* *Objet géographique*. Cette dernière répertorie uniquement l'ensemble des *Objets géographiques*⁷ (proposés par les différentes tables spatiales générées par l'importation des *Layers* grâce à un *utilitaire d'import* dans un *SGBDR à cartouche spatiale*) (voir Figure 6.10) qui ont effectivement servi à indexer des *métadonnées*.

L'avantage de cette approche est que seuls les *descripteurs spatiaux* qui ont servi à l'indexation (lors de la saisie de métadonnées) sont présents dans cette table unique grâce à laquelle on va pouvoir exécuter les *requêtes spatiales* de recherche de métadonnées (au lieu d'être réparties et d'avoir à parcourir toutes les tables spatiales issues de l'importation de *Layers* dans lesquelles les *objets géographiques* qui servent effectivement à indexer la métadonnée sont mélangés à d'autres inutiles pour la résolution des requêtes). En dupliquant la partie des objets des *Layers* qui correspond aux descripteurs de l'*ISO 19115* dans une table qui leur est consacrée et donc moins volumineuse, le *temps de parcours* et donc le *temps de réponse* du système est amélioré au prix d'une légère redondance d'information sans incidence notable sur l'espace disque consommé.

⁷Cette table comporte donc potentiellement une *collection* d'*objets géographiques* de différentes natures.

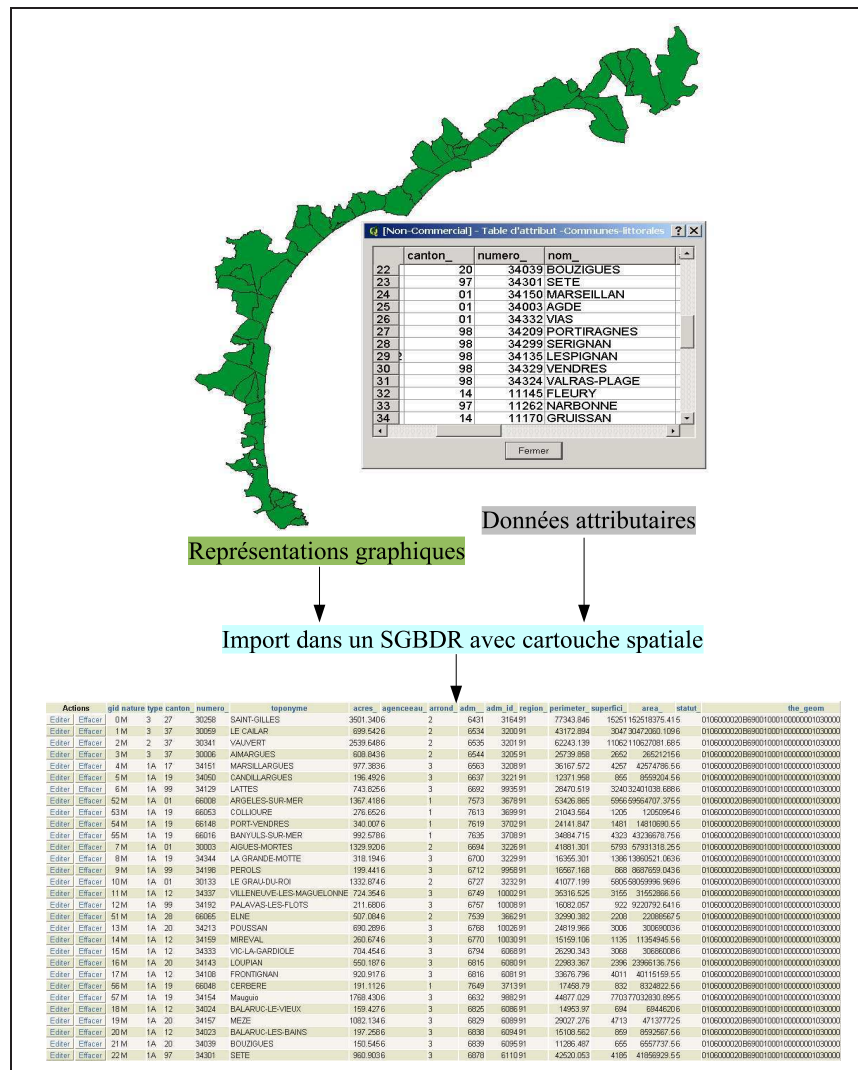


FIG. 6.10 – Structure graphique et tabulaire de l'information géographique

Contraintes de remplissage de la table *Objet géographique* En pratique, lors de la saisie d'une fiche de métadonnée, l'utilisateur choisit un (ou plusieurs) *Objet(s) géographique(s)* issu(s) des *tables spatiales (Layers)* qui ont été importées (*uploadées*) dans la *BDS*.

Par exemple s'il choisit l'*Objet géographique étang de Thau* dans le *Layer étang* :

- si cet *objet géographique* n'a pas déjà servi à indexer d'autres métadonnées, l'objet avec les attributs qui correspondent aux descripteurs de l'*ISO 19115* sont alors automatiquement transférés depuis la *table spatiale* (le *Layer étang*) vers la *table Objet géographique* où ils sont stockés sous la forme d'un nouvel enregistrement. Un nouvel enregistrement est également ajouté à la *table jointure MD-OG*.
- si cet *objet géographique* est déjà présent dans la *table Objet géographique* (ie qu'il a déjà été utilisé pour indexer une ou plusieurs autres fiches de métadonnées) nous ajoutons seulement un nouvel enregistrement à la *table jointure MD-OG* (afin de ne pas stocker plusieurs fois dans la *table Objet géographique* le volumineux attribut qui décrit

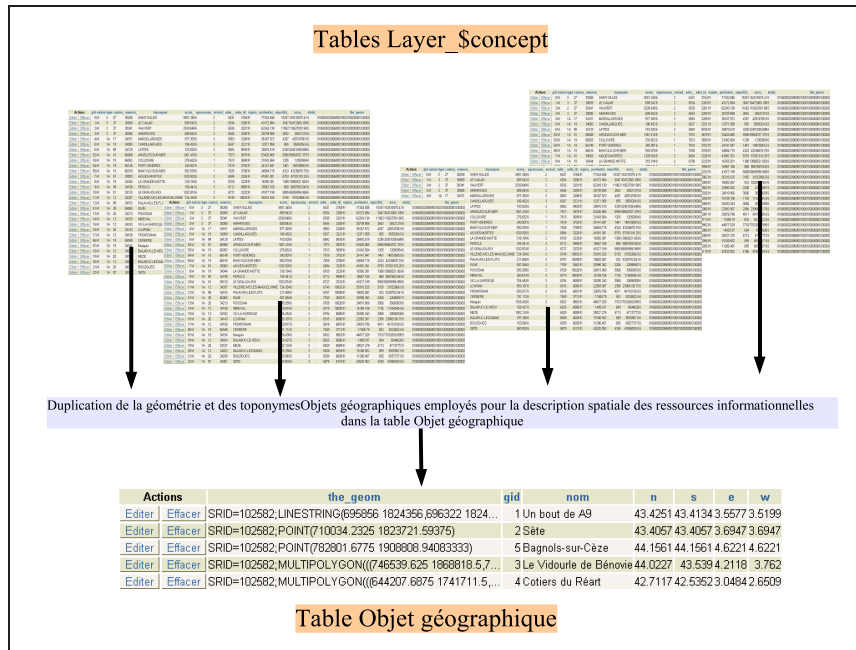


FIG. 6.11 – Duplication des géométries et toponymes des enregistrements des tables *Layer_\$concept* ayant servi à la description spatiale d'une ressource

la géométrie⁸).

Ainsi, de manière transparente et en mode *cartographique*, nous permettons à l'utilisateur de choisir parmi les *objets géographiques* des *Layers* du *référentiel spatial* (du type *Layer_\$concept*) celui ou ceux qui seront utilisés pour générer les descripteurs de l'*ISO 19115* que nous gérons à l'aide des tables *Objet géographique* et *jointure MD-OG* pour les besoins spécifiques du *SMD*.

Nous présentons maintenant quelques utilisations possibles de ce modèle pour résoudre des *requêtes spatiales* sur la *métadonnée* stockée dans notre *SGBDR*.

6.2.2 Les requêtes spatiales désormais possibles sur la métadonnée

Nous avons décrit comment nous avons choisi de rendre obligatoires les descripteurs spatiaux de l'*étendue géographique de la ressource décrite* et ce quel qu'en soit le type (voir partie 6.1.1). Cela signifie que chaque fiche de métadonnée est composée d'un jeu d'*éléments de métadonnées* ou attributs alphanumériques associés à une représentation cartographique. Or ces propriétés permettent d'exploiter les fiches de métadonnées à l'aide d'un *SIG-outil* (logiciel *SIG* ou de la *cartouche spatiale* d'un *SGBD*) qui sont dotés de fonctionnalités d'*analyse* et de *requêtes spatiales*. On pourrait donc dire qu'une des utilisations possibles de cette infrastructure de données spatiales (*SDI*) mise en place est l'exploitation d'un *SIG sur les ressources informationnelles de la GIZC en LR*.

Il est alors possible d'appliquer des fonctions d'*analyse spatiale* à tous les *objets géographiques* utilisés pour indexer la *portée spatiale* des *fiches de métadonnées* mutualisées. Ceci répond bien à une attente du domaine puisque, comme le note [45] : *l'interrogation du catalogage par*

⁸La colonne en charge de ces valeurs porte le nom *the_geom* dans *Postgis*.

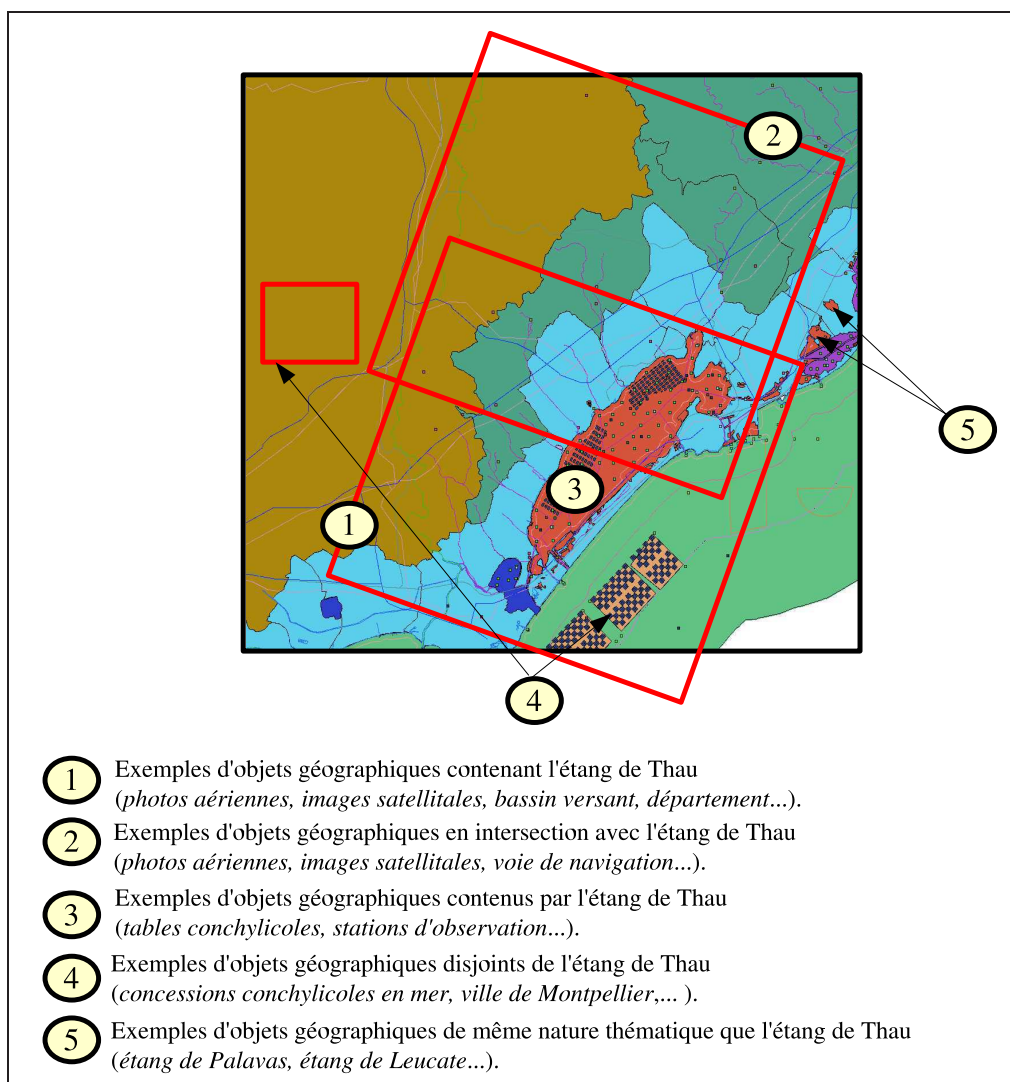


FIG. 6.12 – Exemple de *requêtes spatiales* possibles pour consulter les fiches de métadonnées disponibles sur l'*étang de Thau*

requêtes spatiales, depuis une sélection territoriale sur une carte, est également souhaitable.

La Figure 6.12 représente un ensemble d'*objets géographiques* proposés par le *référentiel spatial* qui ont, potentiellement, tous pu servir à *décrire* des *ressources* dans différentes *fiches de métadonnées*.

Nous présentons quelques exemples, formulés en *langage naturel*, de différentes *requêtes spatiales* possibles qui illustrent l'apport spécifique de notre *référentiel spatial* et de notre *MPD* pour la consultation d'un *SMD* pour la GIZC. Nous nous plaçons dans le cas où un utilisateur souhaite consulter les *fiches de métadonnées* dont la description spatiale correspond à l'*étang de Thau*.

L'expression des différentes requêtes spatiales à effectuer peut être formulée ainsi :

"sélectionner l'ensemble des fiches de métadonnées dont l'étendue géographique a été décrite

par le toponyme *étang de Thau*" ce qui équivaut d'après le modèle proposé à "*sélectionner l'ensemble des enregistrements de la table *Objet Géographique* dont la géométrie est contenue dans l'étang de Thau*".

"*sélectionner l'ensemble des fiches de métadonnées dont l'étendue géographique a été décrite à l'aide d'objets géographiques qui sont contenus dans l'étang de Thau*" ce qui équivaut d'après le modèle proposé à "*sélectionner l'ensemble des enregistrements de la table *Objet Géographique* dont la géométrie est contenue dans l'étang de Thau*". L'expression soulignée dans la requête correspond à une *fonction spatiale*, elle peut donc être remplacée par d'autres expressions telles que : *qui contiennent, qui ont une intersection avec... qui ont la même nature thématique que*.

Ces requêtes permettent de sélectionner les différentes catégories d'objets distingués dans la Figure 6.12, en effet, selon les critères de recherche principaux de l'utilisateur (par exemple le type de ressource recherchée) la nature de la fonction spatiale la plus pertinente pour la requête varie. On peut, lorsque le profil de l'utilisateur le permet lui soumettre le choix de la fonction spatiale qui lui paraît le plus adaptée. On peut bien sûr extrapoler l'exemple sur une requête portant simultanément sur plusieurs objets, ce qui complique les choix possibles et la restitution des résultats.

De manière comparable, si l'utilisateur utilise un *rectangle englobant* pour sa requête de consultation les possibilités sont identiques exceptée bien sûr les recherches sur critères thématiques (fiches indexées par un même toponyme, ou un objet géographique de même nature thématique...)

On peut noter l'existence de *fonctions spatiales* intéressantes pour notre domaine d'application comme celles qui traitent du voisinage et permettent par exemple de sélectionner des *Objets Géographiques* contenus dans un rayon défini autour de l'*Objet géographique* désigné par l'utilisateur.

Une fois les requêtes exécutées, on sélectionne les identifiants des fiches de métadonnées associées à ces *Objets géographiques* grâce à la table *jointure MD-OG* et on présente la liste (en modes *textuel* ou *cartographique*) des fiches de métadonnées qui correspondent à la *requête spatiale* exécutée, par exemple dans un ordre correspondant aux différentes catégories de relations spatiales utilisées dans les requêtes.

A présent nous allons présenter l'intérêt d'implémenter ce modèle dans le *SGBDR Postgres* avec sa *cartouche spatiale Postgis*.

6.2.3 Intérêt de *Postgres* et *Postgis*

Nous décrivons d'abord brièvement l'intérêt d'utiliser des *SGBDR* disposant d'une *cartouche spatiale*, puis nous donnons quelques exemples de traduction en *SQL* des exemples de *requêtes spatiales* donnés dans la partie précédente et désormais possibles sur la *métadonnée* (en complément des requêtes sur les *mots clés*). Nous donnons quelques détails de la mise en pratique de ces recommandations à l'aide du *SGBDR Postgres/Postgis* que nous avons choisi.

Intérêt général des *SGBDR* à cartouches spatiales

D'un point de vue technique, afin de gérer dans un même *SGBDR* des *descriptions textuelles* et *géométriques* auxquelles on peut appliquer des *fonctions d'analyse spatiale*, comme nous le suggérons explicitement dans notre *modèle conceptuel*, il est nécessaire de disposer d'une *cartouche spatiale*. Cette dernière est une extension du *SGBDR* qui permet de gérer des *relations topologiques* entre les *géométries* stockées sous forme alphanumérique.

Avec une telle infrastructure il est donc possible de gérer, en plus de notre base de *métadonnées*, notre *référentiel sémantique* (avec les *concepts thématiques* et *spatiaux* ainsi que les *objets géographiques* qui décrivent l'*extension* de ces derniers). Ainsi, en plus des *critères sémantiques* que nous avons déjà décrits dans le chapitre 5, les *requêtes SQL* sur les *métadonnées* peuvent désormais également intégrer des *critères topologiques* grâce aux *fonctions spatiales* du *SGBDR*.

En outre, nous montrons en partie 7.2.1 comment ce type de *SGBDR* facilite la mise en place d'une *interface cartographique* pour une utilisation plus conviviale des *descripteurs géométriques* dans le *SMD*.

Différents *SGBDR* permettent l'implémentation de ce type d'*infrastructure de données spatiales* depuis les solutions propriétaires (*ESRI* avec *ArcSDE* ou *Oracle* et son *extension spatiale Oracle Spatial Extension*), jusqu'aux solutions libres⁹ (*MySQL* depuis la version 4.1¹⁰ avec *MySQL Spatial Extensions* ou *Postgres* avec sa *cartouche spatiale Postgis*, solution que nous avons retenue).

Intérêt particulier de *Postgres/Postgis* pour implémenter le MPD

Postgres grâce à *Postgis* implémente les *fonctions spatiales*¹¹ et les *types d'Objets géographiques* recommandés par l'*OGC* (dans le cadre de l'implémentation de *SFS*). La version *Postgis 1.0* est actuellement soumise à la certification de l'*OGC*, nous décrirons ultérieurement certains bénéfices de cette standardisation en parties 7.3.2 et 8.2.2.

Pour ce qui est de la gestion des *relations spatiales* entre les *objets géographiques*, il suffit d'importer différents *Layers* dans *Postgis* pour bénéficier d'un ensemble complet, bien que non exhaustif, de *relations topologiques* standardisées par l'*OGC* déjà intégrées au système. Néanmoins, dans les versions à venir, les relations disponibles seront plus nombreuses. En outre, pour la mise en place d'*interfaces cartographiques*, le choix de *Postgis* est judicieux puisque certaines applications libres de droits comme *MapServer*¹² proposent même des développements particuliers pour faciliter leur implémentation avec *Postgis* (nous présentons ceci plus en détail en partie 8.2.2).

De plus, pour ce qui concerne la réalisation pratique présentée en partie 6.2.1, *Postgres* et *Postgis* sont particulièrement adaptés. En effet, *Postgis* propose deux types de tables qui correspondent à celles que nous recommandons (voir Figure 6.5) :

- la table *geometry_columns*. Cette table native de *Postgis* correspond à une *méta-table* qui établit la liste des *Layers* présents dans la *BDS* (à l'aide de l'attribut *f_table_name*, voir Figure 6.13, équivalent à l'attribut *ID_Layer* indiqué dans le *MPD* en Figure 6.9) et *type* de la classe *Layer* respectivement avec ses attributs *f_table_name* valués par des *concepts spatiaux* et *type* (voir illustration en Figure 6.13),
- les *tables spatiales* du type *Layer_\$concept* automatiquement générées par l'utilisation d'outils d'import d'IG de *Postgis*¹³ (voir illustration en Figure 6.14).

Postgis permet donc une mise en œuvre simple du remplissage de notre *référentiel spatial* en faisant correspondre lors de l'importation d'un *Layer* le nom de la *table spatiale* générée

⁹La performance de *Postgres/Postgis* rivalise avec voire dépasse les systèmes propriétaires [173, 31].

¹⁰*MySQL* en est actuellement à la version 5.0.

¹¹Actuellement, les *fonctions spatiales* implémentées dans *Postgis* ne couvrent pas l'ensemble de celles standardisées par l'*Open Gis Consortium*.

¹²<http://mapserver.gis.umn.edu/>.

¹³*Postgis* propose plusieurs outils d'imports ou de chargement de données (*data loader*) selon les formats d'entrée : par exemple *shp2pgsql* pour les fichiers de type **.shape* ou *ogr2ogr* pour les fichiers de type **.map*.

Concepts spatiaux

Nom de la colonne qui gère la géométrie (I) : *the_geom* par défaut

f_table_schema	f_table_name	f_geometry_column	coord_dimension	srid	type	attrelid	varattnum	stats
public	bassin_versant	the_geom		2102582	MULTIPOLYGON	41049	12	NULL
public	ville	the_geom		2102582	POINT	41089	3	NULL
public	contour_lagunes	the_geom		2102582	MULTIPOLYGON	41115	5	NULL
public	resultat	the_geom		2102582	GEOMETRYCOLLECTION	NULL	NULL	NULL
public	departement	the_geom		2102582	MULTIPOLYGON	41024	10	NULL

FIG. 6.13 – Exemple de contenu de la méta-table *geometry_columns* dans *Postgis* : exemples d'enregistrements

Toponyme (III)			Géométrie (I)					
id_dep	nom	num_dep	num_reg	the_geom	n	s	e	w
62	HERAULT	34	91	SRID=102582;MULTIPOLYGON(((736560 1867834,736626 ... 43.9651 43.2147 4.2067 2.5398				
47	AUDE	11	91	SRID=102582;MULTIPOLYGON(((578595 1821535,578665 ... 43.4549 42.6448 3.2458 1.6954				
79	LOZERE	48	91	SRID=102582;MULTIPOLYGON(((664399 1988469,664419 ... 44.9676 44.1131 4.0147 2.9765				
54	GARD	30	91	SRID=102582;MULTIPOLYGON(((737595 1925845,737624 ... 44.4429 43.4715 4.8676 3.2537				
32	PYRENEES-ORIENTALES	66	91	SRID=102582;MULTIPOLYGON(((598942 1744961,598937 ... 42.9143 42.3288 3.1862 1.7237				

FIG. 6.14 – Exemple de *table spatiale* département du type *Layer_\$concept* générée par l'utilitaire d'import *shp2pgsql* dans une base *Postgis*

avec celui d'un *concept spatial* issu de notre *ontologie* (voir 6.1.2).

Par exemple, la commande *shp2pgsql -d 102582 -c etang.shp etang* permet de générer une *table spatiale* *etang* à partir du *Layer* *etang.shp* en utilisant le *système de projection Lambert II étendu* (*SRID=102582*).

En outre, chacune des *tables spatiales* générées dans *Postgis* comprend obligatoirement une *colonne géométrique* *the_geom* (qui correspond au descripteur *I* de l'*ISO 19115* dont se déduit donc *II*, voir Figure 6.13) et quasi systématiquement un *toponyme* associé à la géométrie (descripteur *III* voir Figure 6.14). En effet, il est rare de créer des *Layers* qui ne possèdent pas, en plus de la colonne obligatoire pour stocker l'information sur la géométrie (I), une colonne complémentaire qui associe un nom à la géométrie représentée. Or, comme *II* se déduit de *I*, cela signifie (dans la mesure où cette couche comporte bien une colonne *toponyme* voir partie 6.1.1) que chaque couche – *Layer* – importée dans *Postgis* permet de générer les trois types de *descripteurs spatiaux* recommandés par l'*ISO 19115*. En d'autres termes : à chaque import de *Layer*, on peut produire autant d'*objets géographiques* (comprenant chacun les trois sortes de *descripteurs spatiaux* proposés par l'*ISO 19115*) que d'enregistrements dans le *Layer* importé. Ces *objets géographiques* correspondent à l'*extension* du *concept spatial* qui donne le nom au *Layer*...et constituent autant de valeurs préenregistrées dans le *référentiel spatial* à la disposition des utilisateurs pour décrire une *fiche de métadonnée*.

En conséquence la *table Objet géographique* peut mélanger des enregistrements correspondant à différents types d'éléments géométriques (*point*, *ligne*, *polygone*...). Elle doit donc être déclarée du type *geometry_collections*.

Nous décrivons à présent la traduction des exemples de requêtes donnés en partie 6.2.2 à l'aide des *fonctions spatiales* présentes dans *Postgis*.

Intérêt de *Postgres/Postgis* pour l'exécution de requêtes spatiales

Grâce aux *fonctions spatiales* proposées par *Postgis* nous pouvons couvrir nos besoins pour les *requêtes spatiales* associées à la *saisie de descripteurs spatiaux* (aussi bien pour les fonctions d'*indexation* que pour celles de *consultation* du SMD), tout en respectant les spécifications de l'*OGC*.

Requêtes spatiales de saisie dans le SMD Compte tenu de notre *MPD*, lorsqu'un utilisateur indexe l'étendue géographique d'une ressource lors de la *saisie* d'une *fiche de métadonnée* (voir Figure 7.10), nous devons distinguer deux cas :

- l'utilisateur décrit l'*étendue géographique* par des descripteurs connus du système (*the_geom* ou *toponyme*) et gérés dans les tables du type *Layer_\$concept* (voir partie 6.2.1). Les valeurs des attributs des enregistrements des tables spatiales utilisées (*the_geom* et *toponyme*) sont alors dupliquées dans la table *Objet géographique* (si elles ne sont pas déjà présentes dans les attributs du même nom). Il reste alors à calculer les limites extrêmes *N*, *S*, *E*, *W* sur la valeur de *the_geom*¹⁴ pour obtenir le troisième mode de description demandés par l'*ISO 19115* (*II* ou *rectangle englobant*, voir partie 6.2.1). La *métadonnée* et les descripteurs concernés sont indiqués dans la table *jointure MD-OG*.
- l'utilisateur ne trouve pas de descripteurs appropriés parmi ceux proposés et préfère alors décrire la métadonnée en proposant une nouvelle *géométrie* (pour le moment un *rectangle englobant* compte tenu de la difficulté technique pour proposer l'*édition* en ligne d'*IG*, voir partie 6.2.4). Pour cela, nous récupérons les coordonnées extrêmes *N*, *S*, *E*, *W* en degrés décimaux du *rectangle élastique* dessiné, à la souris, par l'utilisateur dans l'*interface cartographique* puis on les stocke dans les attributs du même nom à partir desquels nous générons¹⁵ la *géométrie* correspondante au *rectangle englobant* (que nous stockons dans l'attribut *the_geom* de la table *Objet géographique*). Pour avoir les trois descripteurs spatiaux, il reste alors à demander le nom associé aux descripteurs *géométriques* (qu'on va stocker dans l'attribut *toponyme*. Ce *toponyme* peut alors indiquer la nature thématique de la *géométrie* dessinée et permettre de combler le manque du *concept spatial* ou *Layer* correspondant dans notre *référentiel spatial* (voir détail en partie 6.3.1).

Les trois types de descripteurs de l'*ISO 19115* sont alors accessibles dans la table *Objet géographique* (voir partie 6.2.1). Grâce à cela, nous pouvons effectuer des *requêtes spatiales* pour consulter les fiches de métadonnées existantes.

Requêtes spatiales de consultation dans le SMD *Postgis* propose un ensemble complet de *fonctions spatiales* standardisées par l'*OGC* qui permet l'exécution de *requêtes spatiales*

¹⁴ Afin de générer à partir de la colonne *the_geom* des tables *Layer_\$concept* les 4 descripteurs correspondant au *rectangle englobant* (*N*, *S*, *E*, *W*) demandés par l'*ISO 19115* dans notre table *Objet géographique* nous utilisons les *fonctions spatiales* suivantes : *YMAX(geometry)* pour générer la valeur de l'attribut *Nord* (valeur maximale des *Y* d'une *géométrie*) et nous procédons de même pour les valeurs des attributs *Sud*, *Est* et *Ouest* (respectivement avec les fonctions *YMIN(geometry)*, *XMAX(geometry)* et *XMIN(geometry)*).

¹⁵ La fonction spatiale de *Postgis* (standard *OGC*) *GeometryFromText(text,[<SRID>])* crée une *géométrie* à partir de texte (suite de coordonnées de points et d'un entier identifiant un *SRID*).

variées.

Les *fonctions spatiales* que nous avons utilisées pour exécuter les *requêtes spatiales* sont les suivantes¹⁶ : *within(geometry, geometry)* qui retourne un *booléen* pour indiquer le résultat du calcul de l'*inclusion* d'une *géométrie* dans une autre (est *inclus* ou *contient* selon l'ordre des géométries) et *intersects(geometry, geometry)* selon le même principe pour l'*intersection*¹⁷. Les traduction en *SQL* des requêtes en *langage naturel* énumérées dans la partie 6.2.2 (voir Figure 6.12) sont les suivantes :

- "SELECT * FROM Objet Géographique WHERE toponyme='étang de Thau',
- "SELECT * FROM Objet Géographique WHERE within(the_geom, X)=1,
- "SELECT * FROM Objet Géographique WHERE within(X, the_geom)=1,
- "SELECT * FROM Objet Géographique WHERE intersects(the_geom, X)=1,
- ...,

Dans ces requêtes *X* = ('select the_geom from table_étang where toponyme = 'étang de Thau') représente la *géométrie* associé à l'*Objet géographique* étang de Thau géré dans la table *Layer_étang* (du type *Layer_\$concept*, voir Figure 6.9).

Le *résultat* de ces requêtes rassemble un ensemble *Objets géographiques* à partir duquel il est possible d'identifier les métadonnées correspondantes (par la table *INDEX MD_OG*, voir Figure 6.9).

Afin d'optimiser l'usage en ligne de cette *base de données spatiales*, nous avons finalement pris en compte d'autres spécifications de l'*OGC* pour les *Web Services cartographiques* qui, en plus d'aider à la *visualisation* voire à l'*édition* de l'*IG* en ligne, permettent l'*interopérabilité* avec d'autres *services Web de données spatiales* implémentant les mêmes spécifications.

6.2.4 Interopérabilité du serveur de données spatiales mis en place

Le *référentiel spatial* que nous avons mis en place permet donc l'*amélioration* de la *saisie* et de la *consultation* des *métadonnées* par son intégration dans le *MCD* qui étend le *référentiel sémantique* par les parties relatives aux *concepts spatiaux* et à la *représentation cartographique standardisée* des *objets géographiques* qui leurs sont associés.

La mise en ligne du *SMD* nécessite la diffusion, via le Web, de l'*IG* collectée auprès des utilisateurs pour être visualisée et éventuellement modifiée à l'aide de navigateurs Web. Or l'*OGC*, en collaboration avec l'*ISO*, propose des spécifications pour les *Web Services* qui permettent la gestion d'*IG* : *WMS* et *WFS*. Ces spécifications s'appuient sur le standard *Geography Markup Language*¹⁸ (*GML*). En particulier, *WMS* présente un intérêt majeur pour permettre à terme l'*interopérabilité* de notre service avec d'autres implémentant la même spécification [77].

WMS

Compte tenu de :

¹⁶D'autres fonctions, non implémentées pour le moment, proposent des recherches intéressantes pour notre domaine d'application. En particulier, on peut citer la fonction *buffer(geometry, double)* qui peut être intéressante pour gérer les requêtes spatiales telle que celle donné du type 4 donnée en partie 6.2.3.

¹⁷Certaines *fonctions spatiales* de *Postgis* comme l'*inclusion* ou l'*intersection* sont cependant encore approximatives. En effet, ces opérateurs travaillent, pour le moment, sur les *rectangles englobants* des *objets géographiques* et non sur la *géométrie*.

¹⁸le *GML* est un syntaxe *XML* qui met à disposition un mécanisme pour la modélisation et l'échange de géodonnées orientées *vecteur*. Elle fait l'objet de la norme *ISO 19136*.

- nos problématiques d'*interopérabilité des bases de données environnementales*,
- l'importance de l'*espace* dans les problématique de gestion environnementale en général et de GIZC en particulier,
- de la prise en compte des travaux du *TC 211* de l'*ISO* et de sa collaboration avec l'*OGC*,
- de la récente normalisation de la spécification *WMS*¹⁹ de l'*OGC* en norme *ISO 19128* [3],

nous proposons d'intégrer les recommandations de cette nouvelle norme pour optimiser l'implémentation d'une *interface cartographique* dans et surtout en dehors de notre *service de métadonnées* afin de valoriser au mieux l'*IG* contenue dans notre *référentiel spatial*.

En effet, par cette spécification de l'*OGC*, basée sur une *architecture client-serveur* traditionnelle, utilisant le protocole de communication HTTP²⁰, il est possible, de mettre à disposition d'autres serveurs et clients *WMS* les *Layers* que nous avons importés dans *Postgis*. *WMS* permet l'*interopérabilité* avec d'autres *serveurs WMS* de *données spatiales* (en mode *vecteur*) [11] ce qui est particulièrement intéressant dans le cadre de la communauté impliquée dans la GIZC. Il devient alors possible pour un *client WMS* d'élaborer une carte composée de couches téléchargées depuis un réseau de *WMS*, grâce à une requête de type *GetMap*²¹ [11].

Pour cela, la requête standardisée de type *GetCapabilities* fournit au client un fichier *XML* d'informations (*métadonnées*) sur les ressources disponibles sur le serveur [134] qu'il est, en outre, possible de compléter par d'autres métadonnées comme celles de la norme *ISO 19115*. Ceci permet d'optimiser la finesse des descriptions sur l'information disponible et, dans notre cas, de mettre à profit les efforts de structuration de notre *service de métadonnées*.

Ce point est capital car il permet de sélectionner à l'aide de tous les critères proposés par l'*ISO 19115* des couches d'*IG* (vectorielles ?) réparties dans des *serveurs WMS*. La complémentarité de ces deux normes est donc évidente et on peut rappeler que *Geonetwork* souhaite dans la prochaine version (voir partie 4.2.3) exploiter cette complémentarité pour la *description*, la *recherche* et la *visualisation* d'*IG* au sein d'une même application.

Ceci peut présenter à terme un intérêt majeur pour rendre interopérables un réseau de serveurs *WMS* dédiés à notre domaine. Par exemple notre serveur *WMS* à l'échelle régionale avec un serveur *WMS* équivalent en région *PACA* par exemple (mais aussi en *catalogne espagnole...*) permettrait de contribuer à la réalisation de cartes à l'échelle de la *Méditerranée occidentale*. Cette possible intégration d'efforts issus de différentes initiatives à différentes échelles s'inscrit tout à fait dans le cadre des recommandations de la GIZC et du *DD*.

De plus la compatibilité de *WMS* est optimale avec les *SGBDR* à *cartouche spatiale* de type *Postgis* (conforme aux recommandations de l'*OGC* concernant *SFS*).

WMS est la spécification primordiale de l'*OGC* pour les *Web services*, par contre elle présente des possibilités limitées du point de vue de l'utilisateur dans la mesure où elle ne permet pas par exemple l'*édition* des couches qu'elle permet de visualiser, ceci est le lot de la spécification *WFS*.

¹⁹ *WMS*, *Web Mapping Service*, depuis sa version 1.3.0 correspond à l'*ISO 19128*, preuve de la bonne complémentarité entre ces deux organismes.

²⁰ Hypertext Transfer Protocol, abrégé HTTP, littéralement "protocole de transfert hypertexte", est un protocole de communication informatique client-serveur développé pour le World Wide Web. Il est utilisé pour transférer les documents (document HTML, image, feuille de style, etc.) entre le serveur HTTP et le navigateur Web lorsqu'un visiteur consulte un site Web.

²¹ Il existe trois types de requêtes pour un serveur *WMS*, deux sont obligatoires (*GetCapabilities* et *GetMap*), la dernière est optionnelle (*GetFeatureInfo*).

WFS

Cette spécification insiste sur l'*édition d'objets géographiques* qui est particulièrement intéressante pour le travail collaboratif sur le Web. En effet, il est particulièrement attrayant, par l'implémentation de cette spécification, de proposer par une connexion Web l'*édition* (la *saisie*, *modification* ou *suppression*) des *objets géographiques* contenus dans notre *base de données spatiales* pour affiner l'*IG* relative à la GIZC [134].

Il existe différents types de *WFS* selon leurs compétences (*Basic* ou *transactions*). Mais cette spécification est en cours de révision à ce jour et son implémentation est encore délicate [11]. En outre, la *visualisation* sur le Web de l'*IG* que nous avons collectée est impérative dans notre contexte pour permettre la *description spatiale* dans notre *SMD* et l'apport de *WMS* est alors indéniable. L'*édition*, quant à elle peut se faire à distance par d'autres moyen que le Web (connexion *IP* utilisant les formats *WKT* ou *WKB*, *ssh*²² à l'aide de clients *SIG* libres par exemple), ce qui rend l'implémentation de *WFS* secondaire, nous justifions cela plus en détail en partie 8.2.2.

De plus nous verrons que nous proposons d'implémenter ces spécifications en restant dans un *environnement de développement libre* pour le *SGBDR* et le *viewer* (voir partie 7.1.2).

Synthèse section 6.2

- Nous avons générer un *modèle physique de données (MPD)* qui permet d'exploiter le lien entre le *référentiel sémantique* et le *service de métadonnées* (comme cela a été présenté en section 6.1).
- Le *MPD* proposé permet l'exécution de *requêtes spatiales* pour l'*assistance* à la *saisie* ou à la *consultation* dans un *service de métadonnées* qui utilise des *descripteurs spatiaux*.
- Nous recommandons d'exploiter ce lien à l'aide d'un *SGBDR* avec une *cartouche spatiale*. En particulier, nous décrivons les avantages de *Postgres* et *Postgis* qui permettent de gérer les *descripteurs spatiaux* selon une forme standardisée par l'*OGC* adaptée aux *descripteurs spatiaux* de l'*ISO 19115*.
- De plus, *Postgres* et *Postgis* permettent une mise en œuvre simple du *modèle physique de données* proposé et disposent de nombreuses *fonctions spatiales* utilisables pour interroger le *service de métadonnées*.
- Comme pour les métadonnées, l'utilisation de la norme *ISO 19128 (WMS)* pour la mise en place de *serveur de données spatiales* apporte une réponse éprouvée à nos besoins et permet l'*interopérabilité* du *Web service* proposé avec d'autres à l'échelle nationale et à l'étranger [79].

Nous venons de décrire l'intérêt de l'*exploitation* du *référentiel spatial* dans le *service de métadonnées* indépendamment de la nature de l'*information spatiale* qu'il va contenir. A présent, nous proposons, d'une part un *état de l'art* sur les *sources de référence* pertinentes pour déterminer la *nature thématique* de l'*information spatiale* à inventorier auprès de nos

²²Pour Unix : Sécuré SHell. shell permettant de se connecter de façon sécurisée sur une machine distante et d'y exécuter des programmes, toujours de façon sécurisée (via l'utilisation de divers algorithmes de chiffrement, et destiné à remplacer telnet, rsh ou rlogin.

partenaires et d'autre part, une *méthode de structuration* qui s'appuie sur les *relations* du *référentiel sémantique* illustré en Figure 6.5.

6.3 Alimenter le référentiel spatial par de l'IG pertinente

Le *référentiel spatial* présenté dans la section précédente ne sera efficace pour *décrire et consulter les données ou les connaissances* relatives à la GIZC en *LR* que s'il est alimenté par un *inventaire de concepts spatiaux* et d'*objets géographiques* aussi pertinent et exhaustif que possible. Cette exhaustivité peut être appréhendée en faisant s'exprimer les besoins des utilisateurs qui rendent compte des *descripteurs spatiaux* dont ils aimeraient disposer et qui devraient donc idéalement exister (la théorie), d'une part, et à l'existant (la pratique), d'autre part.

La réalisation de cet *inventaire* est donc l'objectif primordial et, pour le réaliser, nous pouvons nous appuyer sur les *relations générales* entre *concepts* que nous avons détaillées dans la partie précédente, en particulier sur celles entre les *concepts thématiques* et *spatiaux* et celles entre les *concepts spatiaux* et les *objets géographiques* qui leur sont associés. Nous prenons également en compte les *sources de référence* dans le domaine et les analyses effectuées dans le cadre de notre projet. En particulier l'*ontologie* que nous avons présentée en partie 5.2.3 permet l'intégration physique des concepts inventoriés et de la *connaissance* qui leur est associée.

6.3.1 Méthodologie générale

Nous sommes partis du constat que les utilisateurs caractérisent spontanément l'*étendue géographique* d'une ressource par des *descripteurs spatiaux* auxquels correspondent des *objets géographiques* (et leurs *attributs*) qu'ils ont identifiés dans le territoire concerné (voir partie 6.1.1). Nous avons donc initié l'*inventaire des concepts spatiaux* de référence (puisque les *objets géographiques* leur sont associés implicitement en tant qu'*extension* voir partie 6.1.2) pour la GIZC du Languedoc Roussillon afin d'alimenter le *référentiel spatial* inclus dans le *référentiel sémantique*.

Cependant nous allons voir que ce travail nécessite de partager un *langage commun* pour désigner de manière homogène les différents *concepts spatiaux* et *objets géographiques* de notre *inventaire* (ce qui implique en particulier la gestion de la *polysémie* et de la *synonymie*, voir partie 5.1) et la mise en forme par une *structuration* qui exprime la *connaissance* du domaine. Pour cela nous pouvons nous appuyer, d'une part, sur le *MCD* (présenté en partie 6.1.2) et, d'autre part, sur le *modèle d'ontologie* (présenté en partie 5.2.3).

Inventaire et structuration des concepts spatiaux

Le *référentiel spatial* est un sous-ensemble du *référentiel sémantique*. Ainsi, l'*inventaire des concepts spatiaux* bénéficie non seulement de l'*inventaire des concepts thématiques* et des *relations ontologiques* présentées dans le chapitre 5 mais aussi de *relations spatiales* qui leur sont propres.

Inventaire des concepts spatiaux Compte tenu des modèles que nous avons présentés, si l'*inventaire des concepts thématiques* était exhaustif, alors celui des *concepts spatiaux* le serait également et cette partie n'aurait pas lieu d'être. En effet les *concepts spatiaux* sont une sorte de *concept thématique* (voir partie 6.1.2). Or, en raison de l'ampleur de cette tâche et des

limites floues des domaines thématiques à prendre en compte, ce n'est malheureusement pas le cas et l'inventaire des *concepts spatiaux* peut donc, au contraire, aider à réaliser celui des *concepts thématiques* (puisqu'ils généralisent les *concepts spatiaux*), en exploitant le lien de *généralisation / spécialisation*. Concrètement nous proposons donc de mélanger les *approches top-down* et *bottom-up* [70], respectivement *approches descendantes* et *ascendantes* (voir partie 5.1.4).

Il est possible d'exploiter les différents types de liens présentés dans la Figure 6.4 entre les classes *concept spatial*, *Layer* et *Objet géographique* pour mettre en œuvre l'inventaire des *concepts spatiaux* pertinents dans notre domaine par deux moyens complémentaires :

1. une collecte des *Layers* existants chez les *acteurs*. Ces *Layers*, comme nous l'avons expliqué en partie 6.2.3, regroupent, au moins implicitement, l'ensemble des *descripteurs spatiaux* demandés par l'*ISO 19115*, à partir desquels nous pourrions générer les *objets géographiques* correspondants dans notre *référentiel spatial*. En outre, il faut tenir compte du fait qu'on peut associer à un même *Layer* un ou plusieurs *concepts spatiaux* (en raison de la *polysémie* associée à certains *concepts*) et qu'un *concept spatial* important peut en pratique ne pas avoir de *Layer* produit pour le cartographe (remarques 1 et 2). La collecte même exhaustive des *Layers* n'implique donc pas l'exhaustivité de l'inventaire des *concepts spatiaux* existants. Par cette première méthode, l'inventaire obtenu est donc une vue objective de l'existant (en terme d'*IG* en mode *vecteur*) et par conséquent non exhaustive concernant l'inventaire des *concepts spatiaux* théoriquement indispensables.
2. un inventaire des *concepts spatiaux* et *toponymes* qui doivent figurer dans le *référentiel spatial* idéal, indépendamment de l'existence des *Layers* associés. Ceci permet de dresser un état de l'existant et des priorités pour générer les *Layers* manquants et combler les lacunes.

En pratique, ces deux approches peuvent être concrètement alimentées par l'analyse de plusieurs sources :

- sur la base d'approches et de travaux de référence sur le sujet ainsi que dans le cadre de notre projet (approche écosystémique, voir Figures 6.15 et 6.16(a), exercice de *hiérarchisation du territoire*, mutualisation des *références bibliographiques*...).
- l'utilisation du travail déjà réalisé sur le *modèle générique d'ontologie* (présenté dans le chapitre 5). Il s'agit concrètement d'*étiqueter* les *concepts spatiaux* parmi les *concepts thématiques* déjà inventoriés (répartis dans les classes / objets qui spécialisent les quatre principales classes de *concepts* de notre *modèle* et constituent les *typologies* présentées en partie 5.3).

Les *concepts spatiaux* inventoriés à partir d'autres sources peuvent alors être *intégrés* dans les *typologies* de notre *ontologie* en procédant soit par une approche ascendante (*bottom-up*) soit par une approche descendante (*top-down*) (voir définitions en partie 5.1.4).

Ceci fait, en plus des *relations ontologiques* que nous avons énumérées en partie 5.1.4, les *concepts spatiaux* bénéficient de *relations spatiales* supplémentaires tel que cela est présenté par le *modèle conceptuel* présenté en partie 6.1.2.

Structuration des concepts spatiaux Pour ce qui concerne la structuration des *concepts spatiaux*, en dehors du fait que ces derniers sont une sorte de *concepts thématiques* (les relations de *généralisation* et de *spécialisation* présentes dans l'*ontologie* leur sont donc applicables), les autres types de *relations ontologiques* comme la *synonymie* (*lagune de Thau*, *étang de Thau*...) ou la *polysémie* sont là encore tout à fait pertinentes pour ce type de *concepts*, de

même que les relations d'*agrégation* ou d'*association*. Nous avons amorcé cette démarche en nous appuyant là encore sur l'analyse et la formalisation de l'état de l'art ou des entretiens dans le cadre de notre projet. En particulier, à l'instar des *concepts thématiques*, il s'avère que, selon les disciplines, différentes descriptions (*textuelles* mais aussi *géométriques*) des *objets géographiques* cohabitent au sein du projet *Syscolag*.

Pour ce qui concerne les *relations spatiales* qui leur sont spécifiques : en important des *Layers* dans *Potsgres/Postgis*, comme nous l'avons décrit dans la section précédente (voir partie 6.2.3), nous alimentons l'*inventaire* par autant de *concepts spatiaux* que de couches importées et autant d'*objets géographiques* que d'enregistrements dans l'ensemble des *Layers*. En outre, nous disposons automatiquement de *fonctions spatiales* standardisées par l'*OGC* qui sont des *relations topologiques* utilisables entre les *concepts spatiaux* et *objets géographiques* associés.

Toutefois il est possible d'utiliser d'autres sortes de *relations spatiales* existantes afin de *structurer* les *concepts spatiaux* et *objets géographiques* et d'exploiter pleinement le *modèle conceptuel* de notre *référéntiel spatial*. En particulier, il est envisageable de proposer de nouvelles relations entre *concepts spatiaux* du type : *est voisin de*, *est au dessus de*, *est adjacent*... (voir partie 6.1.2).

Après ce que nous venons de détailler, *collecte* et *structuration* des *concepts spatiaux de référence pour la GIZC*, il ne faut cependant pas occulter que les *objets géographiques* qui décrivent leur *extension* et composent les *Layers* qui leur sont associés nécessitent, pour intégrer un *référéntiel spatial*, un travail commun supplémentaire pour s'entendre sur la qualité de leurs *limites géométriques* (*I* ou *II*).

Variabilité de la géométrie des objets géographiques du système

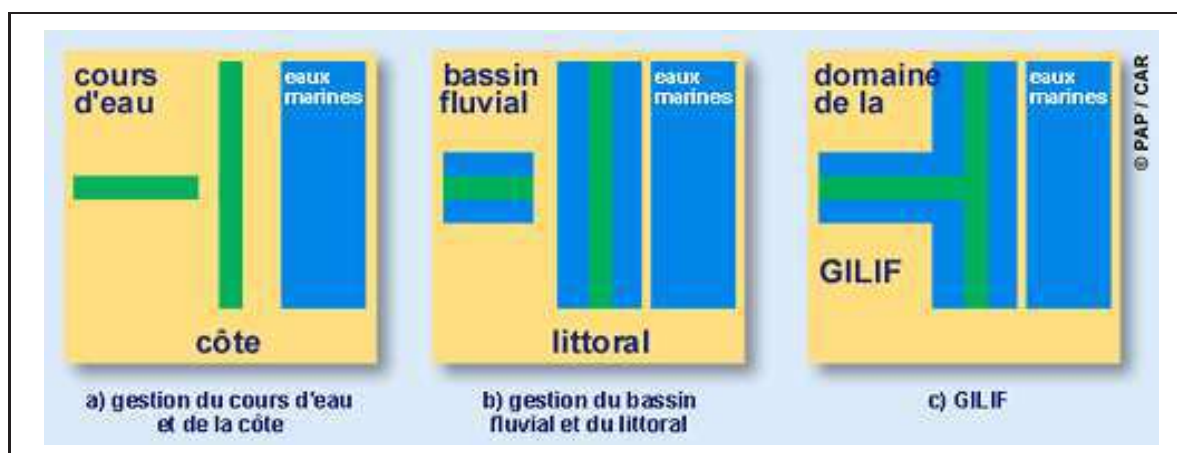


FIG. 6.15 – L'approche écosystémique au service de l'inventaire de l'IG pertinente

Dans un premier temps, il s'agit de définir les *limites spatiales du système* que nous étudions : *ie* définir l'*objet géographique* le plus large qui représente notre domaine d'étude *ie* la *zone côtière* et la nature des *objets géographiques* qui la composent. Cet exercice n'est pas trivial [80, 50].

L'*approche écosystémique* apporte là encore une réponse pertinente (voir Figure 6.15 et Figure

6.16(a)). Elle propose en plus des *eaux marines*, la prise en compte des *eaux continentales* (*cours d'eaux* et *bassins versants* associés) pour l'importance de leurs *interactions* avec les *eaux côtières*, ce qui est implicite dans la notion de *continuum terre-mer* (voir Figure 6.15). Cependant les limites spatiales des *objets géographiques* associés à ces *sous-systèmes* (*concepts spatiaux* correspondant à des *éléments statiques* ou *dynamiques* : *eaux marines*, *eaux continentales*, *bassins versants*, *courants marins*...) varient. Et, récursivement il en est de même pour les *objets géographiques* des *sous-systèmes* qu'ils contiennent eux-mêmes.

En effet, selon la nature des *éléments* et des *interactions* prises en compte, les *limites spatiales* des *objets géographiques* qui leur sont associés peuvent varier d'un *contexte* à l'autre, voire dans un même contexte : "*Le domaine côtier est le siège de processus dynamiques dont les limites sont définies par l'espace d'interaction entre les milieux marins et terrestres qui l'influencent. Ces limites ne sont pas mesurables instantanément et seule l'intégration des processus dans le temps et dans l'espace permet d'en préciser le contour*" [108]. Pour illustrer cette citation on peut prendre l'exemple des limites spatiales des *objets géographiques* associés aux *concepts spatiaux* suivants : un *courant marin* (variable en fonction des *saisons*, des *régimes pluviométriques*...) ou encore le *trait de côte*, une *lagune*...

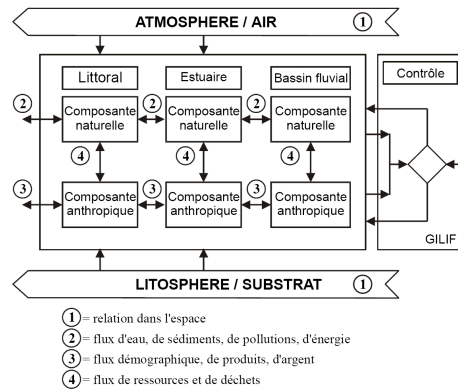
Ceci souligne l'importance de la *dimension temporelle* pour la modélisation des analyses liées à la GIZC. Cela signifie que l'*objet géographique* qu'on peut associer à un même *toponyme* n'est pas unique. Une limite spatiale a donc une pertinence par rapport à une période donnée. Nous sommes conscients de cette nécessité et proposons donc, en perspectives, d'intégrer la *dimension temporelle* (voir partie 8.2.3) sous forme d'attributs de la classe *objet géographique* afin de marquer le caractère éphémère d'une représentation géométrique / spatiale [58, 83]. En outre, il faut également considérer que les limites spatiales associées à un *objet géographique* peuvent également varier selon les différentes *représentations disciplinaires* (voire *individuelles*) de l'*espace*. Par analogie avec la *dimension temporelle* on pourrait parler de *dimension cognitive* pour les *représentations spatiales* possibles d'un même espace (voir diversité des représentations possibles pour des informations semblables en Figure 6.16).

Pour conclure, on peut donc dire que pour chaque *concept spatial* inventorié, le travail est complexe puisqu'on doit lui associer, dans la mesure du possible un *Layer*, dont le contenu, ie les *objets géographiques*, correspond à une représentation partagée par les utilisateurs. A ceci s'ajoute la mise en relation des concepts inventoriés avec d'autres par le biais des relations de différentes natures de notre ontologie spatiale (voir Figure 6.8)

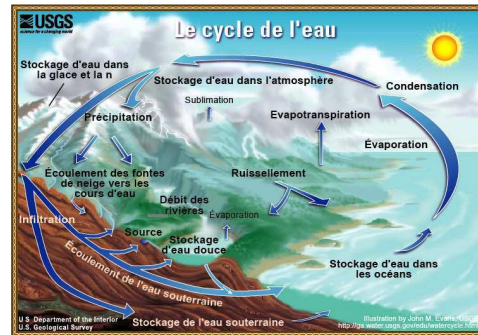
Pour répondre à l'ensemble des points soulevés dans cette méthodologie générale, nous avons donc, d'une part consulté les *sources de référence* du domaine et d'autre part consulté nos partenaires pour prendre en compte les *différentes perceptions disciplinaires de l'espace littoral* dans notre projet.

6.3.2 Les sources de référence utilisées pour l'inventaire des concepts spatiaux pour la GIZC

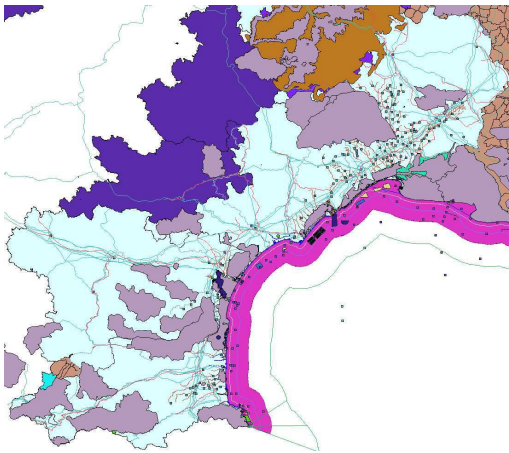
Il n'existe pas actuellement d'inventaire de référence sur les *concepts spatiaux* (*Layers*) et encore moins sur les *objets géographiques* ou les *concepts thématiques* de référence spécifiques à la GIZC [115, 132]. Il existe cependant un grand nombre d'initiatives à l'échelle *européenne* qui partagent l'objectif de définir des *SIG* de référence pour la zone côtière et qui contribuent donc largement à cet objectif [29]. A l'échelle *nationale*, la littérature confirme la nécessité de mettre en place un *référentiel spatial* partagé par les acteurs de la zone côtière [108, 110, 154] et des initiatives comme le *GT littoral* (*Groupe de Travail Littoral*) [8, 9] inventorie finement



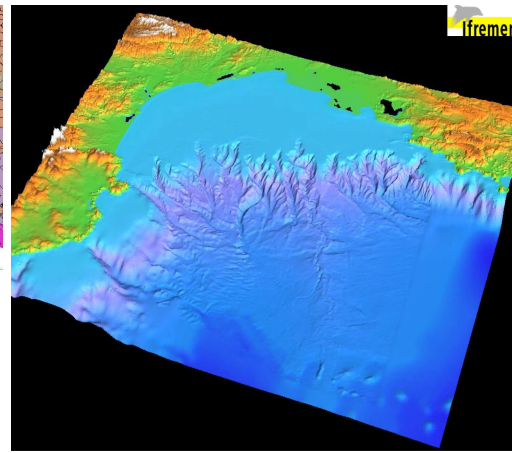
(a) Approche écosystémique et objets géographiques associés [143]



(b) Bloc diagramme 3D non géoréférencé de la zone côtière



(c) Représentation 2D géoréférencée d'Objets géographiques de la zone côtière



(d) Exemple d'IG 3D géoréférencée permettant d'intégrer la dimension verticale

FIG. 6.16 – Exemples de descriptions de *concepts spatiaux* en mode *textuel* (6.16(a)) ou *graphique* (6.16(b)) pouvant être utilisés pour cartographier la zone côtière (6.16(c)) et perspectives 3D (6.16(d))

les principales *thématiques* des *objets géographiques* de référence du domaine de la zone côtière. Ces travaux s'appuient généralement sur la prise en compte des points de vue des utilisateurs [48].

GT littoral Le *GT littoral* propose un inventaire de l'IG existante sur la zone côtière (aussi bien en modes *vecteur* que *raster* et correspond donc à l'approche 1 de la partie 6.3.1 basée sur l'inventaire de l'existant), structuré par *catégories thématiques* :

- Géographie maritime : trait de côte, bathymétrie et isobathe, géologie sous-marine, sédimentologie...
- Faune et flore : aire de distribution, frayère, nourricerie, voie de migration...
- Hydrodynamique et hydrologie : courant, ports de référence, apports ponctuels, apports diffus...
- Activités humaines et usages : amers et balises, objets sous-marins (épaves...), cultures

marines, activités récréatives...

- *Administration et réglementation : limites administratives, limite aval des eaux territoriales, limite aval de la ZEE, limite côtière des communes, eaux fluviales, eaux estuariennes, PNR, ZICO, ZNIEFF, PLU, SMVM...*

Les termes cités ci-dessus ne sont qu'un extrait du large *inventaire* de Layers (et *concepts spatiaux* associés) réalisé par ce *groupe de travail* qui retranscrit pour chacun d'eux, dans la mesure du possible, une *définition de référence*, un *propriétaire* ou *responsable* de la *ressource*...et préconise à terme l'emploi de *métadonnées* (la norme *ISO 19115* en particulier) pour documenter ce *géocatalogue* et d'un *prototype de serveur* permettant d'y accéder [45] (à ce sujet nous renvoyons sur la pertinence de la réponse proposée par la partie 6.2.4). En effet, l'*IG* décrite est *distribuée* dans les *SGDB* des différents acteurs mobilisés.

Il est intéressant de noter pour la *catégorie thématique* de la *faune* et de la *flore* que les *concepts spatiaux* évoqués relèvent plus de *propriétés* générales aux espèces que de spécialisation des *concepts thématiques faune* ou *flore*. Ceci n'est pas étonnant dans la mesure où, par essence, une espèce végétale ou animale ne peut à elle seule constituer un *concept spatial*.

Parmi les points importants, recommandations et objectifs qui se dégagent de l'étude, figurent :

- l'adoption d'un *langage* et d'un *vocabulaire commun* à l'ensemble des acteurs,
- l'*inventaire qualitatif* des données existantes et des besoins,
- la *hiérarchisation* des *besoins* en données de référence,
- la *structuration* et la *cohérence des données* de référence en tenant compte de leur *topologie* et non des seules contraintes de représentation cartographique,
- la réflexion sur la *cohérence entre les données maritimes et terrestres* qui est bien une problématique de GIZC (voir partie 2.1.2).

Cependant cet inventaire n'est pas exhaustif et concerne des données plus pertinentes à l'échelle *nationale* que *régionale*. Les analyses ou implémentations plus régionales complètent ces travaux.

Inventaires régionaux En fonction des caractéristiques régionales des zones côtières, certaines couches d'*IG* ont une importance particulière. On peut citer, pour le cas du *Languedoc Roussillon*, l'exemple des *lagunes* qui sont une originalité de cette zone côtière. On retrouve donc, en plus des caractéristiques géographiques générales de la zone côtière, certaines spécificités voir particularismes géographiques locaux qui complètent (spécialisent selon une *approche top-down*) les inventaires d'ordre plus général réalisés au plan national (approche générique, similaire à une ontologie *top level* [70], voir partie 5.1.4) et contribuent donc à dresser un *inventaire* plus exhaustif de l'*IG* pertinente pour décrire les problématiques de GIZC à l'échelle régionale.

A ce titre, le rapport [48], rattaché aux analyses du *GT littoral*, partage ces objectifs et souhaite donner un aperçu des informations qui devraient être communes à tous les acteurs de la zone côtière en s'appuyant sur l'interrogation de 39 individus représentatifs de la variété des acteurs mobilisés dans la GIZC, dans la région des *Pays de la Loire*. Cinq thèmes ont été mis en avant par les acteurs :

- *Limites physiographiques : trait de côte, zéro IGN...*,
- *Limites de l'hydrodynamique : courants de marée, marnage...*,
- *Occupation du sol : activités industrielles, activités agricoles...*,
- *Limites administratives : Domaine Public Maritime, Zones de pêche réglementée...*,

- *Protection des milieux : espaces littoraux remarquables au titre de la Loi littoral, Zone de Protection Spéciale ZPS, ZNIEFF, espaces acquis par le conservatoire du littoral...*

Il est intéressant de noter que ces *catégories thématiques* rejoignent celles mises en avant par les experts du *GT littoral* et qu'il est donc possible de partager un vocabulaire commun entre les différents acteurs mobilisés dans la GIZC. En outre, l'auteur insiste sur le fait que la majorité de l'*IG* recherchée existe, le problème étant qu'elle est distribuée chez les différents acteurs et non documentée.

Il semble donc, là encore que les verrous pour la réalisation de l'*inventaire des concepts spatiaux* de la zone côtière relèvent plus de la *gestion / mutualisation de l'information* et de la *sémantique* que de l'existence même des ressources.

En complément de ces sources de référence, nous avons analysé et formalisé dans le contexte de notre programme différentes perceptions des *concepts spatiaux* et *objets géographiques* de référence pour la zone côtière régionale du *Languedoc Roussillon* (entretiens menés auprès de nos partenaires, analyse de références bibliographiques...).

6.3.3 Les entretiens auprès des partenaires du projet Syscolag

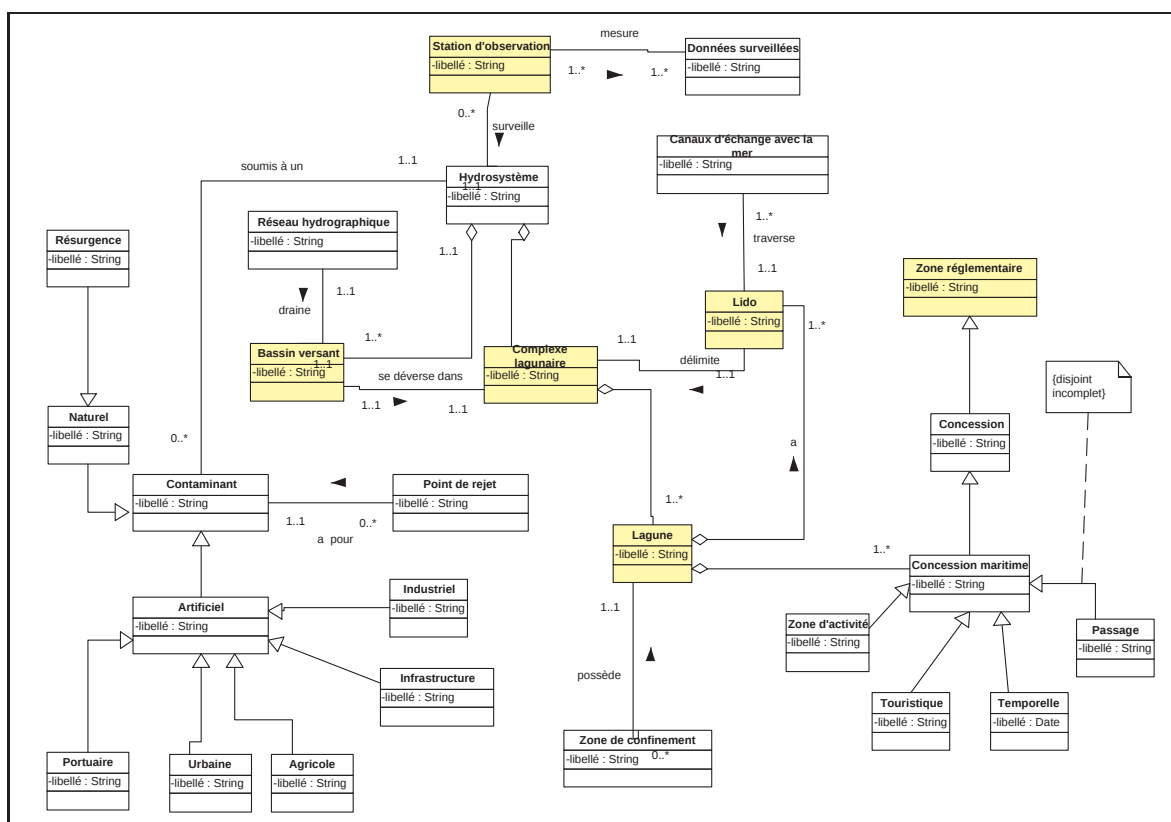


FIG. 6.17 – Exemple de d'analyse et de formalisation UML d'un entretien conduit auprès de l'IFREMER

Une enquête par *entretiens* a été menée pour *inventorier* et *structurer* les *concepts spatiaux* et *objets géographiques* utilisés par les différentes disciplines scientifiques mobilisées dans le

cadre du programme *Syscolag* [91, 23]. Ces *entretiens* ont été suivis d'une *analyse* et d'une *formalisation* utilisant le langage *UML* [91]. La Figure 6.17 illustre la formalisation de l'analyse d'un entretien mené par [91] auprès de l'*IFREMER* sur la problématique concrète des *objets géographiques de référence* pour la gestion de la *qualité des eaux de l'étang de Thau*. La justification de la nature du lien entre *concepts thématiques* et *spatiaux* que nous avons développée en partie 6.1.2 se retrouve dans les diagrammes issus de ces analyses. En effet, si des *instances* apparaissent, elles sont toujours rattachées au *concept thématique* qui les généralise.

Nous avons en outre vérifié au cours de ces analyses que les *objets géographiques* inventoriés (voir partie 6.3.1) peuvent avoir plusieurs *géométries* pour un *toponyme* donné. En effet, dans notre projet, de la même manière que la *description thématique* diffère d'une discipline à l'autre, la perception de l'espace n'est pas homogène [15] : là où les *sciences humaines*, en caricaturant, trouvent un intérêt particulier comme descripteurs d'emprise spatiale aux *objets géographiques* de types *zonages réglementaires* ou *administratifs*, les *naturalistes* s'intéressent d'avantage à des *objets géographiques* du type *limites géomorphologiques* ou d'*écosystèmes*. Ainsi nous avons constaté pour les *géologues* la vision suivante des limites du territoire du *Languedoc Roussillon* : du *Cap Creus* jusqu'à la *faille de la Durance* d'*Ouest* en *Est*, et celle des sciences politiques : celle de la *région en tant que limite administrative*. Il n'est en effet pas envisageable pour la *géologie* d'arrêter les limites de ses *cartes de formations géologiques* aux *limites administratives*. Qu'en serait-il pour un *historien* ? Nous avons donc à gérer *différentes représentations cartographiques*, *I* ou *II* (voir partie 6.1.1) pour une même terminologie (un même *toponyme*, *III*) qui peuvent dans certains cas atteindre un consensus mais pas dans d'autres.

Il apparaît ainsi que des *toponymes* identiques employés par des disciplines différentes ne vont pas couvrir la même emprise géographique (limites floues). En effet, si certains *objets géographiques* ont des limites clairement établies comme un *département* ou une *commune*, d'autres non définis réglementairement, ont des *représentations spatiales* différentes d'une discipline à l'autre et parfois plus finement d'une personne à l'autre au sein d'une même communauté, on touche ici aux *représentations cognitives de l'espace* [15, 70, 71, 122], équivalent spatial de la *polysémie thématique*. Pour ces derniers, un consensus sur les limites paraît plus délicat à obtenir.

Nous avons d'autre part exploité les *concepts spatiaux* et *toponymes* fournis par les utilisateurs de la communauté *Syscolag* à travers la mutualisation de leurs *références bibliographiques*. Cet exercice a mis en avant la diversité des *toponymes* possibles pour décrire un même *objet géographique* (par exemple :) ou des *synonymes* possibles pour un même *concept spatial* (par exemple : *étang*, *lagune*, *bassin*...).

Sur la base de ces analyses nous pouvons désormais implémenter / alimenter le *modèle conceptuel* de notre *référentiel sémantique* et alimenter, par les *concepts spatiaux* inventoriés, les travaux engagés en partie 5.3 sur le *modèle d'ontologie* proposé pour la GIZC. Ainsi les classifications des *concepts spatiaux* que nous avons récoltées reprennent celles des *classifications thématiques*.

6.3.4 Intégration des analyses dans notre modèle d'ontologie

Pour établir une *structuration thématique* des *concepts spatiaux* inventoriés (et implicitement des *objets géographiques*) de référence (selon la démarche prônée par l'*approche écosystémique*), nous nous sommes appuyés sur le *modèle d'ontologie* présenté dans la partie 5.2.3 (voir Figure 6.18). Il s'agit de *spécialiser* ces quatre classes générales de *concepts* par l'introduction

des *concepts spatiaux* que nous allons inventorier.

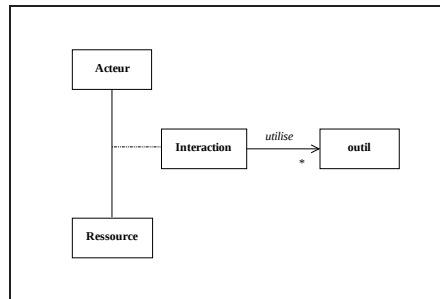


FIG. 6.18 – Le modèle d'ontologie à quatre classes

- Pour la classe *Acteur*, il est possible de prendre en compte les *concepts spatiaux* suivants : *organismes* étant donné qu'on peut géoréférencer les organismes en *Languedoc Roussillon* pour proposer des consultations de ces derniers par *requêtes spatiales*, idem pour les *acteurs* en considérant qu'ils sont localisables, entres autres, par leur *adresse professionnelle* et par leurs *zones d'études*. Toutefois ces *concepts spatiaux* sont plus des espaces associés aux acteurs (des *propriétés* de ces *classes*) qu'à proprement parler des spécialisations de la classe *acteur*.
- Pour la classe *Ressource*, les *concepts spatiaux* sont réparties dans les classes qui la spécialisent.
- Pour la classe *Interaction*, les *concepts spatiaux* sont d'une manière générale des *espaces* associés à la réalisation des *interactions*, selon l'idée que les interactions en particuliers les *activités* (et *usages*) sont consommatrices d'espace. Par exemple l'*agriculture*, la *pêche*, la *courantologie*, les *inondations*... sont autant de *concepts spatiaux* qui peuvent donner leurs noms à des *Layers*. En outre la spatialisation des *usages* est primordiale pour la GIZC, en particulier pour la gestion des *conflits d'usages* [80] et le modèle d'ontologie que nous proposons peut apporter des réponses concrètes à cette problématique [18].
- Pour la classe *Outil*, il n'y a *a priori* pas d'espaces associés qui ne soit pas pris en compte par ceux déjà cités dans les *interactions*.

Une fois que les *concepts spatiaux* sont *étiquetés* au sein des *typologies*, il est alors possible de générer une *arborescence* qui ne contient plus que des *feuilles* correspondant à des *concepts spatiaux*. Ceci est particulièrement intéressant pour générer une *arborescence* de gestion des *Layers*.

Génération dynamique de typologies de concepts spatiaux et de toponymes

Nous proposons dans cette partie de montrer comment nous pouvons bénéficier de l'*inventaire* de *concepts spatiaux* et de *toponymes* intégrés dans les *typologies* que nous avons proposées en partie 5.3 pour générer des *typologies* qui ne contiennent plus que des *concepts spatiaux* et leurs *instances* (*feuilles* de l'*arborescence*).

Problématique et objectifs de la démarche Nous devons permettre à l'utilisateur une navigation aisée dans les *concepts spatiaux*, les *Layers* ou les *objets géographiques* que nous avons inventoriés en utilisant une *structuration* spécifique à ces éléments.

Problématique Autant la désignation de *descripteurs spatiaux* (*I*, *II* et *III*, voir Figure 6.2) en mode *graphique* est simple à gérer grâce aux outils *SIG* qui gèrent leur visualisation, autant celle des *concepts spatiaux* (ou *Layers*) et des *toponymes* (*III*) en mode *textuel* est complexe puisqu'il s'agit alors de les faire désigner à l'utilisateur par une structuration pertinente. Ceci est particulièrement important, par exemple, pour une gestion efficace de l'affichage des *Layers* dans l'*interface cartographique*.

En effet :

- Pour désigner un *concept spatial* (ou *Layers*) ou un *toponyme*, il n'est pas envisageable, compte tenu du volume potentiel de *concepts thématiques* contenu dans notre *ontologie*, de présenter des *typologies* à l'utilisateur qui mélangent *concepts thématiques* et *concepts spatiaux* sans altérer l'*ergonomie* de l'outil par la présence incongrue de *concepts thématiques*.
- Il n'est pas non plus envisageable d'agréger linéairement dans un même *menu déroulant* l'ensemble des *concepts spatiaux* (et *toponymes* associés) contenus dans tous les *Layers* qui ont été importés (ceci est d'autant plus vrai que le nombre de *layers* disponibles est important, ce qui est le cas pour la GIZC).

Objectifs Comme dans les *SIG*, les *interfaces cartographiques* qui proposent la visualisation d'un nombre conséquent de couches d'information géographique ont généralement recours à une *structuration thématique* des *layers* disponibles [145] (voir exemple donné dans l'état de l'art en partie 2.3.2), dont la forme la plus élaborée correspond généralement à une *arborescence* (de type explorateur de fichiers structuré par thématiques).

Nous proposons d'utiliser cette solution, grâce aux liens exprimés en Figure 6.5. La logique de navigation dans les *référentiels thématique* et *spatial* reste ainsi homogène.

Méthode proposée L'introduction de différentes *relations ontologiques* entre les *concepts spatiaux* et les *toponymes*, notamment *hiérarchiques*, peut s'appuyer sur notre *ontologie spatiale* en exploitant le lien entre *concepts thématiques* et *spatiaux* expliqué dans la partie 6.1.2 et illustrée dans le diagramme en Figure 6.5.

Pour cela, à partir du *référentiel sémantique* (Figure 6.19(a)), nous proposons de créer une *vue* qui ne contient plus que les *concepts spatiaux* et les *relations* de *généralisation* / *spécialisation* (voir Figure 6.19(b)). À partir de là, nous pouvons effectuer une *requête* pour remonter depuis les *feuilles* jusqu'à la *racine* (en réincorporant les *concepts thématiques* plus *généraux* nécessaires à la reconstruction d'une *typologie* unique, voir Figure 6.19(c)).

Pour générer une *arborescence* dédiée à la gestion des *Layers*, il suffit d'appliquer cette même *requête* uniquement aux concepts spatiaux qui ont un *Layer* associé, c'est à dire les *concepts spatiaux* gérés dans la table *Layer* de notre *MPD*.

Pour conclure, compte tenu de la nature des ressources inventoriées, nous proposons de distinguer quelques caractéristiques particulières de l'*IG* par rapport aux ressources informationnelles générales de la GIZC.

6.3.5 Caractéristiques de l'IG associée à la GIZC

Nous avons présenté en partie 2.1.3 quelques caractéristiques générales des diverses *ressources informationnelles* mobilisées dans les processus de GIZC. Nous proposons désormais de distinguer quelques caractéristiques propres à l'*IG* mobilisée dans la GIZC.

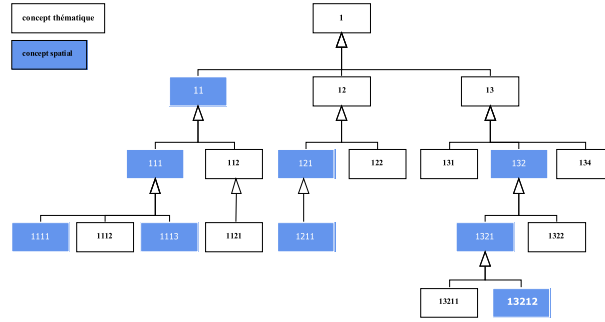
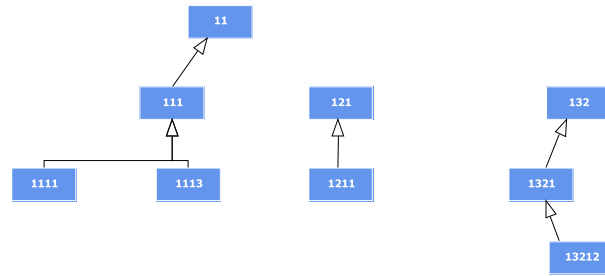
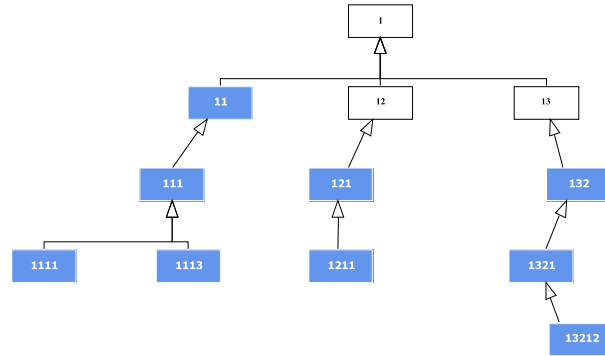
(a) Intégration ou étiquetage des *concepts spatiaux* dans les *typologies* de *concepts thématiques* de l'ontologie(b) *Typologies* de *concepts spatiaux*(c) Structuration thématique des *concepts spatiaux*

FIG. 6.19 – Utilisation des liens entre *concepts thématiques* et *spatiaux* pour générer *dynamiquement* une *structuration thématique* des *concepts spatiaux* et *toponymes* du référentiel *spatial*

A la difficulté de rassembler en un même *SIG* des *données terrestres et marines* produites et gérées différemment au plan national, s'ajoute celle d'intégrer des données issues de différentes disciplines. En effet, bien que manipulées par les géomaticiens, l'*IG* n'est pas pour autant moins distribuée que les autres types de ressources mobilisées dans la GIZC et ne dispose pas, dans notre cas au moins, de *SMD* dédiés pouvant interopérer.

Les problèmes de contraintes d'accès, de diffusion (exemple des positions géographiques des *réécifs artificiels*) ou d'hétérogénéité de référentiels thématiques et spatiaux déjà présents n'en sont que plus complexes (exemple du *zéro hydrographique* et du choix d'un référentiel commun pour le *continuum terre-mer*) [42, 92, 106].

A cela peut s'ajouter des problèmes d'échelle afin d'harmoniser la précision des ressources présentées en fonction des différents utilisateurs aux besoins distincts.

Le choix de la norme *ISO 19115* par la richesse de ses éléments de métadonnées permet cependant de bien gérer cette diversité de modes d'utilisation et correspond à un besoin explicite de service de métadonnées dans le domaine [8, 128]. L'utilisation d'un *référentiel sémantique* peut, pour sa part, répondre à la *complexité inhérente à la mise en cohérence sémantique et géométrique de ces données hétérogènes* [106].

La qualité de l'IG collectée diffère donc d'un acteur à l'autre et la mise en place d'un *référentiel spatial* nécessite un travail commun conséquent d'harmonisation de la terminologie et des méthodes de production utilisées.

Synthèse section 6.3

- Pour alimenter le *référentiel spatial* que nous avons présenté dans la section 6.1, il existe des *sources de référence* dans le domaine sur lesquelles nous pouvons nous appuyer pour *inventorier* et *structurer* les *concepts spatiaux* et *objets géographiques* de référence pour la GIZC.
- En particulier le *GT littoral* propose un *inventaire* de l'IG de référence pour la zone côtière qu'il est intéressant d'intégrer à nos développement en enrichissant la structuration de l'IG proposée. En effet, cet inventaire issu d'un groupe de travail *pluridisciplinaire* respecte globalement la *structuration thématique* que nous avons mise en avant dans notre *ontologie* et les inventaires régionaux confirme sa pertinence.
- Nous proposons de réutiliser le *modèle d'ontologie* que nous avons mis en place dans le chapitre 5 pour accueillir les *concepts spatiaux* inventoriés et les *relations* qu'ils entretiennent.
- Nous proposons finalement une méthode de *structuration thématique de l'IG* qui s'appuie sur le lien entre nos *concepts thématiques* et *spatiaux*.

Conclusion Chapitre 6

La gestion de l'information spatiale prend une importance croissante dans les systèmes d'information et de connaissances au service du *Développement Durable* ou de la GIZC pour lesquels l'espace est un concept central et l'*information géographique*, une aide à la décision, quel qu'en soit le support [119].

Par l'emploi de standards (*ISO 19115* et *OGC*), nous disposons de *descripteurs spatiaux* qui respectent les attentes des utilisateurs puisqu'ils permettent aussi bien la *description spatiale* en *mode textuel* que *cartographique*. En outre, par ce biais, nous assurons l'*interopérabilité* de notre système avec des services équivalents au plan national ou à l'étranger, sans perdre l'adaptation aux caractéristiques de notre domaine d'application attendue par les utilisateurs. L'emploi d'un *SGBDR spatial* associé à une *interface cartographique* permet de gérer plus facilement la *description* et surtout la *recherche des métadonnées par critères spatiaux*. Nous appuyons ainsi la *médiation spatiale* et l'utilisateur bénéficie d'un affranchissement partiel des problèmes de descriptions sémantiques par une alternative ou un complément à l'utilisation d'*ontologie* uniquement pour des descriptions textuelles.

L'*information géographique* pertinente du domaine que nous avons inventoriée à ces fins consti-

tue un *référentiel spatial* lié à la connaissance thématique du domaine.

Le service proposé respecte donc à la fois l'indépendance des organismes pour la gestion de leurs ressources et la nécessité de partager des outils communs dédiés à l'échange nécessaire en vue d'une bonne *intégration de l'information et de la connaissance* pour la résolution des problématiques complexes de GIZC.

Cette *méthodologie générique* peut être déclinée pour différents problèmes de gestion de ressources environnementales, comme en témoigne la démarche commune de ce projet mené en collaboration avec l'équipe *ROSELT*.

Enfin, la possibilité d'être interopérable avec d'autres serveurs de données spatiales (*qui respecte la spécification WMS de l'OGC*) sur la zone côtière respecte le concept de GIZC, la géographie naturelle et le fonctionnement d'écosystèmes dont les limites dépassent nécessairement l'influence des espaces réglementaires définies par la sphère anthropique.

Synthèse du Chapitre 6

- La maîtrise sémantique de la *description spatiale* des ressources informationnelles de la GIZC est essentielle pour le *partage d'information et de connaissance* dans ce domaine. La norme *ISO 19115* fournit des *descripteurs spatiaux* pertinents et standardisés. Nous proposons de rendre ces descripteurs *obligatoires* dans le *profil Syscolag* afin d'améliorer la *qualité des descriptions et des recherches de métadonnées*.
- Nous explicitons les liens entre *concepts thématiques* et *spatiaux* ainsi qu'entre *concepts spatiaux* et *Objets géographiques* et intégrons ces nouvelles relations dans notre *référentiel sémantique*. Le sous-ensemble du *référentiel sémantique* qui en résulte constitue un *référentiel spatial* pour la GIZC.
- Ce *référentiel spatial* est utilisable pour guider la gestion du *service de métadonnées* et par les *acteurs* de la GIZC.
- Nous avons générer un *modèle physique de données* qui permet l'exécution de *requêtes spatiales* pour l'*assistance* à la *saisie* ou à la *consultation* dans un *service de métadonnées* qui utilise des *descripteurs spatiaux*.
- Nous recommandons d'exploiter ce lien à l'aide du *SGBDR Postgres* avec sa *cartouche spatiale Postgis* qui permettent de gérer simplement les descripteurs *spatiaux* selon une forme standardisée par l'*OGC* adaptée aux *descripteurs spatiaux* de l'*ISO 19115* et disposent de nombreuses *fonctions spatiales* utilisables pour interroger le *service de métadonnées*.
- L'utilisation de la norme *ISO 19128 (WMS)* pour la mise en place de *serveur de données spatiales* apporte une réponse éprouvée à nos besoins et permet l'*interopérabilité* du *Web service* proposé avec d'autres à l'échelle nationale et à l'étranger [79].
- Pour alimenter le *référentiel spatial* il existe des *sources de référence* dans le domaine sur lesquelles nous pouvons nous appuyer pour *inventorier* et *structurer* les *concepts spatiaux* et *objets géographiques* de référence pour la GIZC.
- Nous proposons de réutiliser le *modèle d'ontologie* que nous avons mis en place dans le chapitre 5 pour accueillir les *concepts spatiaux* inventoriés et les *relations* qu'ils entretiennent. Ceci permet la mise en œuvre d'une méthode automatique de *structuration thématique de l'IG* qui s'appuie sur le *référentiel sémantique*.

Chapitre 7

Servir l'information et la connaissance

Nous avons, dans les chapitres précédents, décrit les différentes parties de l'infrastructure informatique que nous proposons. Un *service de métadonnées* composé par un *jeu d'éléments normalisés* par la norme *ISO 19115* et triés dans un *profil* spécifique au projet *Syscolag*. Dans la partie 4.3.1, nous avons présenté comment la nature des éléments retenus est variable de même que leur intérêt pour la *GIZC*. En particulier, les deux éléments fondamentaux pour ce domaine sont ceux qui décrivent les *portées thématique* et *spatiale* des ressources informationnelles. Ainsi avons-nous expliqué, dans les chapitres 5 et 6, l'intérêt d'un *usage contrôlé* des *concepts thématiques* et *spatiaux* par la mise en œuvre d'un *référentiel sémantique* qui intègre et structure les *éléments* de descriptions (*concepts* et *relations*) propres aux *dimensions thématiques* et *spatiales*. En particulier, voir chapitre 6, la *dimension spatiale* est essentielle dans les processus de *GIZC* et le programme *Syscolag* souhaite l'utiliser pour appuyer sa démarche de partage d'*information* et de *connaissance* par une utilisation conviviale du *support cartographique*.

En dehors de l'usage de ces descripteurs auxquels nous avons consacré l'essentiel de notre travail, on trouve, dans le *profil Syscolag*, d'autres *éléments de métadonnées* d'importance moindre pour la qualité de la description en *GIZC* mais qu'il est souhaitable d'*automatiser* lors des processus de *saisie* ou de *consultation* de *fiches de métadonnées* (faute de quoi la lourdeur du *SMD* peut rebuter les utilisateurs).

Pour les types particuliers d'*éléments de métadonnées* que nous avons identifiés dans le contexte de la *GIZC*, nous proposons donc une *assistance à la saisie* à l'aide d'applications (*composants*) qui permettent la *sélection de valeurs préenregistrées* dans le *référentiel sémantique* (pour les *descripteurs thématiques* et *spatiaux*), dans l'*annuaire électronique* (pour la saisie des *informations relatives aux acteurs mobilisés dans la GIZC* et impliqués par les ressources informationnelles gérées dans notre *SMD*) et dans une base de *références bibliographiques* (pour la saisie des *éléments de métadonnées principaux* déjà gérés dans des *références bibliographiques*). Concernant les *descripteurs spatiaux*, nous avons également précisé en partie 6.2 que compte tenu de la difficulté de la description *géométrique géoréférencée* des *objets géographiques*, l'assistance d'une *interface cartographique* conviviale est indispensable pour imposer la saisie des *éléments de métadonnées* correspondants.

Les *éléments de métadonnées* restants sont gérés de manière entièrement *générique* sans contrôle de la valuation autre que celui des *CodeLists* proposés par l'*ISO 19115*.

D'un point de vue technique, pour *gérer l'information* associée à ces différents éléments dans notre *base de métadonnées*, la solution retenue s'appuie sur des *interfaces graphiques* consti-

tuant un *portail Web* dont les pages *HTML* sont interfacées par un *langage de script* avec notre *SGBDR*. Pour la *simplicité* de son *administration* et son *évolutivité*, la *génération* du contenu de ces pages est *dynamique* et indépendante du type des ressources décrites. Ainsi, conformément aux recommandations fondamentales du *DD*, est-il possible de *diffuser* le contenu de notre service de *métadonnées* en *libre consultation* auprès du *grand public* et, pour les *utilisateurs identifiés*, de mettre à disposition les *formulaires de saisie* qui permettent, concrètement la *mutualisation* (dans notre *SGBDR* centralisé) de l'*information* sur les *ressources dispersées* chez les différents *acteurs* mobilisés dans la GIZC régionale. Ceci est donc un premier pas vers l'*interopérabilité* des bases de données environnementales de nos partenaires.

Cependant, l'intérêt des *composants* développés en complément du *SMD* réside peut être autant dans leur utilisation en tant qu'application à part entière que dans celle qui en restreint l'usage à la *saisie* ou à la *consultation* de *fiches de métadonnées* et qui n'exploite qu'une partie de leur intérêt potentiel pour l'utilisateur. En effet, comme nous l'avons montré dans la partie 6.3, il serait dommage, après avoir collecté l'*IG de référence pour la GIZC* de n'en proposer la consultation que par le biais de l'*interface cartographique* qui assiste la *saisie* des *descripteurs cartographiques* dans le *SMD* alors que les *atlas cartographiques* sur le Web présentent un intérêt reconnu pour le partage d'*IG* entre acteurs d'un domaine. Il en va de même pour le reste du contenu du *référentiel sémantique*, de l'*annuaire électronique* ou de la *base de référence bibliographique* qui peuvent faire l'objet d'applications séparées du *SMD*.

Dans ce chapitre, nous proposons, dans la première section, de décrire brièvement l'environnement de développement de nos applications. Nous présentons nos choix, orientés par le respect de standards ouverts, que ce soit pour le *SGBDR Postgres/Postgis* (comme cela a été déjà décrit), le *langage de script PHP* utilisé pour la *génération dynamique* des *formulaires HTML* de notre *SMD* et les outils de visualisation qui exploitent le contenu de notre *référentiel sémantique* (bibliothèque *Java* pour la partie *sémantique* générale et langage *SVG* pour la partie spécifique à la *cartographie*). Dans les sections 7.2 et 7.3, nous présentons les *interfaces hommes machines* (*IHM*) (essentiellement des *interfaces graphiques*) qui permettent la réalisation des différents *cas d'utilisation* que nous avons énumérés dans le chapitre 3 (voir partie 3.3). Nous rendons compte du travail qui a consisté à exploiter le contenu de notre *SGBDR* en détaillant les différentes *interfaces graphiques* à la disposition des *utilisateurs* sur notre *portail Web*. La section 7.2 décrit les *interfaces graphiques utilisateurs* (*GUI* pour *Graphical User Interface*) qui permettent la gestion du *SMD* (*saisie* ou *consultation*) par l'intégration, selon la nature des *éléments de métadonnées* concernés (voir partie 4.3.3 Figure 4.12), de *composants* permettant l'*assistance* ou le *contrôle* de leur *valuation* en s'appuyant notamment sur le contenu du *référentiel sémantique* par la mise en œuvre d'une *ontologie (spatiale)* (en modes *textuel* et *graphique*, en particulier un *vocabulaire contrôlé* et une *interface cartographique*) mais aussi d'un *annuaire électronique* et d'une *base de référence bibliographique*. Dans la section 7.3, nous présentons l'intérêt d'une utilisation indépendante de ces *composants* en les enrichissant de nouvelles fonctionnalités pour optimiser la *gestion (restitution ou administration)* de la *connaissance* que nous avons inventoriée sur la GIZC et stockée dans le *référentiel sémantique*. En particulier nous présentons une *base terminologique*, un outil de *visualisation du réseau sémantique* et un *atlas cartographique* spécifiquement dédié à l'exploitation du contenu du *référentiel spatial*.

7.1 Choix techniques pour servir l'information et la connaissance contenues dans notre SGBDR

Nous présentons dans cette section les choix techniques qui permettent d'exploiter le contenu de notre *SGBDR* (*métadonnées* et *connaissances* dont la nature a été détaillée dans les chapitres 4, 5 et 6) et de réaliser les différentes *interfaces graphiques* nécessaires pour les divers *cas d'utilisation* présentés dans la partie 3.3.

A ces fins et compte tenu des objectifs énoncés dans le chapitre 3, les *langages* et *applications* utilisés varient selon la nature des *cas d'utilisation* et des *éléments de métadonnées* concernés mais le souci de rester dans un *environnement de développement libre* est respecté.

7.1.1 Choix techniques pour le *SGBDR* et le *langage de script*

Nous justifions brièvement ici les choix de notre *SGBDR* et du *langage de script* qui nous permet d'en exploiter le contenu afin de générer les pages *HTML* du *portail Web* qui héberge le *SMD* et les différents *composants* développés.

Le SGBDR Postgres / Postgis

Nous n'allons pas détailler à nouveau les raisons qui justifient le choix du *SGBDR Postgres*, nous renvoyons pour cela à la partie 6.2.3 et rappelons simplement que ce logiciel est libre, performant et dispose d'une *cartouche spatiale* (*Postgis*) essentielle pour intégrer le contenu de notre *référentiel spatial* puisqu'elle permet la gestion dans le même *SGBDR* de l'*information spatiale* en complément de l'*information alphanumérique* traditionnelle gérée par *Postgres*. En outre, par cette approche, nous bénéficions d'outils libres et variés pour l'administration de ce *SGBDR* (par exemple : *PhpPgAdmin* comme client Web ou les logiciels "*stand alone*" *PgAdmin III*, *PgAccess*...).

Grâce à *Postgres/Postgis*, il est donc possible de réunir en un *SGBDR* unique, l'ensemble des fonctionnalités que nous voulons proposer aux utilisateurs sur notre *portail Web*. Pour générer les pages *HTML* correspondantes aux différentes tables de la base de données, nous utilisons le langage libre *PHP*.

Le Langage PHP

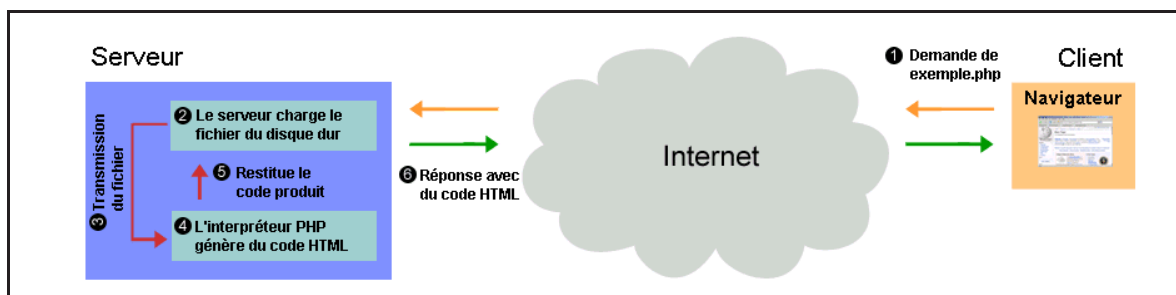


FIG. 7.1 – Le choix de *PHP* comme langage de script pour notre portail Web [179]

PHP est un *langage de script interprété et libre* qui est utilisé ici pour la génération de pages Web (contenant du code $(X)HTML^1$, *CSS*², *JavaScript*³, ...) au *contenu dynamique* (voir illustration en Figure 7.1).

PHP a été utilisé dans sa version 4 qui présente un intérêt pour sa simplicité d'apprentissage (important dans le contexte de notre implémentation par rapport au public visé, voir partie 3.1) et a permis la collaboration avec l'équipe *ROSELT* à l'origine du logiciel *MDWeb* sur lequel nous nous sommes appuyés pour mettre en place un *prototype de démonstration*. *MDWeb* permet la *gestion* d'un *service de métadonnées* à l'aide des choix techniques que nous venons d'énumérer (voir détails complémentaires en partie 4.2.3). La Figure 7.2 donne un aperçu des *formulaires HTML* générés *dynamiquement* par *PHP* (en fonction du *jeu d'éléments de métadonnées* retenus dans un *gabarit* sélectionné par un *producteur de métadonnées*, voir partie 4.3.2). L'*interface graphique* présentée peut aussi bien servir à la *saisie* qu'à la *consultation* de *fiches de métadonnées* dans le *SMD* et l'*administration* des tables de la *base de métadonnées* sous jacente peut, comme nous l'avons précisé, s'appuyer sur des logiciels libres. Le détail des développements réalisés est donné dans [43, 44, 46].

Au sein de ce cadre général de développement, certains *cas d'utilisation*, liés à la valuation d'*éléments de métadonnées* particuliers, nécessitent le développement de *composants* spécifiques, utilisant des langages voire des applications dédiées, pour la mise en œuvre d'*GUI* de *contrôle* ou d'*assistance* à la *saisie* (en particulier ceux qui concernent la *visualisation* du contenu de notre *référentiel sémantique* que ce soit pour les *concepts thématiques* ou *spatiaux*).

7.1.2 Choix techniques pour la gestion des référentiels sémantique et spatial

Compte tenu des efforts réalisés pour *analyser* et *formaliser* l'*information* et la *connaissance* sur la GIZC, il est intéressant de prendre en compte l'émergence d'applications existantes pour les *représenter*, notamment à l'aide d'outils de *visualisation* conviviaux. Basiquement, il s'agit de restituer, par une *représentation graphique*, l'*information* accumulée sur les *concepts* que nous avons *inventoriés* et *structurés*.

Plus précisément, l'*information* contenue dans le *référentiel sémantique* mélange différentes sortes de *concepts* (en particulier *concepts thématiques* et *spatiaux*) qui peuvent, selon les cas, être utilisés pour *contrôler la valuation des éléments de métadonnées* correspondants⁴ ou simplement *consultés* en tant que *références terminologiques* afin de partager leurs propriétés *sémantiques* et participer ainsi à l'élaboration d'un *langage commun*. Selon le *cas d'utilisation* concerné et la nature du *concept*, les *composants* développés sont différents et leurs *GUI* utilisent des *représentations* en modes *textuel* et *graphique* (*géoréférencé* dans le cas des *concepts spatiaux*).

¹Extensible HyperText Markup Language. Reformulation de la norme HTML 4.0 en XML, incluant trois DTD correspondant à celle de HTML 4.0 (strict, transitional et frameset, pour la version 1.0). L'intérêt est de permettre un passage en douceur du HTML classique à XML [168].

²Cascading Style Sheet. Modèle de feuille de style utilisé dans les navigateurs web (à partir des versions 4 pour MSIE et Navigator). Elles permettent (entre autres) de modifier l'aspect de tout un site en ne touchant qu'à un seul fichier (la feuille de style), et donc d'alléger grandement la gestion d'un site [168].

³Langage de script assez simple, développé par Netscape et initialement baptisé Livescript (dont le nom a été modifié pour bénéficier de l'élan positif déclenché par Java. . . Sans que les deux langages aient un quelconque rapport) [168].

⁴*Éléments de métadonnées* représentant *mot clé* et *toponyme*.

FIG. 7.2 – Formulaire (X)HTML généré dynamiquement par PHP pour un *gabarit* défini dans la base de métadonnées

Choix techniques pour les *interfaces graphiques* du référentiel sémantique

Les *GUI* concernées par la visualisation du *référentiel sémantique* doivent permettre l'exécution des *cas d'utilisation* relatifs à :

1. l'*assistance* à la *valuation* des *descripteurs thématiques* ou *spatiaux* lors de la *saisie* et de la *consultation* du *SMD*,
2. la *gestion* (*consultation* et *administration*) du *référentiel sémantique*.

Selon le *cas d'utilisation* concerné, les techniques employées pour présenter le contenu du *référentiel sémantique* dans les *GUI* diffèrent :

1. Pour la *saisie* des *descripteurs thématiques* ou *spatiaux*, nous proposons de nous appuyer sur une technique simple basée sur la présentation d'une *arborescence* de *concepts* sélectionnables par l'utilisateur et générée en *HTML* avec *Javascript* (voir détail dans [46]). Ce *composant* est utilisable indifféremment pour la *valuation* de ces *descripteurs* lors de la *saisie* ou de la *consultation* du *SMD*, voire pour la *valuation* ou l'*administration* du *référentiel sémantique*.
2. Pour la *navigation* de l'utilisateur dans le contenu général du *référentiel sémantique*, les techniques issues de la *cartographie sémantique*, notamment les *topic maps*, offrent des solutions de visualisation intéressantes sous la forme de *réseaux sémantiques*. Les outils disponibles sont nombreux et on peut distinguer grossièrement, pour notre cas, ceux qui

FIG. 7.3 – Exemple de *GUI* pressentie pour renseigner les mots clés

exploitent le contenu des *SGBD(R)* de ceux qui s'appuient sur des fichiers basés sur des syntaxes *XML* (comme *RDF*, *XTM*⁵ ou encore *OWL* dont nous avons donné un exemple en Figure 5.17(c) partie 5.2.1). En particulier, pour cette dernière catégorie, une application comme *omnigator* d'*ontopia*⁶ permet la visualisation d'information formalisée, par exemple, en *RDF* (voir Figure 7.4(a)).

Les objectifs visés à long terme influent sur le choix final du formalisme de représentation des *connaissances*. Pour le moment, notre prototype de consultation du *référentiel sémantique* s'appuie sur la *librairie Java Prefuse*⁷ pour la visualisation du *référentiel sémantique* (stocké dans *Postgres*), sous la forme d'un *réseau sémantique*. Nous illustrons plus en détail l'application développée par [107] en partie 7.3.1.

Les *concepts spatiaux* sont une sorte de *concept thématique* qui hérite donc de ces méthodes de visualisation mais il est également possible de leur associer une *représentation géométrique géoréférencée* qui permet leur visualisation sous forme *cartographique*.

Choix techniques pour la représentation cartographique de l'*IG* contenue dans le *référentiel spatial*

La visualisation cartographique de l'*IG* contenue dans le *référentiel spatial* va permettre la réalisation de deux *cas d'utilisation*. D'une part, le *support cartographique* va être intégré dans les *GUI* du *SMD* pour aider l'utilisateur à *décrire* et/ou à *rechercher*, de manière conviviale, des ressources informationnelles (voir partie 7.2.1), d'autre part, la *connaissance* contenue dans le *référentiel spatial* va être mieux restituée par la mise en place d'une application de visualisation dédiée que nous appelons *atlas cartographique* (voir partie 7.3.2). Ces deux applications utilisent un même *composant*, spécifique à notre programme, qui permet la visualisation de l'*IG* du *référentiel spatial* à l'aide du langage libre *SVG*.

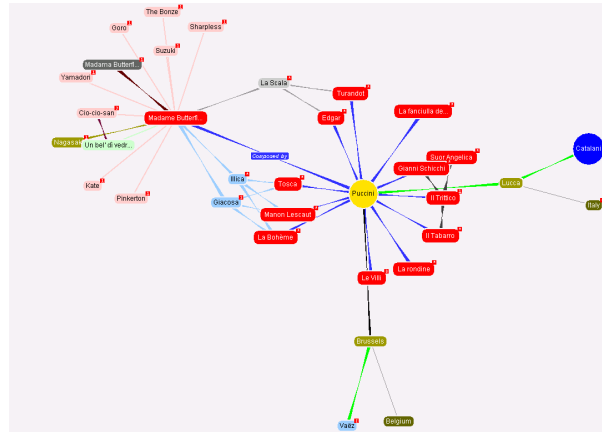
Le choix du format graphique en *mode vecteur SVG* (*Scalable Vector Graphics*)⁸ nous a

⁵XTM : XML-based Topic Maps ou norme *ISO 13250* [1].

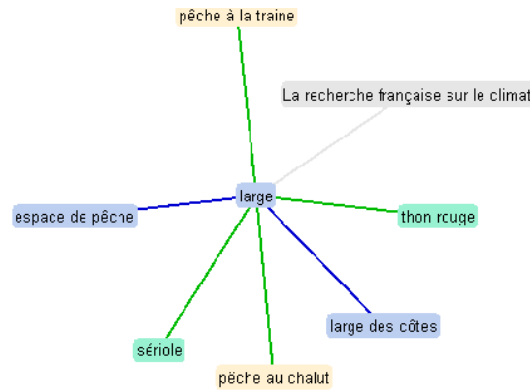
⁶<http://www.ontopia.net/omnigator/models/index.jsp>.

⁷Prefuse, <http://prefuse.sourceforge.net/>

⁸SVG est utilisé en version 1.1, est un langage libre équivalent au *Flash* (langage propriétaire quant à lui). C'est une grammaire *XML* définie par le *W3C* pour permettre la description d'images 2D vectorielles en



(a) Illustration de visualisation d'un réseau sémantique avec l'application *omnigator* [140]



(b) Exemple de visualisation de notre référentiel sémantique sous forme d'un réseau sémantique à l'aide de l'application *Java* utilisant la librairie *Prefuse*

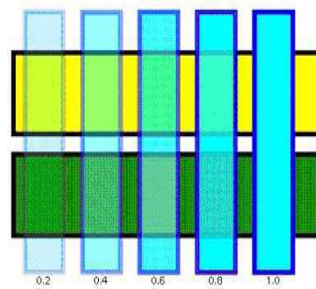
FIG. 7.4 – Exemples d'outils de visualisation de réseau sémantique

paru intéressant pour ses qualités de *rendu visuel* (intérêt pour l'utilisateur), la possibilité de *géoréférencer* les représentations graphiques et sa *légèreté* qui facilite l'*intégration* de l'application dans les *formulaires* du *SMD*. De plus, d'un point de vue graphique, l'implémentation de l'*algorithme des peintres* améliore la lisibilité des cartes et résout certains problèmes liés à la notion de *couche active* par la gestion des transparences (voir Figure 7.5(a)).

Nous utilisons la *cartouche spatiale Postgis* de *Postgres* pour importer les *objets géographiques* de notre *référentiel spatial* dans des *tables spatiales* (qui portent le nom de *concepts spatiaux*, voir partie 6.2.3). Ceci permet de les manipuler avec des *fonctions spatiales* (*relations topologiques* standardisées par l'*OGC*, voir partie 6.1.2) ainsi que de générer une sortie *SVG* à l'aide de la fonction *Assvg* de *Postgis* (intéressante pour sa vitesse d'exécution élevée).

En outre, ce *composant* respecte les spécifications de l'*OGC* pour les serveurs Web de données spatiales (*WMS* voir partie 6.2.4) de manière transparente pour l'utilisateur. Le langage *EC-*

utilisant *XML*, les *CSS*, le *PNG* et le *JPEG*.



(a) Effet de superposition et transparence avec *SVG*



(b) Représentation cartographique géoréférencée de notre *référentiel spatial* avec *SVG*

FIG. 7.5 – Utilisation du langage SVG pour des représentations graphiques en mode *vecteur*

*MAScript*⁹ a servi à l'interface avec *Potgres/Postgis*. Le détail technique de ces fonctionnalités est accessible dans [39, 32].

Il était également envisageable d'utiliser d'autres *interfaces cartographiques* qui génèrent également du *SVG* et possèdent des fonctions de consultation de l'*IG* plus riches (*Mapserver*¹⁰, *Geoserver* ou encore *Geotools* qui implémentent les normes *WMS* et *WFS* de l'*OGC*) mais la lourdeur de ces applications qui tendent à proposer de véritables *SIG* en ligne est trop importante par rapport aux cas d'utilisation relatifs à la saisie d'*éléments de métadonnées* (nombreuses fonctionnalités inutiles dans ce cas). Nous revenons sur ce point en parties 7.3.2 et 8.2.2.

⁹ *JavaScript* est aujourd'hui défini par la norme ECMA-262, aussi connue sous l'appellation ECMAScript.

¹⁰ *Mapserver* possède une sortie graphique *SVG* à laquelle on reproche actuellement une génération trop lente en sortie.

Synthèse Section 7.1

- L'utilisation du *langage de script PHP* associé au *SGBDR Postgres* et à sa cartouche spatiale *Postgis* pour la *génération dynamique* d'*GUI* en *(X)HTML* assure un environnement de *développement libre* pour la mise en place de notre *portail Web*.
- Pour ce qui concerne la visualisation du *référentiel sémantique*, les *composants* qui ont été développés s'appuient sur des langages et applications spécifiques.
- Un autre *composant*, plus léger en vue d'une *intégration* dans le *SMD*, permet l'*assistance à la valuation* (par des *concepts*) des *éléments de métadonnées* *mot clé* et *toponyme*.
- Un *composant* permet la *visualisation* des *concepts* contenus ainsi que des *relations* qu'ils entretiennent sous la forme d'un *réseau sémantique*.
- Pour ce qui concerne la *visualisation* du *référentiel spatial*, le langage *SVG* a été utilisé pour ses qualités de représentations (carto)graphiques en mode *vecteur* et sa complémentarité idéale avec la fonction *Assvg* de *Postgis*.
- Un *composant* spécifique, utilisant *SVG* et la spécification *WMS* de l'*OGC* (voir partie 6.2.4) permet la *visualisation* des *concepts spatiaux* et *objets géographiques* contenus dans *Postgis* associés au sein d'une *interface cartographique* utilisable dans ou en dehors du *service de métadonnées*.

Après avoir décrit les caractéristiques techniques générales de développement des *GUI* du *portail Web* concernant le *SMD* et les *composants* spécifiques exploitant le contenu de notre *référentiel sémantique* (pour la *gestion des valuations* des *éléments de métadonnées* essentiels pour la *GIZC* ou la *consultation* des *connaissances* contenues dans ce référentiel), nous présentons, dans les sections suivantes, des exemples d'*interfaces graphiques GUI* qui les mettent concrètement en œuvre pour résoudre nos cas d'utilisation : d'abord intégrés aux *formulaires HTML* pour la gestion du *SMD* (section 7.2) puis, utilisés séparément, pour porter à *connaissance* des utilisateurs le contenu de notre *référentiel sémantique* (section 7.3).

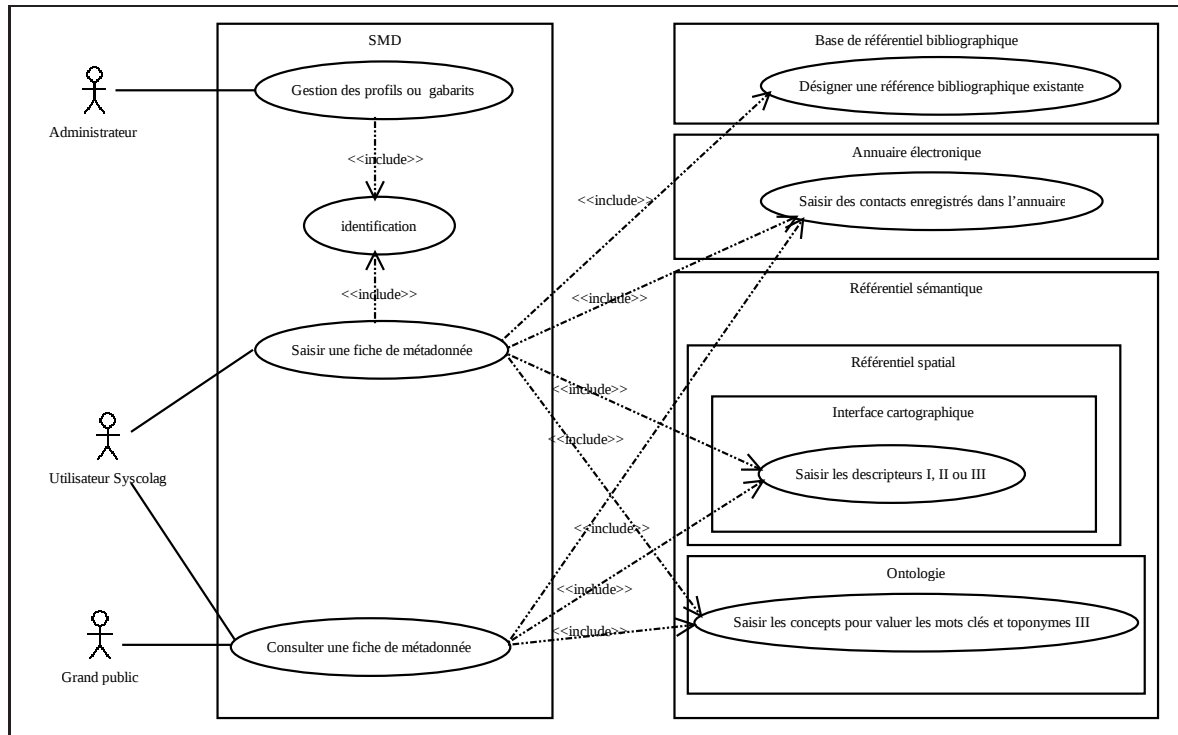
7.2 Interfaces graphiques de gestion du service de métadonnées

Nous présentons dans cette section comment nous avons intégré dans les *formulaires de saisie* et de *consultation* du *SMD* les *GUI* associées aux *composants* présentés dans la section 7.1.

Ainsi, il est possible de *contrôler* ou d'*assister* la *valuation* des *éléments de métadonnées* essentiels pour la *GIZC* contenus dans les *formulaires* de *saisie* et de *consultation* du *SMD*. En outre, nous proposons également des *GUI* complémentaires qui facilitent la tâche complexe d'*administration* relative à la saisie des éléments (retenus par nos partenaires dans le *profil Syscolag* et les *gabarits* qui en découlent, voir détail en partie 4.3.1).

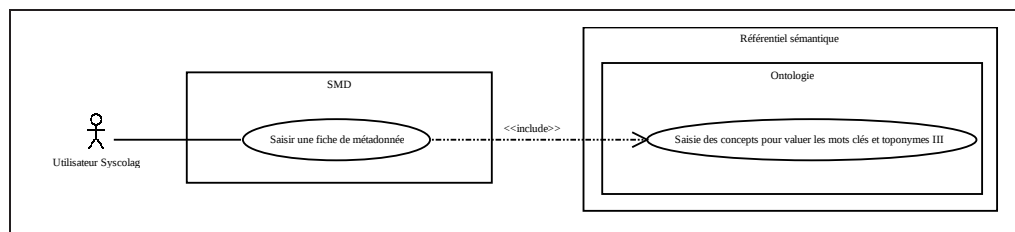
7.2.1 Interfaces graphiques de saisie des fiches de métadonnées

Nous allons à présent distinguer les *GUI* associées aux différents *cas d'utilisation* qui permettent aux *producteurs*, *grand public* et *administrateur* de réaliser la *gestion* du *SMD*. En particulier, nous détaillons comment sont renseignés (lors de la *saisie* ou de la *consultation*)

FIG. 7.6 – Cas d'utilisation associés au *SMD*

les *éléments de métadonnées* dont la *valuation* est *assistée* ou *contrôlée* par des *composants* représentés par des *packages UML* dans les *cas d'utilisation* représentés en Figure 7.6.

Saisie en mode texte des mots clés et toponymes contenus dans le référentiel sémantique

FIG. 7.7 – Cas d'utilisation associés à la saisie en mode texte dans le *SMD* des mots clés et toponymes contenus dans le référentiel sémantique

Ce *cas d'utilisation* (voir Figure 7.7) concerne aussi bien la *valuation* en mode *textuel* de l'élément de métadonnée *mot clé* par des *concepts thématiques* et *spatiaux* que celle de l'élément de description de l'étendue géographique par des *toponymes* (*III*, voir partie 6.1.1). Afin de permettre un *contrôle du vocabulaire* peu contraignant pour l'utilisateur et léger pour le système, nous présentons un *inventaire des concepts* et *toponymes* associés dépourvu de détail sur leurs *propriétés* (que nous gérons dans le *référentiel sémantique*, voir détail en partie

5.1) hormis l'emploi des *structurations thématiques* pour faciliter la navigation au sein de l'*inventaire*. Pour cela nous proposons de nous appuyer sur les relations de type *BT*, présentées en partie 5.1.3, pour hiérarchiser les *concepts thématiques* et *spatiaux* (voir Figure 7.8 et 7.12(a)) et sur les relations d'*instanciation* entre *concepts spatiaux* et *toponymes*. En effet, dans un formulaire de métadonnées, afin de *valuer* les *éléments de métadonnées mot clé* et *toponyme*, il n'est pas nécessaire de restituer l'ensemble de la *connaissance* associée à chaque *concept* puisque l'objectif consiste uniquement à *désigner le concept* ou le *toponyme* (et non à *prendre connaissance du référentiel sémantique*).



FIG. 7.8 – Exemple de GUI pressenti pour renseigner les mots clés avec des *concepts thématiques* ou *spatiaux* et des *toponymes*

Nous avons vu en partie 6.1.2 que les *concepts spatiaux* sont une *sorte de concept thématique*. Ils peuvent, en outre, être liés par des *relations spatiales* et être représentés *cartographiquement*. L'*interface cartographique* que nous avons développée exploite ces nouvelles propriétés des éléments (*concepts spatiaux* et *objets géographiques*) gérés dans le *référentiel spatial*.

Saisie des descripteurs géographiques avec l'interface cartographique

L'objectif de cette application est d'*assister* et, ainsi, d'*optimiser* la qualité de *valuation* des trois sortes d'*éléments de métadonnées* dédiés à la *description de l'étendue géographique concernée par la ressource* (voir partie 6.1.1 Figure 6.2).

Comme cela est illustré en Figure 7.9 nous proposons l'utilisation d'une *interface cartographique* qui *facilite* et *contrôle* la saisie de ces *valuations* et s'appuie sur le *composant* présenté en partie 7.1.2.

L'*interface cartographique* est l'alternative à la *description textuelle* présentée en partie 6.1.1. Elle exploite les *propriétés géométriques* des *Objets géographiques* de notre *référentiel sémantique* et permet de renseigner les *éléments de métadonnées* relatifs à la *description de l'étendue géographique de la ressource* en mode *cartographique* de manière *géoréférencée*. Ceci est possible pour tous les *concepts spatiaux* qui possèdent un *Layer* associé (voir partie 6.1.2) permettant la saisie des éléments de métadonnées *I* (*géométrie précise*), *II* (*rectangle englobant*) et *III* (*toponyme*) en mode *cartographique* (voir partie 6.1.1 et Figure 6.2).

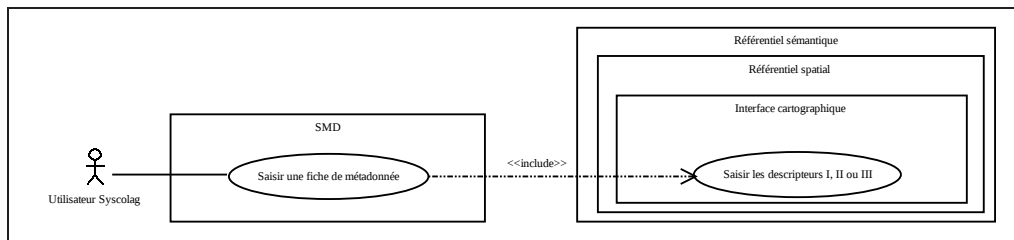


FIG. 7.9 – Cas d'utilisation associés à la saisie des descripteurs géographiques avec l'interface cartographique dans le *SMD*

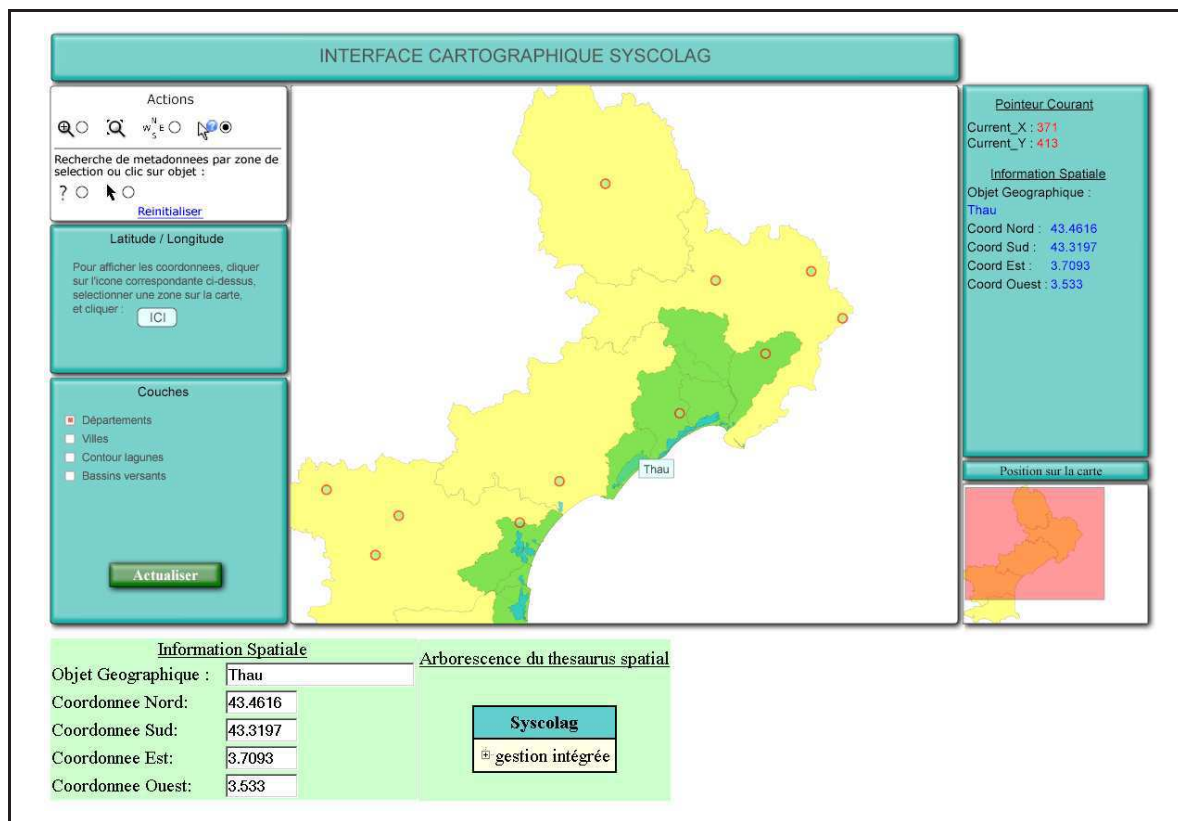


FIG. 7.10 – Description géographique d'une ressource dans l'*ISO 19115*

Sans ce type d'interface, il n'est pas envisageable de demander la *valuation* des *éléments de métadonnées* correspondants sous la forme demandée par l'*ISO 19115* et il est donc impossible, par la suite, d'effectuer des *requêtes spatiales* sur les métadonnées (voir partie 6.2.2). Son importance est donc capitale.

Dans ce cas, afin de s'intégrer dans les formulaires de *saisie* ou de *consultation* du *SMD*, nous utilisons également une version légère du *composant* pour la rapidité et le confort de description de ces éléments.

La Figure 7.12(b) présente l'*interface cartographique* dont dispose l'utilisateur pour décrire la *portée géographique* d'une ressource lors de la *saisie* d'une *fiche de métadonnée*.

Deux cas de *requêtes spatiales de saisie* se distinguent selon la nature des descripteurs employés par l'utilisateur (voir détail en parties 6.1.1 et 6.2.3) :

- l'utilisateur décrit l'*étendue géographique* par des descripteurs connus du système (*descripteurs I ou III*, voir Figure 7.10), dont on déduit le *descripteur II*,
- l'utilisateur ne trouve pas de *descripteurs I ou III* appropriés parmi ceux proposés et préfère alors décrire l'*étendue géographique* par le *descripteur II*¹¹. A l'inverse du premier cas, nous allons générer le *descripteur I* correspondant à ce *rectangle englobant* et demander à l'utilisateur si ce nouvel *objet géographique* représente un *toponyme* (*descripteur III*) absent de notre *référentiel spatial* (et donc une *thématique* absente de notre *référentiel spatial* d'après les liens entre *toponyme*, *concept spatial* et *concept thématique*, voir Figure 6.4), sinon on peut donner à *III* la valeur par défaut *rectangle englobant*.

Au bilan, dans les deux cas, nous obtenons simultanément les trois types de descripteurs de l'*ISO 19115* sans que l'obligation de *saisie* soit trop contraignante.

Si les *descripteurs thématiques* et *spatiaux* sont essentiels pour l'*indexation* des ressources informationnelles mobilisées dans la GIZC, nous avons également développé des *GUI de saisie* pour répondre à la nécessité d'*automatiser* les *éléments de métadonnées* de moindre importance mais qui, sinon, alourdissent la *saisie* (voir détail en partie 4.3.1).

Saisie de *Contacts* par le biais de l'*annuaire électronique*

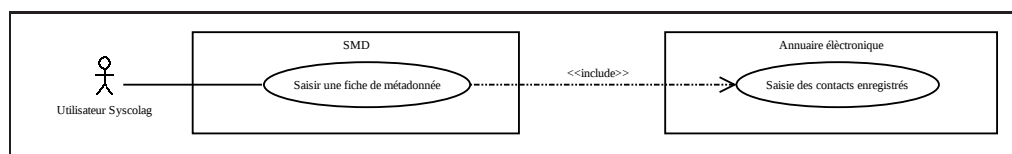


FIG. 7.11 – Description géographique d'une ressource dans l'*ISO 19115*

Comme nous l'avons développé en partie 3.1.1, il s'agit de faire bénéficier les utilisateurs du *SMD* de *valeurs préenregistrées*¹² pour les *éléments de métadonnées* qui permettent la description des *contacts* associés aux métadonnées (*acteurs / organismes*, voir illustration en Figure 7.11). En effet, nous avons vu en partie 4.3.1, que le nombre d'*éléments de métadonnées* associés à la description d'un *organisme* et / ou d'un *acteur* est considérable (42 éléments, voir partie 4.3.1) et que la gestion des valuations qui leur sont associées est possible par le biais d'un *annuaire électronique*. Cela participe au *confort* de la *saisie* en même temps qu'à sa *qualité*. En outre, l'inventaire des *acteurs* mobilisés dans la GIZC en *Languedoc Roussillon* ne pose pas *a priori* de difficultés de réalisation (compte tenu d'un réseau régional d'*acteurs* et d'*organismes* somme toute limité). Une illustration du *composant* utilisé est présenté en Figure 7.12(c).

En dehors des *éléments de métadonnées* relatifs au *contact*, il est intéressant de s'appuyer sur les *bases de références bibliographiques* pour proposer la *saisie automatique* des *éléments de métadonnées principaux* qui correspondent aux champs des *références bibliographiques* déjà valués.

¹¹A terme, il serait souhaitable de proposer la digitalisation de *géométries* complexes sur l'interface cartographique. L'intégration d'*IG* en mode raster dans le *référentiel spatial* étant alors nécessaire.

¹²Dans des tables en dur, voir partie 4.3.3.

Saisie de *Citations* par le biais de la *base de référence bibliographique*

Là encore il s'agit de s'appuyer sur une source potentielle d'*automatisation* de la saisie des *métadonnées* en utilisant la propriété suivante : idéalement à toute *métadonnée* doit correspondre une et une seule *référence bibliographique*. On peut en effet considérer que cette dernière est une *vue* d'une *fiche de métadonnée* qui ne concerne que les *éléments de métadonnées principaux* relatifs à la *citation* de la ressource. Pour preuve, ce jeu d'éléments est assez proche de celui proposé par le *Dublin Core* comme cela a été présenté dans la partie 4.2.2 en Figure 4.5. Le Tableau 7.1 propose quelques correspondances exploitables pour automatiser le remplissage des *éléments de métadonnées* correspondant à des champs de *références bibliographiques*. Une illustration du *composant* utilisé est présenté en Figure 7.12(d).

Champ Bibtex	Élément ISO 19115 correspondant
Author	CI_Citation.citedResponsibleParty
Title	CI_Citation.title
Year	CI_Citation.editionDate
Journal	CI_Citation.series
Editor	CI_Citation.series
Address	CI_Contact.Address
Abstract	MD_Identification.abstract
ISBN	CI_Citation.ISBN
ISSN	CI_Citation.ISSN
...	...

TAB. 7.1 – Tableau de correspondances entre les champs de références bibliographiques déclarés dans *Bibtex* et les éléments de l'*ISO 19115*

Pour conclure cette partie, nous rappelons que les *interfaces graphiques* que nous avons utilisées permettent de renseigner plus facilement certains *éléments de métadonnées*. Ceci est aussi bien valable pour la *saisie* des *fiches de métadonnées*, comme nous venons de l'exposer, que pour la *consultation* où ces éléments interviennent également comme nous allons le présenter dans la partie suivante.

7.2.2 Interfaces graphiques de consultation du SMD

Nous présentons dans cette partie des *interfaces graphiques* qui s'appuient également sur les *composants* présentés dans la partie 7.1 et permettent une *consultation* plus conviviale du contenu de la *base de métadonnées*. Lors de la *consultation* du *SMD*, l'utilisation des *composants* pour la *valuation* des *éléments de métadonnées mot clé, toponyme, contact* est tout à fait similaire à celle que nous avons présentée pour les cas de *saisie* de *fiches de métadonnées*. Elle permet l'élaboration de *requêtes multi-critères* et nous ne revenons pas sur ce qui a déjà été présenté. Nous présentons plutôt la place importante réservée à l'emploi de l'*interface cartographique* pour la *formulation des requêtes spatiales* et la *consultation des réponses* sélectionnées par le système. Cette place est d'autant plus centrale que, dans ce cas, la *consultation* des *fiches de métadonnées* est ouverte au *grand public* (voir cas d'utilisation présenté en Figure 7.13). Ainsi nous privilégions la *médiation spatiale*.

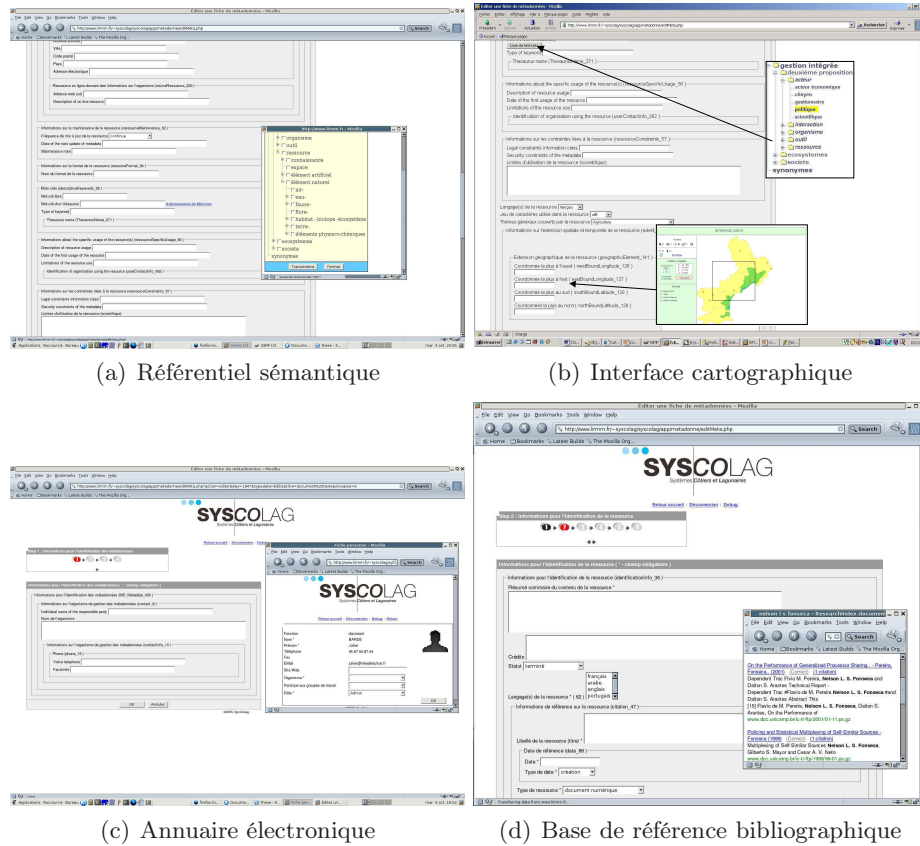


FIG. 7.12 – Insertion des différents *composants* de *contrôle* ou d'*aide* à la *valuation* des *éléments de métadonnées* essentiels pour la GIZC dans les *formulaires HTML* associés aux *gabarits de saisie* du SMD

Elaboration de requêtes multi-critères

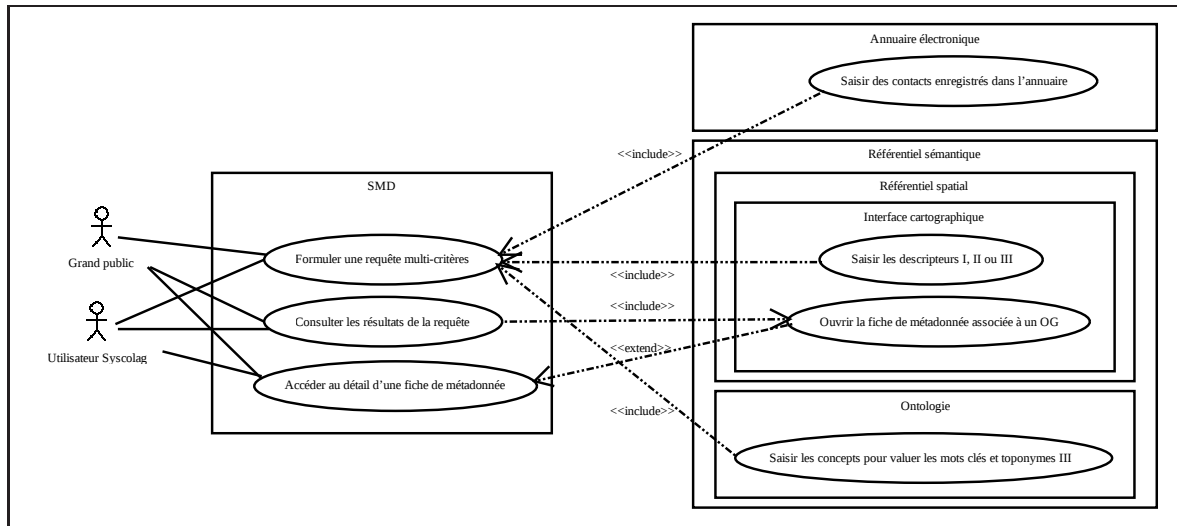
Afin d'optimiser la *consultation de métadonnées*, nous avons décidé de dédier une *GUI* spécifique d'aide à l'*élaboration de requêtes multi-critères* qui s'appuie sur les *éléments de métadonnées* essentiels pour la GIZC.

Cette interface permet donc la construction de *requêtes* qui s'articulent autour de critères permettant de répondre aux questions suivantes (voir illustration en Figure 7.14) :

- *qui ?* avec le support de l'*annuaire électronique* (gestion de l'aspect *contact*),
- *quoi ?* avec le support de l'*ontologie* (gestion des aspects *thématique*, *spatial*, *temporel* en mode *textuel*),
- *où ?* avec le support de l'*interface cartographique* (gestion de l'aspect *spatial* en mode *cartographique*),
- *quand ?* (gestion de l'aspect *temporel* en mode *numérique*),
- *comment ?* (gestion de l'aspect *méthodologique*).

La *saisie des valuations* à l'aide des *composants* s'effectue de manière comparable à celle employée pour la *saisie d'une fiche de métadonnées* (voir partie Figure 7.12).

A l'image de l'importance de l'*espace* dans la GIZC (voir partie 2.1.2), l'*interface cartographique* occupe une place importante dans cette *interface graphique* et permet d'assister, de

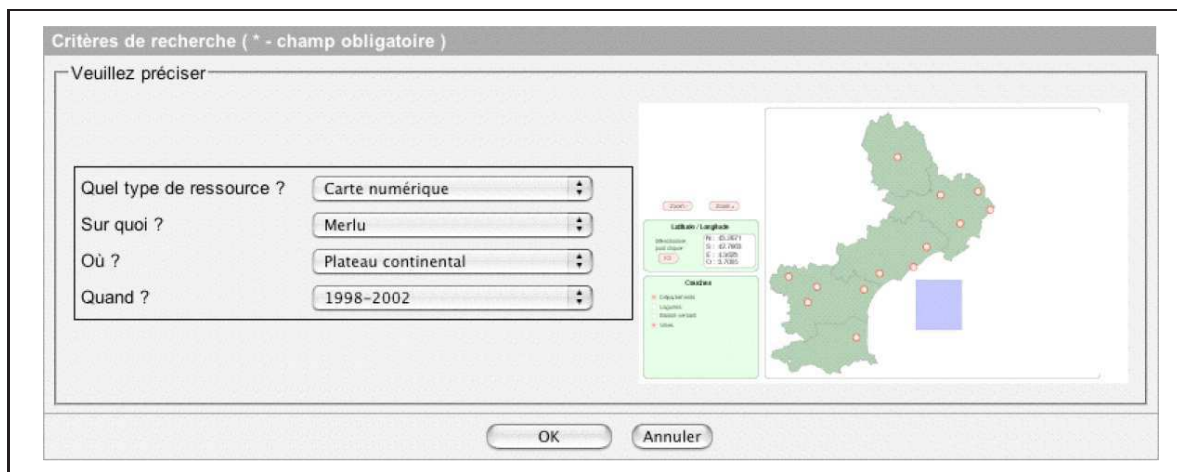
FIG. 7.13 – Cas d'utilisation associés à la consultation du *SMD*

manière transparente, la formulation des *requêtes spatiales* rendues possibles par le lien que nous avons introduit entre *métadonnées* et *référentiel spatial* (voir partie 6.2.3).

L'idée est de permettre à l'utilisateur des *recherches* (et des *restitutions*) de *métadonnées* sans taper de texte mais en utilisant l'*interface cartographique* afin de résoudre des questions du type : "*quelles sont les ressources répertoriées sur cet espace ?*". Par ce biais nous souhaitons rendre l'outil accessible au grand public.

Conformément à ce qui a été présenté en partie 6.1.1, on distingue deux types de *requêtes spatiales* de *consultation* du *SMD* selon le critère spatial (*I*, *II* et *III*) utilisé pour la formuler :

- Généralement, les *requêtes spatiales* qui s'appuient sur *I* ou *III* sont les plus puissantes puisqu'elles spécifient, en plus des *limites topologiques* (précises ou approximées) de

FIG. 7.14 – Exemple d'*GUI* pressenti pour les recherches multicritères pour la consultation de métadonnées

l'*objet géographique*, le *toponyme* associé et donc la *nature thématique* de l'espace de recherche (*lagune*) voire l'*instance* de l'*objet géographique* de recherche (*lagune de Thau*). On peut donc chercher les *métadonnées* associées à un *Objet géographique* identique - le même *toponyme*- (ou de même *nature thématique*) en plus des recherches sur *critères topologiques*.

- Les *requêtes spatiales* qui s'appuient sur II^{13} , quant à elles, ne permettent que l'utilisation de *critères topologiques* grâce aux *fonctions spatiales* de *Postgis* (voir détail en partie 6.2.3).

Le *critère spatial* apparaît donc comme le critère primordial dans l'*interface graphique* que nous proposons pour consulter les *métadonnées* existantes, nous conservons cette approche dans l'*interface graphique* de réponse à la requête formulée.

Réponse à une requête : synthèse des réponses

Une fois la *requête multi-critères* effectuée, il s'agit de retranscrire ce qui dans l'offre correspond à la demande en permettant d'évaluer point par point la pertinence des réponses en fonction des critères clés demandés (*qui, quoi, quand, où, comment*). En particulier il est alors possible de s'appuyer sur une *restitution cartographique des résultats* (même si la formulation de la *requête multi-critères* n'a pas utilisé de critères spatial) (voir cas d'utilisation décrits en Figure 7.15).

Il s'agit d'afficher, dans l'*interface cartographique*, les différents types d'*objets géographiques* qui décrivent les *descripteurs spatiaux* de l'*ISO 19115* dans les *métadonnées* répondant à la requête exécutée (voir partie 6.2.2). En parallèle, nous résumons les réponses sous *forme textuelle* à l'utilisateur (voir illustration proposée en Figure 7.16(a) et 7.16(b)).

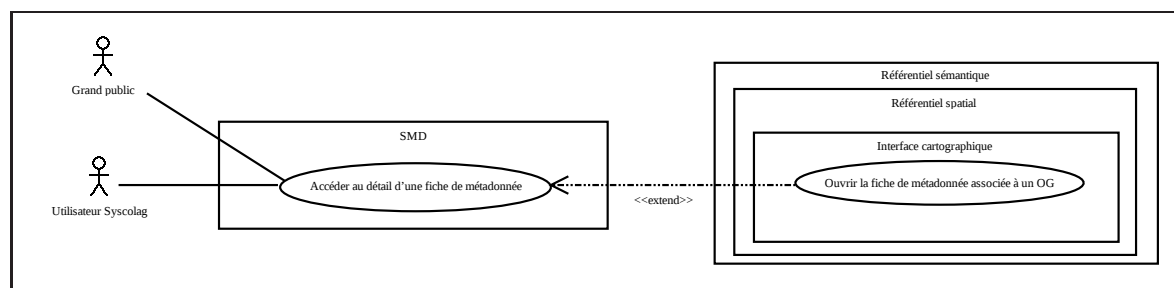


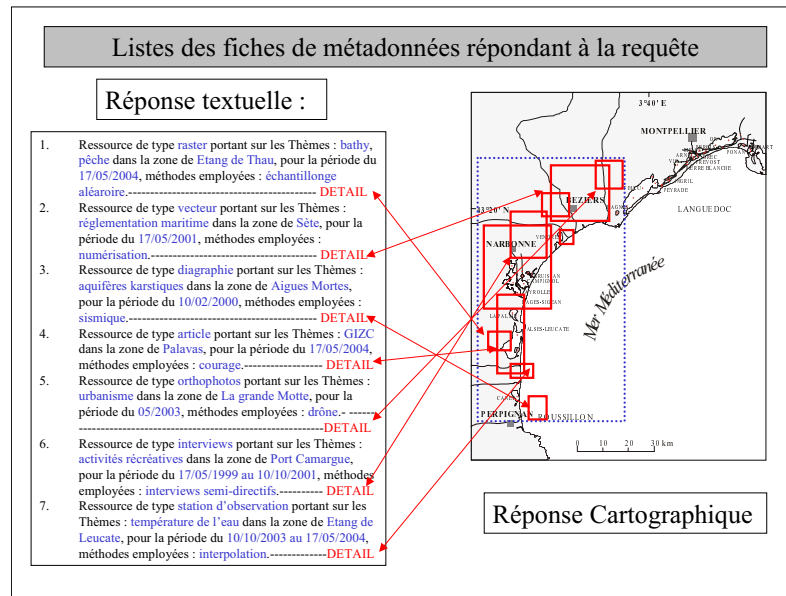
FIG. 7.15 – Cas d'utilisation associés à la consultation des résultats d'une requête

Sur la base des renseignements fournis par cette *interface graphique*, l'utilisateur est en mesure d'effectuer un tri parmi les propositions et de choisir, parmi elles, au détail de quelle(s) fiche(s) de métadonnée il souhaite accéder. La manière d'accéder à la fiche détaillée est laissée à la préférence de l'utilisateur (par un *hyperlien textuel* ou (*carto*)*graphique*).

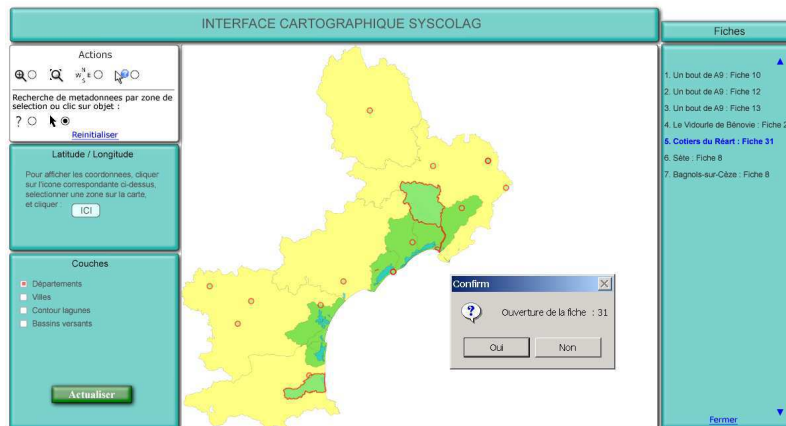
Consultation d'une fiche détaillée

Une fois que l'utilisateur a analysé, grâce aux *GUI* précédentes, les *fiches de métadonnées* pertinentes parmi celles proposées en réponse à sa *requête*, il peut alors accéder à leurs des-

¹³Pour la *consultation*, la note sur l'intérêt d'une *digitalisation* de *géométries complexes* pour effectuer des *requêtes spatiales* de *saisie* plus fines est encore valable.



(a) Réponses textuelles et cartographiques à une requête multi-critère, maquette



(b) Réponses textuelles et cartographiques à une requête multi-critère, prototype en cours

FIG. 7.16 – Maquette 7.16(a) et prototype 7.16(b) d'interfaces graphiques de réponses à une requête de consultation du SMD

criptions détaillées.

Nous ne décrivons pas ici les *GUI* qui présentent le détail d'une *fiche de métadonnées* puisqu'elles correspondent à celles d'un *formulaire de saisie* dont les *champs* sont complétés par les *valeurs* qui ont été attribuées lors de la *saisie* de la fiche et sont *stockées* dans le *SGBDR*. Elles ont donc une forme tout à fait similaire à celle de la Figure 7.12.

Afin de générer les *formulaires de saisie* ou de *consultation détaillée* des *fiches de métadonnées*, un *composant* simplifie l'*édition de nouveaux profils et gabarits* pour la *génération dynamique* des *formulaires* qui leur sont associés.

Administration du SMD : éditeur de gabarit

Une première adaptation de la norme *ISO 19115* à notre contexte a été effectuée par la mise en place d'un *profil* et de *gabarits* (voir partie 4.2.2). Dans la mesure où la *typologie des ressources* susceptibles d'être décrites au sein de ce service peut évoluer, il faut permettre l'élaboration de nouveaux *gabarits*. Or la mise en place d'un *gabarit* reste un exercice complexe pour un utilisateur néophyte (sans connaissance de la norme *ISO 19115* et de ses éléments de métadonnées). Sans assistance graphique, l'exercice n'est a priori pas gérable, nous proposons donc la mise en place d'*GUI* réservées à l'*administrateur* du *SMD* pour réaliser cette tâche (voir Figure 7.17).

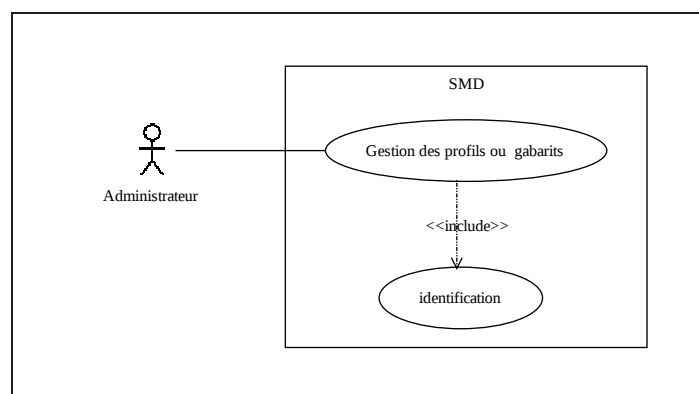


FIG. 7.17 – Cas d'utilisation associé à la génération d'un nouveau *profil* ou *gabarit*

L'*outil d'édition de gabarit* qui a été mis en place (voir Figure 7.18) facilite la tâche complexe de manipulation des rubriques de la norme par une présentation hiérarchisée (qui exploite les relations entre les classes de la norme *ISO 19115*, notamment les liens *BT / NT*, voir partie 4.2.1) des *éléments de métadonnées* que l'administrateur peut retenir en cochant le *bouton radio* qui est associé à chacun d'eux.

En outre, un tel *composant* permet à présent, lors des *entretiens* effectués auprès des utilisateurs, de *visualiser* directement les *formulaires de saisie* ou de *consultation* associés à l'exercice de mise en place de *profils* ou de *gabarits*, ce qui est difficile à imaginer pour l'utilisateur sans cette application.

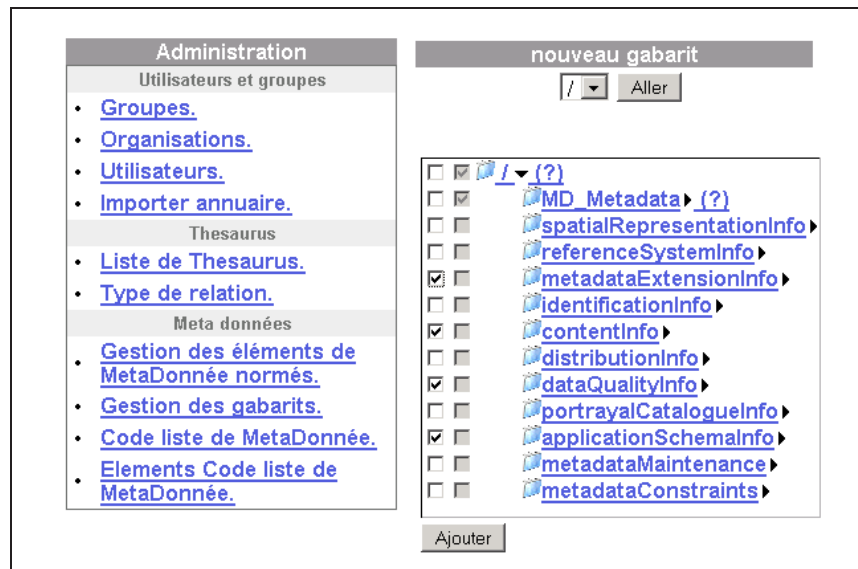


FIG. 7.18 – Outil d'édition de gabarits

Synthèse section 7.2

- Conformément aux souhaits des utilisateurs, les *interfaces graphiques* de *saisie* et d'*aide* à l'élaboration de *requêtes multi-critères* du *SMD* sont construites autour des *éléments de métadonnées* principaux pour la *GIZC*,
- L'*ontologie*, l'*interface cartographique*, l'*annuaire électronique* et la *base de références bibliographiques* peuvent faire l'objet de *composants* développés séparément qui *facilitent* et *contrôlent* la *saisie* des éléments qui leurs sont associés lors de la *production* ou de la *consultation* des *fiches de métadonnées*,
- Comme cela est particulièrement souhaitable pour la *GIZC*, une place privilégiée est réservée à l'*espace* dans le *SMD* par l'emploi d'une *interface cartographique* qui permet à l'utilisateur, de manière conviviale et transparente, dans le *SMD*, la *saisie* complexe des *éléments de métadonnées* qui décrivent la *géométrie* des *objets géographiques* associés aux *métadonnées* ainsi que la formulation de *requêtes spatiales* pour la *consultation*,
- Afin de faciliter également la tâche de l'administrateur, nous proposons un *éditeur de gabarit* pour faciliter la description de nouveaux types de ressources.

Nous avons décrit dans cette partie les *GUI* qui correspondent à la réalisation des tâches de *saisie*, de *consultation* et d'*administration* du *SMD*, objectif initial de ce travail. Pour cela, des *composants* ont été développés pour exploiter le contenu du *référentiel sémantique* pour la *GIZC* (*thématique* et *spatial*). Toutefois, le *modèle conceptuel* du *référentiel sémantique* proposé présente un intérêt qui dépasse la seule problématique d'aide à la *gestion* du *service de métadonnées* et les *composants* peuvent être utilisés *séparément* pour *consulter* la *connaissance* contenue dans le *référentiel sémantique* en dehors de l'utilisation qui en est faite dans le *service*

de métadonnée.

Nous décrivons à présent deux autres cas d'utilisations de ces composants pour exploiter le contenu de notre *référentiel sémantique* : une *base de connaissances* qui exploite l'ensemble du *référentiel sémantique* et un *atlas cartographique* qui exploite la partie *géographique* ou *spatiale* du *référentiel sémantique*.

7.3 Interfaces graphiques de gestion de la connaissance

La *gestion* du *service de métadonnées* proposé a nécessité la mise en place de *composants* complémentaires essentiels pour répondre au verrou *sémantique* du *contrôle de la valuation*, en *saisie* comme en *consultation*, des *éléments de métadonnées* essentiels pour la GIZC (éléments *thématiques* et *spatiaux*). Ceci est indispensable pour la *pertinence* des *fiches de métadonnées* proposées en réponse à une *consultation*. Cependant, l'exploitation du contenu du *référentiel sémantique* par ces *composants* permet plus que la simple mise en place d'un *vocabulaire contrôlé* et la *saisie* de *descripteurs spatiaux* en mode *cartographique* dans le *SMD*.

En effet, la *connaissance inventoriée* et *formalisée* dans notre *SGBDR* peut être restituée de manière beaucoup plus fine et conviviale en dehors du contexte d'un *SMD* par des *composants* utilisés de manière indépendante en tant qu'applications dédiées à la réalisation des *cas d'utilisation* illustrés en Figure 7.19. Nous présentons dans cette section l'exploitation du *référentiel sémantique* comme *base de connaissances* consultable soit par le biais d'un *réseau sémantique* (pour *visualiser* la nature des *relations* inventoriées entre les *concept* collectés) soit par le biais d'une *base terminologique* qui insiste d'avantage sur l'exploitation des *propriétés sémantiques* attribuées aux *concepts*. Puis nous détaillons l'exploitation spécifique du *référentiel spatial* par le biais d'un *atlas cartographique*.

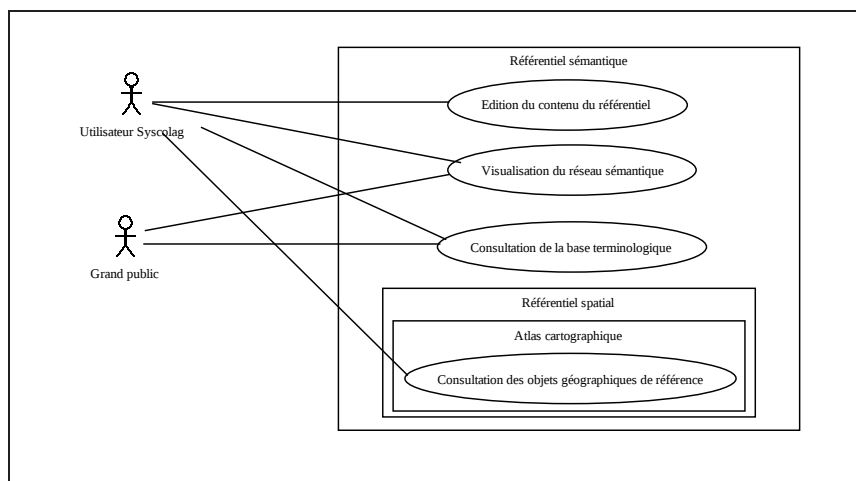


FIG. 7.19 – Les différents cas d'utilisation liés à la visualisation du contenu du *référentiel sémantique*

7.3.1 Ontologie - Réseau sémantique - Base de connaissance

L'*outil de contrôle terminologique* pour la *valuation* des *éléments de métadonnée mot clé* et *toponyme* (voir Figure 7.8) n'est qu'une utilisation partielle du contenu du *référentiel sé-*

mantique pour renseigner ces éléments dans les *GUI* de *saisie* ou de *consultation* du *SMD* qui répond à la problématique présentée en partie 5.1.1.

Toutefois ce *cas d'utilisation* est restrictif et ne représente qu'une des exploitations possibles de l'intérêt du *référentiel sémantique* (voir Figure 7.20). Ce dernier, en tant que tel, peut être dédié à la *consultation* de la *connaissance* qu'il contient soit sous la forme d'une *base terminologique* (comme un *glossaire*, un *thésaurus*, une *ontologie* voire une *base de connaissances*), soit de manière plus conviviale à l'aide un *réseau sémantique*.

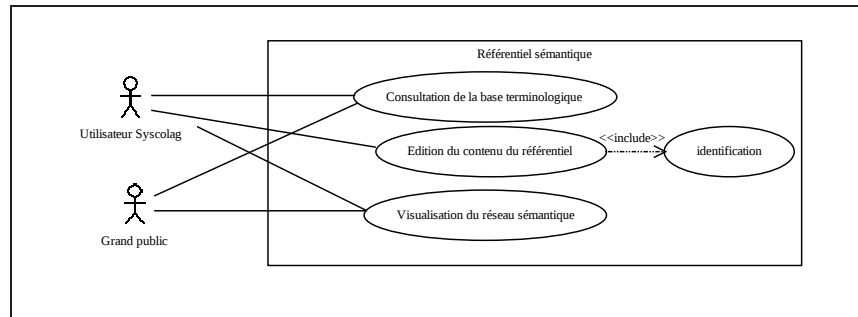


FIG. 7.20 – Les différents cas d'utilisation liés à la visualisation du contenu du *référentiel sémantique* hors représentations cartographiques

Base terminologique

Cette présentation est adaptée à la réalisation des *cas d'utilisation* pour la *consultation* et l'*administration* (en gardant la même structure de formulaire mais en rendant les valeurs des champs éditables directement par des utilisateurs identifiés, comme pour la production de *fiches de métadonnées*). Il s'agit de fournir une information de référence pour les utilisateurs qui souhaitent consulter la *signification* d'un *concept* employé en *GIZC*.

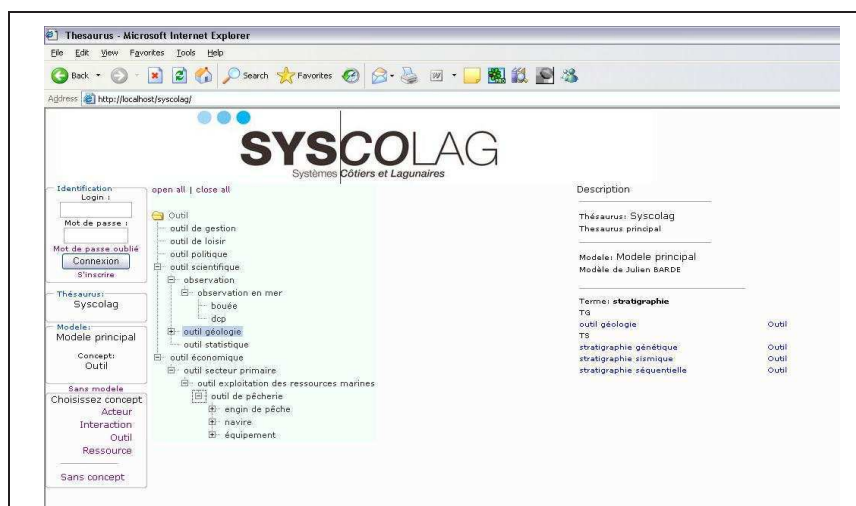
Pour cela nous proposons de restituer exhaustivement les *propriétés sémantiques* des *concepts* contenus dans le *référentiel sémantique* :

- les *définitions* éventuellement multiples (*polysémie*),
- les *synonymes*,
- les autres *relations sémantiques* énumérées dans le *modèle conceptuel* du *référentiel sémantique* en Figure 6.5.

Cette approche est tout à fait comparable aux différentes *bases terminologiques* présentées dans notre état de l'art (en partie 5.1.2) sur les *glossaires*, *thésaurus* ou *ontologies* qui existent dans notre domaine ou à des outils plus génériques comme le *Grand dictionnaire terminologique* [149].

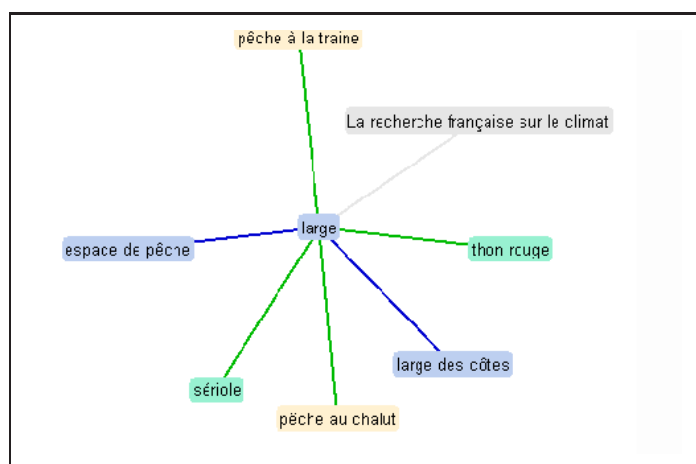
On remarque au passage dans ce *prototype* qu'il est possible pour naviguer dans la *base terminologique* de réutiliser le *composant* présenté en Figure 7.8. Ces *GUI* servent donc de support au *partage* voire à l'*élaboration* d'un *langage commun*. En effet, nous souhaitons permettre aux différents champs disciplinaires mobilisés dans la *GIZC*, si nécessaire, de proposer une *définition* particulière à leur spécialité par les relations de *polysémie*, mais il est souhaitable autant que possible de partager les *concepts* en même temps que leurs *définitions*.

L'alternative à ces cas d'utilisation est une représentation plus conviviale, plus orientée *grand public*, sous la forme d'un *réseau sémantique*.

FIG. 7.21 – Visualisation du *référentiel sémantique* sous la forme d'une *base terminologique*

Réseau sémantique

Par comparaison avec la Figure 7.8, la représentation de la Figure 7.22 n'est jamais qu'une forme plus complète de représentation des *relations qu'entretiennent les concepts* que nous avons *inventoriés* dans notre *référentiel sémantique*. Cette représentation sous forme de *réseau sémantique* offre une navigation plus conviviale dans la *connaissance* que nous avons inventoriée sur le domaine de la GIZC en LR et permet de représenter graphiquement et de manière *exhaustive* les *relations présentes entre les concepts* de notre *référentiel sémantique*.

FIG. 7.22 – Visualisation du *référentiel sémantique* sous la forme d'un *réseau sémantique*

En particulier il serait intéressant de mélanger les deux approches et d'intégrer un *réseau sémantique* dans l'*GUI* précédente pour restituer les différentes *propriétés sémantiques* associées à un *concept* en cours de visualisation dans le *réseau sémantique* (voir Figure 7.23).

Les *cas d'utilisation* que nous avons présentés sont applicables à tous les *concepts*, en particulier aux *concepts spatiaux*. Cependant, nous avons vu que ces derniers sont une sorte

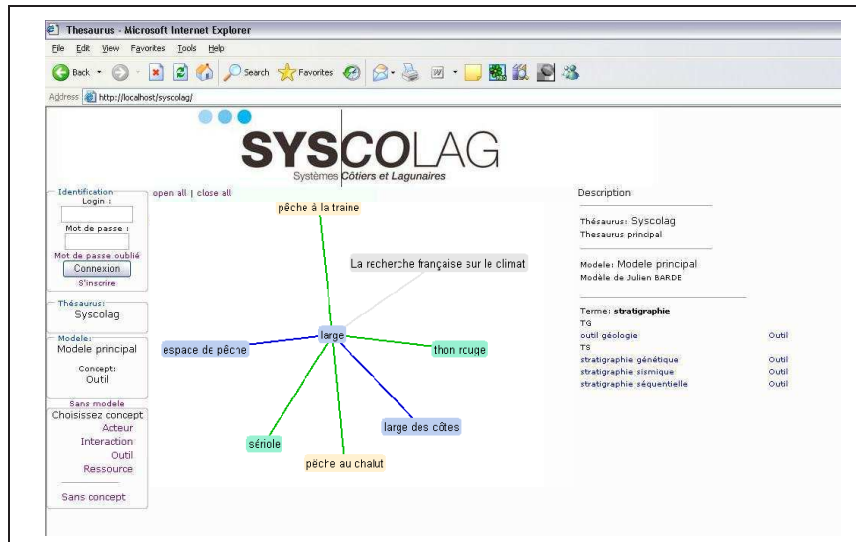


FIG. 7.23 – Exemple d'approche mixte entre *base terminologique* et *réseau sémantique*

de *concepts thématiques* dont les *instances* possèdent des propriétés particulières (voir partie 6.1.2) qui permettent leur *représentation cartographique*. Nous souhaitons donc permettre aux utilisateurs une utilisation optimale des *concepts spatiaux* et *objets géographiques* contenus dans le *référentiel spatial* et qui leur permette en particulier d'exploiter les *relations spatiales* existantes entre eux.

7.3.2 Atlas cartographique

L'importance de l'espace dans la *GIZC* est telle que nous proposons une *vue* du *référentiel sémantique* qui ne retient que les *concepts spatiaux* et *objets géographiques* associés : le *référentiel spatial* (voir chapitre 6).

Intérêt de l'atlas cartographique

Nous proposons, à l'instar de ce qui a été mis en place pour la *consultation* du *référentiel sémantique*, de présenter le contenu du *référentiel spatial* en réutilisant le *composant* avec lequel nous avons généré l'*interface cartographique* au sein du *SMD* en l'enrichissant de fonctionnalités supplémentaires. Ceci constitue ce que nous appelons un *atlas cartographique*. L'objectif de l'*interface cartographique* du *SMD* est l'optimisation de l'utilisation de *notre service de métadonnées*, en *saisie* comme en *consultation*, par une assistance graphique pour renseigner les éléments complexes pour la description de l'*étendue géographique concernée par la ressource* (voir partie 6.1.1). La réalisation de cet objectif a nécessité l'*inventaire* d'un nombre conséquent de couches d'*information géographique* (voir partie 6.3). A ce titre l'information collectée présente un intérêt majeur comme *base d'objets géographiques de référence pour la GIZC* ou *référentiel géographique*, que ce soit pour être utilisée dans un *service de métadonnées* ou dans tout autre type d'application. Il serait cependant dommage de ne pouvoir consulter l'*IG de référence* collectée que par le biais de la *saisie* ou de la *consultation* des *métadonnées*.

Nous proposons donc, pour valoriser cet effort qui n'est pas exploité à sa juste valeur, la mise en

place d'une application Web dédiée à une *visualisation / consultation* plus fine de l'*IG* contenue dans ce *référentiel spatial*. Pour cela, notre prototype exploite le *composant* développé en *SVG* qui a servi pour l'*interface cartographique* du *SMD* présentée en partie 7.2.1 et que nous proposons d'enrichir de nouvelles fonctionnalités (puisque cette fois l'application n'est plus intégrée dans un formulaire du *SMD* et ses contraintes de légèreté sont donc moindres). Cette application, ouverte au *grand public* (voir *cas d'utilisation* présenté en Figure 7.24), offre les fonctions essentielles à la *visualisation* de l'information contenue dans notre *base de données spatiales* et correspond à un *atlas cartographique* [169] pour la *GIZC* en *LR*.

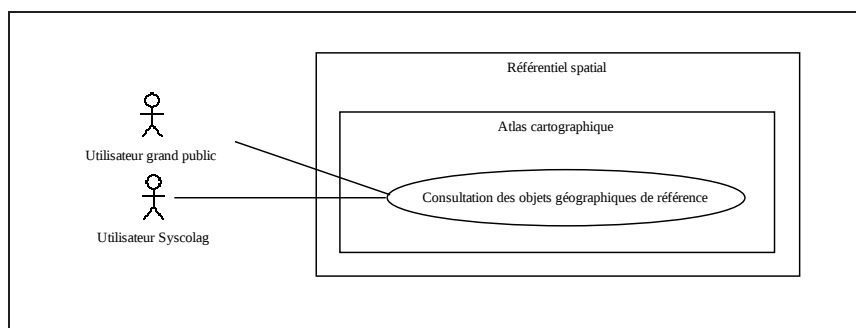


FIG. 7.24 – Consultation du contenu du *référentiel spatial* à l'aide d'un *atlas cartographique*

Fonctionnalités associées à l'atlas cartographique

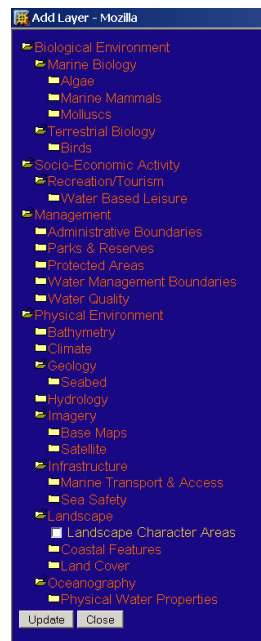
On constate actuellement une émergence considérable d'*atlas cartographiques* dans le domaine *environnemental* et de *côtier* en particulier [58] basés sur des *solutions libres* et qui emploient une *structuration thématique* pour la présentation des *Layers* offerts en consultation (voir Figures 7.25(a)¹⁴ et 7.25(b)¹⁵). Les solutions présentées en Figures 7.25(c) et 7.25(d) utilisent *Mapserver*. On constate au passage le faible rendu visuel de ces applications (bien qu'il existe une sortie *SVG* pour *Mapserver*. Son exécution est cependant lente).

Pour notre part, nous utilisons une *structuration thématique* des *Layers* générée *dynamiquement* à partir de notre *référentiel sémantique* grâce à la méthode proposée en partie 6.3.4. En pratique, nous contribuons par ce biais à diffuser, de manière beaucoup plus large qu'avec le *SMD*, l'*IG* de référence pour décrire les problématiques associées à la *GIZC* en *Languedoc Roussillon* (voir partie 6.3) et permettons, par une *consultation* sur le Web, un *accès facile* et propice à l'*enrichissement quantitatif* et *qualitatif* de notre *inventaire*.

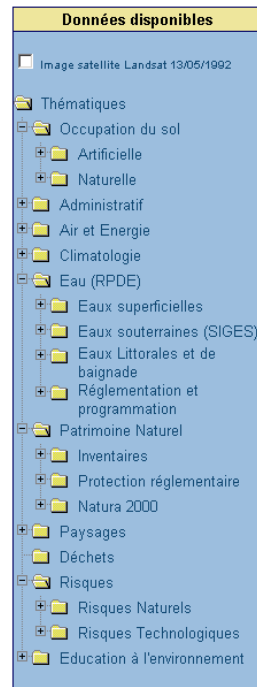
En outre, en nous appuyant sur les spécifications de l'*OGC* (en particulier *WMS*), nous transformons cet *atlas* en *serveur cartographique* et assurons son *interopérabilité* avec d'autres *serveurs* ou *clients WMS* au niveau régional, national voire international (voir partie 6.2.4). Nous élargissons ainsi le spectre de son utilisation en dehors de notre système [79]. Ce point est capital et permet par exemple d'envisager à terme l'*interopérabilité d'atlas cartographique* (autres serveurs *WMS*) sur d'autres zones côtières (par exemple pour les zones voisines : *Catalogne*, *PACA* ou *Ligure*...). Cela peut participer à la constitution ou l'harmonisation de données géographiques côtières de référence et transfrontalières conformément aux

¹⁴Marine Irish Digital Atlas, <http://mida.ucc.ie/pages/atlas/atlas.php>.

¹⁵<http://www.observatoire-environnement.org/OBSERVATOIRE/environnement-cartes.html>.



(a) Structuration thématique des Layers proposés par le MIDA



(b) Structuration thématique des Layers proposés par l'Observatoire de l'environnement en Poitou-Charentes



(c) Atlas cartographique du MIDA



(d) Atlas cartographique de l'Observatoire de l'environnement en Poitou-Charentes

FIG. 7.25 – Structuration thématique des Layers et visualisation de leur contenu par des interfaces cartographiques sur le Web

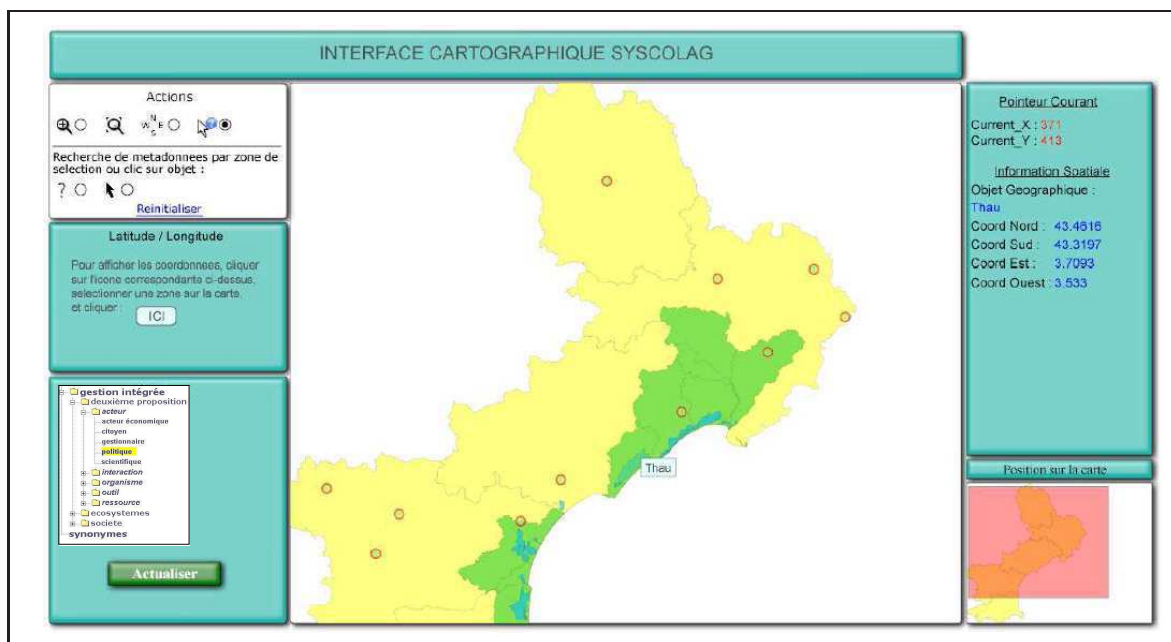


FIG. 7.26 – Extension de l'utilisation de l'*interface cartographique* du *SMD* en *atlas* indépendant

recommandations du *Développement Durable* (voir partie 2.1.1). Il sera ainsi possible d'utiliser cette information pour la mise en place d'*atlas cartographique* à plus grande échelle et de contribuer au croisement des expertises.

En outre, la *base de données spatiales* sur laquelle s'appuie notre *atlas* est exploitable par d'autres applications utilisant une connexion distante (protocoles Web ou autres) : des *interfaces cartographiques* Web qui permettent une *gestion* plus poussée de l'*IG* (édition, spécification *WFS* de l'*OGC*, voir partie 6.2.4) voire l'*administration* de la *BDS* et surtout par des *clients SIG libres et standards* qui ouvrent des perspectives d'accès et de *gestion* partagés de l'*IG* inaccessible jusqu'ici dans un environnement libre et convivial. En outre, l'administration peut s'appuyer également sur les applications énumérées en partie 7.1.1 ou directement en lignes de commandes. Lors de l'*import* du *Layer* lui-même, pour produire une *arborescence de Layers* qui utilise la *structuration thématique* de notre *ontologie*, la seule contrainte de l'administrateur est de s'assurer de l'existence d'un lien entre le *concept spatial* associé au *Layer* et un *concept thématique* géré dans le *référentiel sémantique* (voir Figure 6.4 partie 6.3.4).

Cette démarche est donc un premier pas concret vers la *mutualisation et le partage physique de ressources* et d'outils communs bien que, pour le moment, les *Layers* sont stockés dans un même base de données spatiale, alors que l'implémentation de *WMS* permet la gestion de ressources distribuées (voir détail en partie 8.2.2).

Toutefois l'utilisation à part entière de cet *atlas cartographique* n'empêche pas de proposer des *hyperliens* vers les autres outils comme le *SMD* ou l'*ontologie* (et réciproquement), puisque nous avons vu que le *MCD* que nous proposons intègre ces différents *composants* au sein d'une même base. De cette sorte l'accès aux trois principaux *outils* que sont le *SMD*, la *base terminologique* et l'*atlas cartographique* peut se faire à partir de n'importe quelle entrée (voir

Figure 7.27).

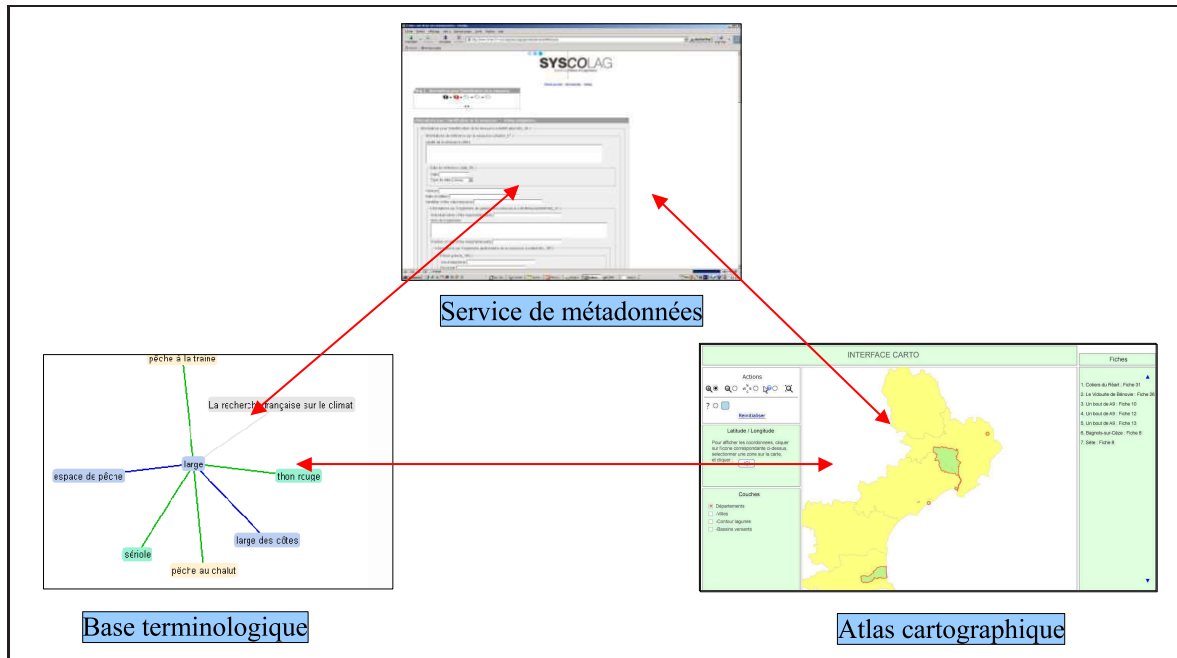


FIG. 7.27 – Passage d'une application à l'autre par liens hypertextes

Synthèse section 7.3

- En complément du *SMD*, nous proposons la *consultation* séparée de notre *référentiel sémantique* pour l'intérêt qu'il présente auprès des acteurs de la GIZC afin d'élaborer un *langage commun*.
- Les *concepts* qui y sont contenus sont consultables par le biais d'une *base terminologique* ou d'un *réseau sémantique*, voire les deux en même temps.
- Les *concepts spatiaux* peuvent être exploités par le biais du *support cartographique*. Le contenu du *référentiel spatial* peut ainsi être porté à connaissance des acteurs et constitue un *atlas cartographique* de la GIZC.
- Quelle que soit l'application consultée par l'utilisateur, il existe des passerelles (*hyperlien*) vers les deux autres applications complémentaires (voir Figure 7.27).

Conclusion du Chapitre 7

Ce chapitre présente les prototypes d'*interfaces graphiques* qui permettent la réalisation des *cas généraux d'utilisation* présentés dans la partie 3.3. En particulier, nous présentons, plus en détail, ceux relatifs à la gestion du *SMD* ainsi que celles des *référentiels sémantique* et *spatial*. Pour cela nous nous sommes appuyés sur un *environnement de développement libre* pour les langages (*PHP* comme langage de script) comme pour le *SGBDR* (*Postgres* et sa *cartouche spatiale Postgis*) ou les applications que nous avons utilisés pour développer des solutions spécifiques. Ces *composants* exploitent les contenus de la *base de métadonnées* et

du *référentiel sémantique*. Selon la nature des *utilisateurs*, des *éléments de métadonnées* ou des *concepts* concernés, il en résulte différentes *interfaces graphiques* utilisables pour le *SMD* ou de manière indépendante pour *servir l'information* et la *connaissance* hébergées par notre *infrastructure*.

En particulier, l'*interface cartographique* utilise une visualisation de l'*IG* stockée sous forme alphanumérique dans *Postgis* par le biais de *SVG* en utilisant la fonction *assvg* de *Postgis* alors que le *référentiel sémantique* est représenté graphiquement par l'utilisation d'un visualisateur de *réseau sémantique* utilisant une *librairie Java* (application *Prefuse*).

Par ce biais les *interfaces* de *saisie* et de *consultation* du *SMD* sont plus conviviales, simples à utiliser et la qualité des saisies s'en trouve améliorée.

Nous respectons donc les objectifs annoncés dans le chapitre 3, en particulier la prise en compte des attentes des utilisateurs et de l'importance de l'espace dans la *GIZC*. En particulier, la *consultation* du *SMD* est possible uniquement par le biais de l'*interface cartographique* et peut se substituer à l'emploi complémentaire de *mots clés* pour alimenter un *moteur de recherche* sur des *requêtes* en *mode textuel* plus traditionnelles. En outre, nous tirons profit de l'effort de mise en place du *référentiel sémantique* en dehors du *SMD*. Il est possible de consulter séparément les *concepts* de référence pour la *GIZC* que nous avons collectés et structurés (voir partie 6). Nous proposons un mode de visualisation du type *base terminologique* ou un mode de visualisation du type *réseau sémantique*. Plus particulièrement, nous proposons la *consultation* du *référentiel spatial* par le biais d'un *atlas cartographique* structuré par le biais de notre *ontologie spatiale* (voir partie 6.3.4).

Synthèse du Chapitre 7

- Nous nous appuyons sur des *composants* réutilisables pour la génération de différentes *GUI* qui correspondent aux *cas d'utilisation* du système que nous proposons.
- Les développements sont effectués dans un *environnement de développement libre* et utilisent différents langages ou applications de référence selon la nature des éléments à traiter.
- Dans le *SMD*, les *GUI* facilitent la gestion d'éléments dont nous souhaitons *contrôler* ou assister la *valuation* en réponse à leur importance, leur complexité ou simplement à la lourdeur de saisie qu'ils engendrent (**concepts** pour la *valuation* des *mots clés*, *descripteurs spatiaux*, éléments de *contact* ou de *références bibliographiques*).
- La place réservée à l'*interface cartographique* est importante en réponse à la volonté de privilégier l'utilisation du *support cartographique* pour le *partage d'information* et de *connaissance*, en particulier nous proposons un *atlas cartographique* pour la consultation détaillée du contenu du *référentiel spatial*.
- Le détail du contenu du *référentiel sémantique* est également accessible, indépendamment du *SMD*, sur le *portail Web* afin de favoriser le partage de la signification des termes entre les *acteurs* de la *GIZC*, soit sous la forme d'une *base terminologique* soit sous la forme d'un *réseau sémantique*.

Chapitre 8

Conclusion et Perspectives

8.1 Conclusion

Le *Développement Durable* (DD) et la *Gestion Intégrée de la Zone Côtière* (GIZC) sont deux concepts récents. Les *plans d'actions* qui visent à les mettre en œuvre sont articulés autour de nombreux *principes* dont la portée opérationnelle est encore floue. Parmi ceux-ci, la *gestion de l'information* et de la *connaissance* occupe une place centrale dans la réalisation des processus de DD et de GIZC mais l'hétérogénéité des *acteurs*, des *disciplines* ainsi que des *ressources* informationnelles et *SGBD* associés complexifie lourdement cette tâche. Pour améliorer la *mutualisation* de l'*information* et de la *connaissance*, un premier niveau de réponse à ces contraintes est de permettre la *localisation* des *ressources informationnelles* réparties chez les *acteurs* mobilisés dans la GIZC avant de permettre leur *partage physique* et d'éventuels *traitements*. En outre, dans ce domaine, les problèmes récurrents de *terminologie* illustrent la nécessité de partager un *référentiel sémantique* qui permettra, entre autres, de contribuer à établir un *langage commun* au bénéfice aussi bien des *acteurs* que des *systèmes informatiques*, notamment pour la description du contenu *thématique* et *spatial* des problématiques de GIZC. La mise en place d'une telle *gestion* de l'*information* et de la *connaissance* s'appuie obligatoirement sur une implication forte de la part des acteurs (concepteurs et futurs utilisateurs) dans la démarche. La prise en compte de ce *facteur humain* est déterminante pour une réponse pertinente de l'outil à cette problématique et peut, d'après ce qui nous a été permis d'observer dans le cadre du programme *Syscolag*, servir d'amorce et de point d'ancrage aux dialogues interdisciplinaires. Au regard de ces contraintes, les apports du Web et du support cartographique comme outils de travail collaboratif sont précieux.

En prenant en compte l'ensemble des *verrous* évoqués ci-dessus et certaines des *réponses* apportées par des programmes de recherche, nous proposons, en réponse à notre problématique, un portail Web hébergeant un *service de métadonnées* (SMD) complété par la mise en place d'un *référentiel sémantique* que nous exploitons à la fois comme un *outil de contrôle de la valuation des éléments de métadonnées* de notre SMD et que nous diffusons pour la mise en place d'un *langage commun* et d'un *outil de partage des connaissances* pour les acteurs mobilisés dans notre programme. En outre, pour réaliser ces objectifs, nous préconisons d'appuyer cette démarche de *mutualisation* par l'utilisation des propriétés de l'*information géographique côté utilisateur* et par l'utilisation de *normes et standards* de référence *côté système* pour assurer la pertinence, la pérennité et l'évolutivité des solutions que nous allons mettre en place.

Concernant la mise en place d'un *service de métadonnées* pour la localisation des ressources existantes, compte tenu de l'importance de l'*espace* dans les processus de GIZC, nous proposons de nous appuyer sur la norme *ISO 19115* relative aux métadonnées pour l'information géographique. Si cette norme est *générique*, elle supporte néanmoins une adaptation assez fine et permet ainsi de répondre aux besoins essentiels *spécifiques* des problématiques de GIZC. En particulier, la norme *ISO 19115* apporte des réponses pertinentes aux principaux *verrous* identifiés, soit directement avec des *descripteurs spatiaux* élaborés et standardisés, soit indirectement en suggérant l'*extension* du *jeu d'éléments* qu'elle fournit, en particulier pour le *contrôle des valuations* de l'élément de métadonnée *mot clé* (elle suggère l'indexation de la ressource à l'aide de *concepts* dont la signification est contrôlée). En outre, sa *généricité* apporte des avantages : une portée internationale et la qualité de ses descripteurs qui assurent l'*interopérabilité* de notre système avec d'autres implémentant ses recommandations. A la suite d'entretiens, nous avons effectué, en outre, des adaptations en fonction de notre contexte d'implémentation par la prise en compte des remarques des partenaires de notre projet (futurs utilisateurs du système proposé).

Pour ce qui concerne l'extension de la norme *ISO 19115*, nous proposons une *méthode générique d'extension de la norme*, que nous avons implémenté dans le cas spécifique de la GIZC en Languedoc Roussillon. En effet, dans le contexte de la GIZC, il nous paraît capital de mettre en place un *référentiel sémantique* pour ce *domaine* qui permet de répondre aux objectifs spécifiques de *contrôle du vocabulaire* employé dans un *SMD* et à ceux, plus généraux, de partage de la *connaissance*. Nous proposons la mise en place d'une *ontologie a priori* plutôt qu'un *glossaire* ou un *thésaurus*, pour bénéficier des perspectives de *gestion* de la *connaissance* ouvertes par un tel outil. Afin d'*inventorier* les *concepts* de référence de notre *domaine* ainsi que les *relations* qu'ils entretiennent, l'analyse des *sources de référence* du domaine associée à celle de l'expertise de nos partenaires nous ont conduit à formaliser un *modèle conceptuel d'ontologie a priori* qui nous sert de *patron* pour l'*intégration de la connaissance experte du domaine*. En particulier, ce modèle permet de retranscrire la nature des interactions entre *éléments anthropiques* et *naturels* qui sont l'*objet* des *principes* du *DD* et de la *GIZC*.

En outre, afin de répondre au souhait de donner à l'*espace* une place centrale dans le *partage d'information* et de *connaissance* associées à la GIZC, nous analysons les *liens* entre *concepts thématiques* et *spatiaux* et enrichissons notre *modèle conceptuel* de sorte qu'il devient possible de gérer en plus des *relations ontologiques* "conventionnelles", des *relations spatiales* qui font du *référentiel sémantique* une *ontologie spatiale*. Ce sous-ensemble du *référentiel sémantique* peut être alimenté, cette fois encore, par les sources de référence du domaine ainsi que par l'analyse de l'expertise de nos partenaires. Il en résulte un *référentiel spatial* pour les *objets géographiques de référence* de la GIZC. Ce référentiel permet notamment d'introduire des *requêtes spatiales* pour la *localisation* de l'*information* et de la *connaissance* inventoriées dans le *référentiel sémantique*, et d'élaborer un *atlas cartographique* pour diffuser l'*information géographique* de référence du domaine.

Nous nous appuyons sur les potentialités de cette *infrastructure de données spatiale* pour proposer des *prototypes* de logiciels qui en exploitent le contenu sous la forme d'un *portail Web* composé essentiellement d'un *service de métadonnées* et d'une *base de connaissance* (au sens *ontologique* du terme) qui présente le contenu du *référentiel sémantique* aux utilisateurs (par des *composants* permettant notamment la visualisation d'une *base terminologique*, d'un *réseau sémantique* et l'emploi d'un *atlas cartographique* pour le cas spécifique des *concepts spatiaux*). En outre, l'ensemble des développements présentés s'appuie sur des applications libres de droits et respecte les normes et standards garants, *a priori*, de la qualité, la longévité et le

respect de l'esprit communautaire et d'indépendance nécessaires pour les solutions apportées dans le cadre d'un mode de gestion lié au *Développement Durable*. En effet, la gratuité des solutions proposées est un argument de poids si on veut les rendre accessibles (surtout dans les cas où gratuité rime avec qualité).

Cependant, derrière ces multiples efforts de développements de *composants* a priori séparés par des exigences techniques diverses (*portail Web, ontologie, base de références bibliographiques, SMD, interface cartographique...*) qui peuvent brouiller / faire perdre de vue l'objectif initial, il convient de rappeler que tous sont au service d'une même application qui consiste à *favoriser la mutualisation de l'information et de la connaissance* associées aux *ressources informationnelles* de la GIZC existantes. Le *modèle conceptuel* que nous proposons unifie les différents *verrous de description de l'information existante* dans les processus de GIZC. Peu importe finalement le nom attribué à cet outil (*service de métadonnées, SI, référentiel, base de données spatiales...*), la *norme* ou le *jeu d'éléments de métadonnées* choisis pour l'implémenter, ne doit subsister / apparaître dans ce modèle que l'*intention de contrôler* nécessairement la *signification (syntaxe) des descripteurs de l'information, des valeurs qui les renseignent et de la connaissance qui leur est associée* si on veut apporter une réponse pertinente. Les développements qui exploitent ce contenu et les techniques employées (*applications, langages, protocoles...*) ne sont qu'une mise en forme potentiellement critiquable de ce *modèle* pour permettre la résolution des *verrous sémantiques* liés au *partage d'information et de connaissance* dans la GIZC. Les développements qui satisfont actuellement la démarche seront peut être obsolètes à court terme mais les *facteurs humains* pris en compte dans la réponse apportée (puisque sous-jacent à ces problèmes de *gestion de l'information et de la connaissance*) ne fluctuent a priori pas aussi vite que l'évolution technologique. Seul l'amélioration du *modèle conceptuel* et du *modèle d'ontologie* proposé pour l'exploiter peuvent réellement changer la pertinence de la réponse. Toutefois, les réponses éprouvées du *Web*, des *métadonnées* et des *standards* ou *normes* employés et l'unanimité autour de la nécessaire mise en place d'*ontologies* pour répondre à ces problématiques (en particulier pour la mise en place du *Web sémantique*), laissent penser que si le nom des futures solutions peut changer, leur nature (*ontologique*), quant à elle, semble confirmée comme en témoigne la convergence des conclusions tirées par les projets de recherches en cours sur ces sujets et des orientations prises par ceux à venir.

En outre, nous avons pu mesurer comment ce travail a permis de *fédérer* les *acteurs* autour d'une problématique commune et a ainsi consolidé l'implication humaine nécessaire dans ce type de projet de *gestion intégrée*. En effet, la mise en place et la pertinence d'un *outil commun* qui doit conjuguer une démarche *volontaire de description de l'information et de la connaissance* existantes en employant un *langage commun* ne peut se réaliser sans l'agrément et l'adhésion des partenaires impliqués dans un tel projet (en particulier pour ce qui concerne l'*intégration de la connaissance experte du domaine*). Le succès de la démarche (ou son échec) dépasse donc la qualité de l'outil proposé. Peut être même à l'inverse peut-on se demander si l'échec d'un outil adapté ne montre-t-il pas celui actuel de la motivation des acteurs dans ces modes de *gestion intégrée*. En effet, en se reposant sur la bonne volonté des acteurs sans contraintes de participation, il est possible que le renoncement prenne le pas sur l'engouement. Ainsi peut-on expliquer l'enjeu majeur de l'*automatisation* du remplissage des *métadonnées* pour pallier la motivation parfois défaillante des *acteurs*. Malheureusement tous les éléments ne sont pas automatisables et il faudra bien argumenter voir encourager la participation des acteurs. Peut-être alors ne faut-il pas trop laisser miroiter que la technologie permettra à terme l'*automatisation* de la production de *métadonnées* de manière transparente pour ses utilisateurs.

Ainsi espérons nous apporter une contribution à la mise en œuvre des principes du *Développement Durable* et de la *gestion intégrée de la zone côtière* afin que la richesse naturelle du *continuum terre mer* perdure *ad vitam eternam*. Amen.

8.2 Perspectives

Ce travail a donc abouti sur une voire des méthodes pour la *mutualisation d'information et de connaissances pour la GIZC* qui doivent à présent être implémentées dans le cadre du projet *Syscolag* en vue d'atteindre les objectifs fixés. En outre, il apparaît que les méthodes proposées sont *génériques* pour les problématiques de *mutualisation* de l'*information* et de la *connaissance* dans la *GIZC*, voire même pour la *gestion environnementale d'écosystèmes anthropisés* associée au *DD*. Le *prototype* est validé du point de vue de la pertinence de ses différents *composants* et permet l'*interopérabilité* des *SGBD* des partenaires mobilisés au sein d'un projet et potentiellement, avec d'autres systèmes identiques ou équivalents en réseau sur le Web (basés sur les mêmes standards). La pertinence du choix des *normes* et *standards* utilisés est confirmée par ceux effectués par des systèmes équivalents en cours qui pointent vers les mêmes solutions (*ISO 19115*, *ISO 19139*, *WMS*, *Postgis*, *SVG* mais aussi *UML*), la collaboration entre les deux grands organismes que sont l'*ISO* et l'*OGC* semblent garantir une certaine longévité au système proposé. En particulier, au niveau européen, les orientations choisies respectent celles qui sont en cours de définitions par la directive *INSPIRE*. Cependant le *prototype* est perfectible du point de vue du *développement* et de l'*intégration* des *sources de référence* pour l'alimenter. Il serait à présent intéressant de déployer ce type d'application en réseau sur le Web, comme autant de *briques* constitutives d'un système *distribué* de *mutualisation* de l'*information* et de la *connaissance* sur la *zone côtière*. Concernant le développement, par la prise en compte de l'*approche objet* (en systématisant par exemple l'implémentation en *XML*) la qualité de l'application semble pouvoir être améliorée. Cependant nous pensons que l'emploi d'un *SGBDR* est, pour l'instant, incontournable pour répondre à certains traitements spécifiques essentiels pour la *GIZC*, notamment ceux liés à la *gestion* de l'*information géographique*. En outre, l'approche proposée devrait bénéficier des nombreux développements en cours pour la *gestion distante d'information géographique*. Concernant le cœur de notre application, le *référentiel sémantique*, les perspectives ouvertes sont larges. Nous avons proposé une *formalisation* de base indispensable pour proposer des formalisations plus complexes. En particulier, la réflexion peut être nourrie par la prise en compte d'approches complémentaires essentielles pour apporter des réponses pertinentes à des questions peu dégrossies mais d'importance dans la *GIZC*, en particulier la prise en compte de la *dimension temporelle*.

Nous proposons, pour finir, de développer chacun de ces points et de proposer des pistes pour l'amélioration ultérieure du système.

8.2.1 Perspectives liées à l'évolution du service de métadonnées

L'agrément sur la nécessité d'un *SMD* pour la localisation des *ressources informationnelles* mobilisées dans la *GIZC* semble consensuel de même que l'utilisation du Web pour en *diffuser* le contenu voire le *gérer*. En employant une norme comme l'*ISO 19115* et son implémentation normalisée par l'*ISO 19139*, l'ensemble des *SMD* liés au *domaine* de la *GIZC* pourront interopérer indépendamment de leur échelle géographique d'application. Ainsi des *SMD* pour des applications plus localisées ou plus larges géographiquement pourront servir à s'alimenter entre eux.

A certaines questions, par contre, seul l'usage apportera des réponses. On peut, par exemple, se demander quel sera en pratique, pour l'utilisateur, le mode de consultation privilégié de la *métadonnée* pour des *requêtes* simples : par le *SMD*, l'*ontologie* ou l'*interface cartographique* voire l'*annuaire électronique*. Plus simplement, à court terme, il s'agira de valider les procédures de *saisie* ou de *consultation* du *service de métadonnées* pour faciliter et ainsi encourager son utilisation.

Interopérabilité des SMD

Le *prototype* qui a été développé pour *Syscolag* répond à une problématique de *GIZC* à l'échelle régionale. Cependant, l'adaptation à cette échelle ne vient en pratique que de la nature de l'*information* et de la *connaissance* respectivement *stockées* dans le *SMD* et le *référentiel sémantique* (en particulier dans le *référentiel spatial*) qui, pour leur part, présentent un *intérêt générique* pour la *GIZC*.

Le système proposé peut donc *a priori* être implémenté quels que soient les *thématiques*, *espace* et l'*échelle*. Ainsi, il serait intéressant à présent de déployer ces *briques* de manière *distribuée* afin de profiter de leur possible *interopérabilité*. En effet, dans le contexte de la *GIZC*, la dimension *distribuée* apportera un bénéfice notable à cette approche puisqu'elle permettra de croiser les expériences (*intégrations* ou *vues* à différentes échelles par exemple) et de répartir dans l'ensemble des *briques* constitutives du réseau les apports provenant d'éléments séparés sur la base des descriptions fournis par les *SMD* (proche d'un *peer to peer*). En effet, des protocoles d'échanges normalisés (comme l'*ISO 239.50*) permettent des *requêtes distribuées* qui ouvrent des perspectives intéressantes de recherche simultanée sur plusieurs *SMD*. Mais les solutions envisageables sont nombreuses puisqu'il est également possible de dupliquer et de *centraliser* (en une base unique sur laquelle sont ciblées les requêtes) les *métadonnées* (voire les *ressources* elles-mêmes) *réparties* dans différentes bases secondaires [115].

Opérationnaliser le lien entre le service de métadonnées et le référentiel sémantique

Nous avons évoqué ce point dans les parties 5.4.2, 6.2 et 7.2. Il s'agit de pouvoir orienter le fonctionnement du service de métadonnées en utilisant la connaissance formalisée dans le référentiel sémantique. Cela peut prendre la forme d'*expansions de requêtes*, comme nous l'avons évoqué dans les parties 5.4.2 et 6.2.2. Ceci devra être effectué, à terme, en fonction des *ressources informationnelles* décrites dans le *service de métadonnées* et des *connaissances* intégrées dans le *référentiel sémantique*. Les combinaisons possibles sont variées et doivent prendre en compte la nature des différentes *relations ontologiques* (*BT*, *RT*, *RTS*..., voir détail en partie 5.1.4) ainsi que celle des *opérateurs logiques* traditionnels (*et*, *ou*, *sauf*...). Dans le cas des *descriptions spatiales*, la question se pose à l'identique avec la prise en compte supplémentaire des *relations topologiques*.

Automatisation. Le chien aboie et la caravane passe.

Les perspectives liées à l'*automatisation* des *métadonnées* sont dépendantes de la nature des ressources décrites. Or, dans la *GIZC*, nous avons vu que celle-ci est très variée (voir partie 2.1.3). On peut imaginer que la fourniture de *métadonnées* avec la *donnée* va se généraliser et ce d'autant plus facilement que la nature des ressources le permet, en particulier pour l'*information géographique* (*images satellitales*, *aériennes* ou *couches vecteur*...). Ceci paraît

d'autant plus plausible pour certaines *données brutes* commerciales dont la *métadonnée* est quasi intégralement automatisable mais déjà un peu moins pour les données résultants de *traitements* ultérieurs à l'*acquisition*, dont la charge de la *description* revient à leurs *producteurs*. Par contre, la question se pose de manière différente pour les données plus conventionnelles, qui, en dehors des champs des *références bibliographiques*, n'ont pas de description fine de leur contenu (type *description géographique, qualité, généalogie, contraintes d'accès...*).

Pour finir, il paraît difficile de croire que les solutions à venir vont résoudre le problème des lacunes de description du *patrimoine informationnel* détenu par les acteurs. Certaines ressources sont, en effet, complètement transparentes pour les systèmes, leur existence n'étant tout simplement pas décrite, parfois même ignorée. En attendant l'automatisation miracle qui ira dans les placards ou les disques durs, décrire les ressources dont l'existence est oubliée, le chien continuera sûrement d'aboyer... jusqu'à ce que la cabane lui tombe dessus.

Les développements

Dans le *prototype* qui a été développé, le mélange entre *programmations objet* et *impérative* (*PHP 4* avec *Java* pour *Prefuse* par exemple) pose des problèmes de développements complexes et sans réel intérêt, si ce n'est de résoudre certaines incompatibilités liés à ces différentes approches. Il serait donc intéressant, à terme, d'unifier les approches des *langages* employés et de simplifier l'outil opérationnel par une approche essentiellement *objet*.

Prendre en compte l'approche objet Le principal reproche qui peut être fait au développement de notre *prototype* concerne la nécessaire adaptation de spécifications utilisant l'*approche objet* pour l'implémentation en *PHP 4* alors que *PHP 5*, *Java* ou *XML* auraient respecté au mieux les recommandations *UML* de la norme *ISO 19115*. En outre, ce choix se justifie par la mise en place d'une *ontologie* qui s'appuie également sur une conception *objet* (voir partie 5.1.4).

Concernant le développement d'une application opérationnelle, à terme, compte tenu de la sortie récente de la version 5 de *PHP*¹, il serait, par exemple, souhaitable de porter les développements réalisés vers cette nouvelle version qui utilise une *approche objet* à l'image de celle adoptée, à l'aide du langage *UML*, dans l'ensemble des recommandations issues des standards que nous utilisons ou de nos analyses. D'autres applications illustrent l'intérêt de l'alternative de la *programmation orientée objet*, en témoigne le nombre d'outils développés en *Java* et *XML-XSL*.

Nous détaillons dans la partie suivante quelques avantages potentiels d'un développement basé sur *XML*.

Privilégier l'utilisation du XML Compte tenu de l'omniprésence d'*XML*, on peut se demander si l'emploi de ce langage et des syntaxes qui en dérivent ne constitue pas finalement la solution la plus efficace pour uniformiser les langages de développement utilisés et les traitements qui leur sont associés :

- format d'échange de l'*ISO 19115* (*XML* et *DTD* normalisée par la norme *ISO 19139 Geographic information - Metadata - Implementation specification* qui définit le modèle *UML* d'implémentation, selon le schéma décrit dans la norme *ISO 19115*, et des règles de mise en œuvre en *XML*),

¹PHP 5 est officiellement opérationnelle depuis Juillet 2004.

- formalisation des *connaissances* (à l'aide des syntaxes *XML* telles que *RDF*, *OWL*, *XTM*...),
- représentation des *objets géographiques* (à l'aide des syntaxes *XML* comme *GML*, *SVG*).

De plus, l'apport de *XML* pour les *Web services* est incontournable.

Par exemple, compte tenu du fait que les fichiers *XML* sont le format d'échange entre les *SMD* implémentant l'*ISO 19115*, il paraît en effet logique de penser que la meilleure solution pour visualiser l'information qu'ils contiennent sur le Web est l'emploi de fichiers de transformation *XSL*². De la même manière, il est possible de générer le *formulaire de saisie* adapté aux *gabarits* par ces techniques et de stocker l'information directement sous sa forme d'échange (le fichier *XML* du gabarit valué). Il n'est donc pas nécessaire de gérer le détail des valeurs associés aux différentes fiches de métadonnées dans une base de données.

Pour les éléments essentiels à la formulation de *requêtes* ou aux *contrôles* de la *valuation*, le problème se pose de manière différente. En effet, pour ces derniers, il est difficile actuellement de croire qu'on pourra se passer des bénéfices liés à l'utilisation de *SGBD* par l'emploi de *XML*.

SGBD Nous avons décrit dans la partie précédente comment il est possible de permettre la *saisie* et le *stockage* des *métadonnées* en s'appuyant uniquement sur des fichiers utilisant les langages *XML* et *XSL*. Cependant, entre la *saisie* d'une *fiche de métadonnées* et sa *consultation* détaillée au sein du *SMD*, il existe un nombre conséquent d'*GUI* associés à des traitements qui permettent, comme nous l'avons vu, la *formulation de la requête* et la *restitution des réponses* articulées autour des valeurs des éléments de métadonnées essentiels pour la *GIZC*.

Nous pensons donc qu'il est justifié pour les *composants* que nous avons développés, afin de *contrôler* ou d'*automatiser* la saisie de certains *éléments de métadonnées*, de continuer à s'appuyer sur un *SGBDR* pour les raisons suivantes :

- la *gestion* des valeurs préenregistrées pour les *éléments de métadonnées* essentiels à la *GIZC* est plus efficace dans un *SGBD*, en particulier pour les *descripteurs spatiaux* et la gestion des *relations topologiques*,
- la *résolution* des *requêtes* de *consultation* est plus efficace avec un *SGBD*,
- l'*interopérabilité* possible avec d'autres applications (*interfaces cartographiques*, *SIG clients libres*...),
- la familiarité des utilisateurs est plus grande avec ces systèmes de gestion qu'avec le *XML*.

Une alternative à un redéveloppement de l'application pourrait être de contribuer à l'amélioration d'un système existant qui répond à ces perspectives : par exemple *Geonetwork*, en conservant notre approche pour les *référentiels sémantique* et *spatial* associés. Cela permettrait de plus de bénéficier des apports de la communauté de développement associée à ces applications de référence et de s'intéresser aux développements des *composants* spécifiques à notre contexte d'implémentation.

Malgré tout, c'est par l'opérationnalisation de ce prototype que nous parviendrons des critiques nouvelles et pertinentes sous la forme de retours utilisateurs, par la confrontation de la pratique à la théorie. Les choix à long terme pourront alors être argumentés.

²XSL (eXtensible Stylesheet Language) est le langage de description de feuilles de style du *W3C* associé à *XML*. Une feuille de style *XSL* est un fichier qui décrit comment doivent être présentés (c'est-à-dire affichés, imprimés, épelés) les documents *XML* basés sur une même *DTD*. (*Document Type Definition*) ou un même schéma. La spécification est divisée en trois parties : *XSLT*, le langage de transformation, *XPath*, le langage de navigation dans un document *XML*, *XSL-FO*, le vocabulaire *XML* de mise en forme [179].

Synthèse des perspectives liées à l'évolution du Service de métadonnées

- Le système proposé peut servir de *brique* constitutive d'un *réseau distribuée* d'applications aux objectifs similaires de *gestion* de l'*information* pour la *GIZC*.
- L'*approche objet* peut permettre la simplification des développements proposés.
- En particulier *XML* semble être une réponse efficace pour unifier les développements et permettre la mise en place de *Web Services*.
- Néanmoins, nous ne remettons pas en question l'usage complémentaire d'un *SGBDR* à *cartouche spatiale* pour la résolution des traitements qui permettent une *saisie* et une *consultation* plus efficace du *SMD*, en particulier ceux qui s'appuient sur des *requêtes spatiales*.
- Une orientation possible pourrait être après les premiers retours des utilisateurs de redévelopper une application qui prennent en compte ces recommandations ou de contribuer à un logiciel existant, dont les orientations de développement correspondent à celles que nous mettons en perspectives.

8.2.2 Perspectives liées à l'utilisation du référentiel spatial

L'inventaire de l'*IG* en mode *vecteur* stockée dans le *référentiel spatial* (voir partie 6.3) est exploité soit par le *SMD* soit par l'*atlas cartographique* (voir parties 7.2.1 et 7.3.2). En outre, cette *IG* est mise à disposition des *clients* ou *serveurs WMS* (voir partie 6.2.4). Compte tenu de la richesse des développements en cours pour la *gestion* d'*IG*, nous pensons que les perspectives de *gestion* de l'*IG* de référence que nous avons collectée, sont prometteuses par le biais de *Web services* ou autres protocoles.

Les apports des standards en termes d'outils clients libres

L'utilisation de *Postgis* permet d'utiliser des outils d'exploitation de l'*information géographique* plus puissants que l'*interface cartographique* que nous avons développée pour répondre aux besoins particuliers du *SMD*.

Les serveurs Web standardisés pour l'*IG* plus riches en termes de fonctionnalités (tels que *Mapserver*, *Geoserver* ou *Geotools* qui implémentent les normes *WMS*, *WFS* voire *WCS*³ de l'*OGC*) améliorent les possibilités d'analyse spatiale en ligne et permettent la mise en place d'un *atlas* sur la *GIZC* en *LR* plus riche en fonctionnalités et dont les capacités d'*interopérabilité* avec d'autres *Web services* équivalents sont plus grandes.

Mais, là encore, il est utile de préciser que les *Web services* ne sont pas la seule manière de permettre une gestion distante de l'*IG* stockée dans une *base de données spatiale*.

Peu importe le flacon. . .

Le point important n'est pas tant le *protocole* utilisé pour permettre la *gestion* de l'*IG* que la *gestion* de l'*IG* elle-même. Peu importe le flacon. . .

Les *connexions distantes* sont en effet possibles en dehors de l'usage de *Web services* et pour les *acteurs* de la *GIZC*, la question de la réalisation technique ou du *protocole de connexion* utilisé n'a finalement que peu d'intérêt. En outre, grâce à *Postgis* nous pouvons également bénéficier d'outils variés et très performants tels que les *SIG clients* comme *Quantum GIS* (voir Figure 8.1), *UDIG*, *Jump* ou encore *grass* qui permettent une *connexion à distance* sur

³Spécification pour le catalogage de services.

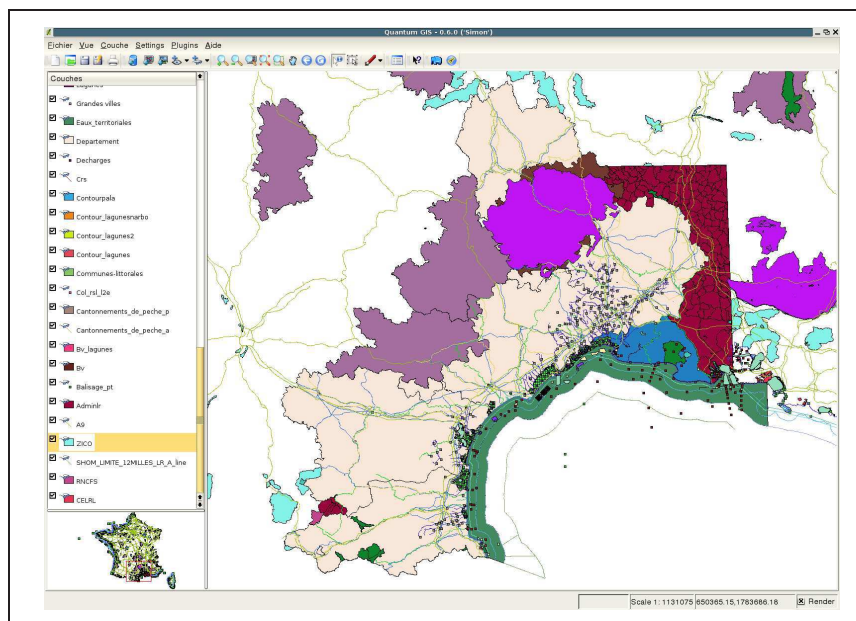


FIG. 8.1 – Exemple de connexion à une base *Postgis* avec le logiciel libre *Quantum GIS*

un *serveur de données spatiales* géré par *Postgis* ou *WMS*, soit directement sur une *BDS* contenue dans *Postgis*.

Il devient alors possible, depuis un *client SIG* sur un poste local, d'exécuter les *fonctions d'analyse spatiale* traditionnelles sur toutes les couches contenues dans les *serveurs de données spatiales* distants voire d'y *éditer* (*ajouter, modifier ou supprimer*) l'*information géographique* contenue (par l'emploi de la *spécification WFS* par exemple). Ceci constitue un vrai apport pour les disciplines impliquées dans la *GIZC* qui ne sont pas encore familières de l'utilisation des *SIG*, essentiellement en raison du *prix des licences de logiciels propriétaires et des données géographiques de référence*. De plus, la qualité de ces *clients SIG* égale celle des logiciels propriétaires pour les utilisations courantes.

En dehors des questions techniques, nous proposons, dans la partie suivante, de conclure par quelques perspectives sur les bénéfices potentiels de cette démarche.

Les bénéfices de la démarche

Il convient de relever ici le fait que la constitution d'un *référentiel spatial* consultable en ligne implique le *partage physique* de l'*IG* en mode *vecteur* en une même *base de données spatiale*. C'est un premier pas important vers le *partage physique* des ressources en plus de leur *localisation*. Par le biais de *Postgis* associé à *WMS* et *WFS* ou à des *clients SIG*, il devient possible d'atteindre un niveau supplémentaire d'*interopérabilité* entre les systèmes puisque les ressources qui y sont contenues deviennent *physiquement* manipulables par ces *Web Services*. Le fait de les dupliquer et des les centraliser dans *Postgis* pour constituer un *référentiel spatial* simplifie la mise en place de cette *interopérabilité* mais on pourrait tout aussi bien proposer aux partenaires d'utiliser une *base de données spatiale* similaire dans chaque organisme en les laissant libres de gérer leurs ressources comme ils le souhaitent.

Le résultat de l'adhésion des utilisateurs à cette démarche sera donc enrichissant pour évaluer la

faisabilité du *partage physique* des ressources en plus de leur simple *localisation*. Cela permettra d'argumenter la faisabilité et l'intérêt d'une telle approche lorsque la question se posera dans des cas similaires.

On peut imaginer que la mise à disposition de l'*information géographique* contenue dans le *référentiel spatial* suscitera des commentaires profitables à la *quantité* et à la *qualité* de l'*IG* collectée et mise à disposition sur des *clients SIG* auprès de nos partenaires :

- un espace de discussion sur les lacunes dans l'inventaire des couches existantes et sur les couches manquantes,
- la validation d'un système de projection commun (pour le moment en *Lambert II étendu*) indispensable pour la cartographie du *continuum terre-mer*,
- un serveur de données spatiales interopérable et décrit par l'*ISO 19115 : géocatalogue*,
- la possibilité d'analyses spatiales sur les ressources et l'aide à la décision par un véritable *SIG* (gestion des *conflits d'usage* par exemple),

Ainsi la *base de données spatiale* a une importance spécifique par rapport au *SMD*, puisqu'elle permet la diffusion *physique* de l'information existante et mutualisée, dans les limites des droits d'accès, grâce à des outils qui restent dans un *environnement de développement libre*. Ceci est un mode de diffusion de l'information qui va dans le sens des recommandations du *DD* et de la *GIZC*.

Synthèse de la Section 8.2.2

- L'*IG* collectée dans le *référentiel spatial* peut être mise à disposition grâce aux spécifications de l'*OGC* pour les *Web Services*.
- Il existe d'autres possibilités que les *Web services* pour la *gestion* de l'*IG* contenue dans *Postgis*.
- Quelle que soit la solution retenue, cette démarche constitue un premier pas concret vers le *partage physique* de ressources informationnelles et une interopérabilité des bases qui les contiennent.
- Par le traitement physique de l'*IG*, le *référentiel spatial* devrait révéler des *lacunes qualitatives* ou *quantitatives* dans les ressources existantes et partagées.

8.2.3 Perspectives liées à l'utilisation de l'ontologie

Parmi les outils proposés, les perspectives ouvertes par l'utilisation et l'évolution du *référentiel sémantique* et du *modèle d'ontologie* seront sûrement les plus riches, compte tenu de l'importance de ces champs de recherche et de la demande des utilisateurs du domaine. Un effort d'homogénéisation des sources employées pour alimenter le référentiel sera nécessaire pour assurer la lisibilité de la solution apportée en dehors de notre contexte d'implémentation, comme nous l'avons fait pour le service de métadonnées (à l'aide d'un *jeu d'éléments normalisés*). En outre, le modèle proposé peut être amélioré par l'intégration de nouveaux *concepts* ainsi que de nouvelles *propriétés*. En particulier, l'intégration des *dimensions temporelle et verticale* dans le *référentiel sémantique* devrait permettre la formalisation de connaissances plus complexes, tels que des *scénarios* de *GIZC*.

Interopérabilité syntaxique pour le partage de la connaissance

Nous avons présenté en parties 5.2.1 et 5.2.3 comment le *modèle d'ontologie* que nous proposons est compatible avec la formalisation des connaissances employée par les langages

du *Web sémantique*. Cependant, l'implémentation actuelle de ce modèle est effectuée dans un *SGBDR*, compte tenu des besoins spécifiques à notre *service de métadonnées*. Toutefois, nous avons présenté un *prototype* qui exploite le contenu du *référentiel sémantique* sous la forme d'un *réseau sémantique* (voir partie 7.3.1). Compte tenu des passerelles possibles par l'utilisation d'*UML* et la conception même du *modèle d'ontologie* proposé, il pourrait être intéressant de valoriser le travail effectué en implémentant ce modèle par le biais d'un langage dédié à l'élaboration d'*ontologie*, par exemple, *XTM* ou *OWL* (ce dernier est utilisé par la *FAO* dans l'*ontologie FOS*, voir partie 5.1.4). L'*interopérabilité syntaxique* permettrait alors d'envisager des bénéfices plus concrets en terme d'*interopérabilité sémantique* en partageant les *concepts* et les *relations* de *domaines* impliqués dans la *GIZC*.

Synergie avec le contenu des sources de référence

Nous présentons dans cette partie, l'intérêt de la réutilisation des *sources de référence* existantes (notamment les *glossaires*, *thésaurus* et *ontologies* des domaines impliqués dans la *GIZC*, voir section 5.1) et d'une collaboration du *programme Syscolag* avec les organismes en charge de leur gestion.

Démarche générale Afin de proposer une *base de connaissance* aux perspectives d'utilisation évoluées, il est nécessaire d'attribuer aux *classes* des *propriétés* et de les *instancier* (ie *valuer* les *propriétés*) puis de diffuser l'information / les connaissances produites sous une forme qui respecte le profil des utilisateurs visés (voir détail en partie 5.1.4).

Pour cela il peut être souhaitable de considérer à nouveau l'apport des *sources de référence* afin de réutiliser la part de *connaissance* qui y est formalisée mais aussi, en retour, d'alimenter la part de *connaissance* manquante en profitant du temps épargné par l'intégration de celle déjà existante. Certaines parties de l'*ontologie* proposé pourraient ainsi jouer un rôle de *briques* basées sur les systèmes internationaux et être exploitables à des échelles locales.

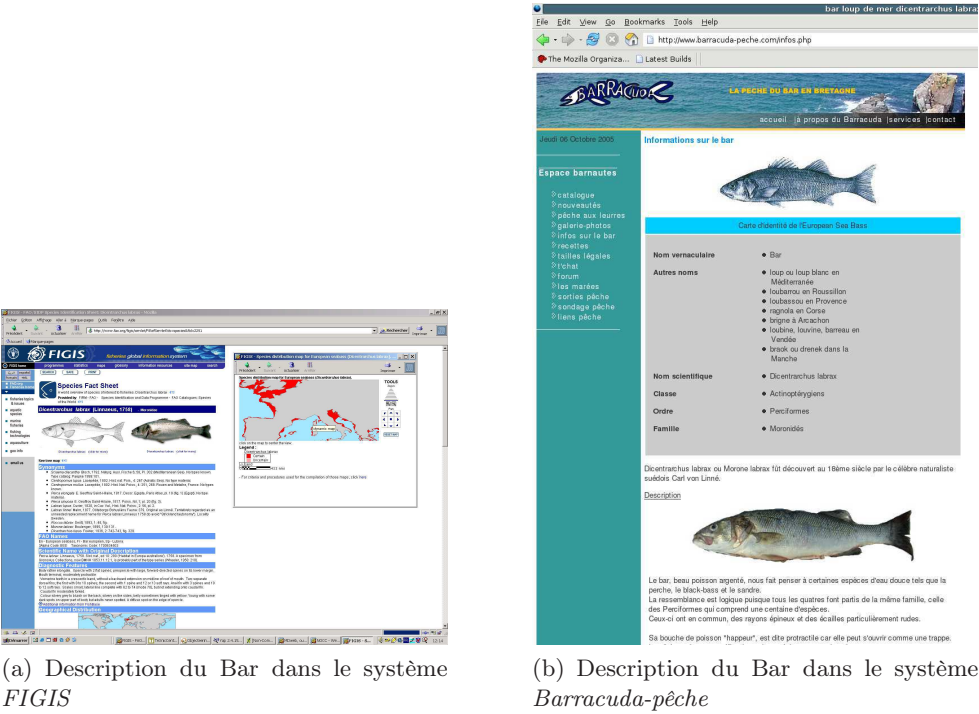
Par exemple, pour amorcer la démarche, une proposition pourrait être, dans un premier temps, de dresser une liste d'*outils terminologiques de référence* pour le domaine de la *GIZC* à utiliser pour définir les *propriétés sémantiques* de chaque *concept* employé librement par un utilisateur pour indexer une *ressource informationnelle* dans le *service de métadonnées*.

En complément, les *concepts* collectés pourraient servir de support pour instancier de nouvelles *relations ontologiques* vers d'autres *concepts*.

Nous proposons dans la partie suivante un exemple de sources possibles pour le cas spécifiques des *concepts* liés à la *Faune*.

Exemples de la Faune Pour illustrer les recommandations précédentes, la Figure 8.2(a) illustre, dans le système *FIGIS* (voir partie 2.3.2), une *interface graphique* dédiée au détail des *propriétés* associées aux *espèces* de poissons répertoriées dans ce système. Les propriétés associées aux espèces pourraient tout à fait être reprises telles quelles comme *propriétés* de la *classe poisson* (voire généralisées à toute *espèce animale* ou *végétale*) dans notre *ontologie*.

Le bénéfice de cette approche est d'absorber une *connaissance* de référence fournie par *FIGIS* : (les *concepts* et leurs *propriétés*). Cela s'inscrit directement dans la tendance actuelle de réutilisation d'*ontologies* de référence. L'*intégration* dans notre système permet un gain de temps et de qualité pour l'élaboration de notre *ontologie*. Si les systèmes de référence ne gèrent pas toutes les *espèces* nécessaires à notre espace géographique d'application, alors la description de nouvelles espèces peut en retour bénéficier aux systèmes de référence tels que



Poisson Figs
Synonymes : string
Nom FAO : string
Nom scientifique et description originale : string
Éléments de détermination : string
Distribution géographique : string
Habitat et Biologie : string
Taille : string
Pêcheries, exploitation : string
Noms locaux : string
Bibliographie : string

(c) Attribution de propriétés à la classe *Poisson*, d'après le système FIGIS

Poisson
Description : string
Habitat : string
Reproduction : string
Nourriture : string
Nom vernaculaire : string
Nom scientifique : string
Classe : string
Ordre : string
Famille : string

(d) Attribution de propriétés à la classe *Poisson*, d'après un site de pêche de loisir

FIG. 8.2 – Sites de diffusion de la connaissance sur le Web vers différents acteurs : scientifiques 8.2(a) ou pêche de loisirs 8.2(b) et formalisation de ces connaissances en vue d'une d'intégration dans notre ontologie 8.2(c) et 8.2(d)

FIGIS. L'harmonisation du contenu du *référentiel sémantique* proposé avec celui des *sources de référence* devient ainsi possible au bénéfice de tous. De la Figure 8.2(a)⁴ on peut extraire des *propriétés* pour la *classe* générale *faune* en vue d'une intégration de ces propriétés dans l'*ontologie* que nous proposons (voir Figure 8.2(c)).

⁴<http://www.fao.org/figis/servlet/FiRefServlet?ds=species&fid=2291>.

De la même manière, la Figure 8.2(b) correspond à un site dédié à la *pêche de loisir*⁵. Il est intéressant de relever la finesse d'information apportée par un tel site, qui en plus de la *polysémie*, rend compte des différents *propriétés* de chacune des *espèces* inventoriées (voir Figure 8.2(d)).

Ces deux exemples illustrent bien l'intérêt commun qu'il peut y avoir à partager des *connaissances* fines pour des acteurs variés : *scientifiques* ou simples *citoyens*. En effet, on peut remarquer la convergence sur la nature des *propriétés* attribuées aux *espèces*. En outre, compte tenu de la valeur générale de certaines propriétés, on pourrait aussi bien les attribuer à la classe *Faune/Flore* de notre ontologie (voir Figure 8.3) (elles seraient alors *héritées* dans la classe *espèce* et l'ensemble de ses *objets* (en particulier *poisson*).

Nous proposons donc la prise en compte de telles sources et la participation aux systèmes de références existants pour assister l'attribution de *propriétés* pertinentes aux différentes *classes* de notre *ontologie*. Ainsi, nous pensons pouvoir également aider en retour à la description de ces sites.

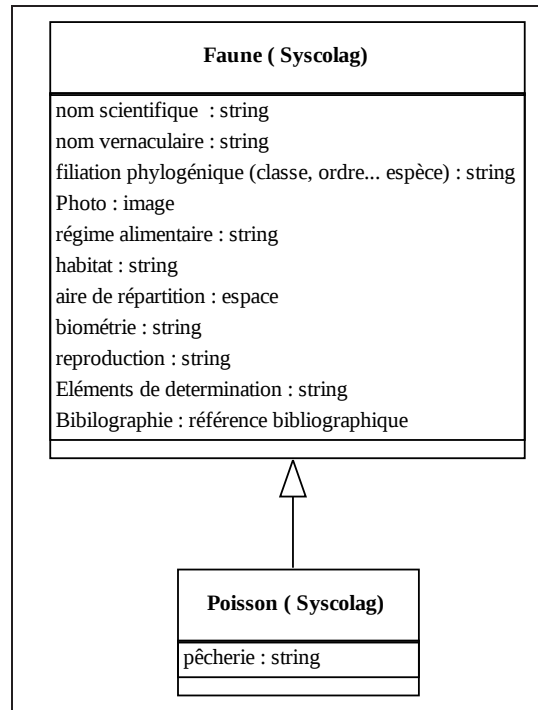


FIG. 8.3 – Exemple d'utilisation de sources de référence du domaine (voir Figure 8.2) pour l'attribution de propriétés aux classes de notre ontologie

A présent, nous allons proposer quelques perspectives de recherche pour améliorer la description des *concepts* par la prise en compte de *propriétés* d'importance capitale dans la *GIZC*.

Intégration de la dimension verticale

Bien que non gérée actuellement par notre *référentiel spatial*, nous soulignons l'intérêt de la *description de l'étendue verticale* en zone côtière. La gestion de cette dimension est possible

⁵<http://www.barracuda-peche.com/infos.php>.

dans la norme *ISO 19115* grâce à la classe *EX_VerticalExtent* du *Data type EX_Extent* où sont gérés l'ensemble des *descripteurs spatio-temporels*.

Pour le *milieu marin*, la gestion de la dimension verticale est essentielle. En outre, si par l'implémentation de l'*ISO 19115*, il est possible de gérer la description verticale à l'aide de descripteurs numériques (de manière tout à fait semblable à ce qui est fait pour les dates), il serait intéressant, de pouvoir décrire en complément la dimension verticale en mode textuel. Par exemple, les termes *benthique*, *pélagique*, *intertidal*...expriment bien une portion de l'espace vertical. C'est également le cas de termes dans d'autres disciplines telles que la *géologie* : *aquifères*, *bassin profond*, *forage*...sont également liés à une dimension verticale.

On peut imaginer, à terme, que la maîtrise de cette troisième dimension dans les *SIG*, avec la généralisation de l'usage des *MNT* et de la *modélisation 3D* (les *SIG* outils gérant la topologie 3D commencent à apparaître), permettra de proposer à l'utilisateur des descripteurs d'objets géographiques qui intègrent directement la dimension verticale par une *géométrie précise* ou *approximée* (limite *inférieure* et *supérieure*), voir illustration en Figure 8.4.

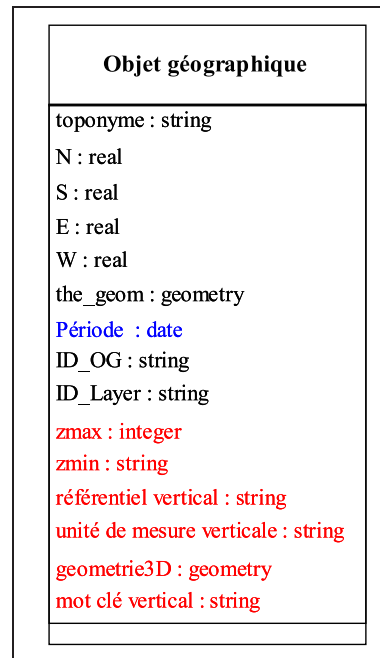


FIG. 8.4 – Intégration de la dimension verticale et temporelle sous la forme d'attributs de la classe *Objet géographique* de l'*ISO 19115*

La *dimension verticale* pourrait ainsi affiner le *référentiel spatial* proposé (voir Figure 8.5).

Intégration de la dimension temporelle

Nous n'avons pas apporté de réponse spécifique à l'*intégration de la dimension temporelle* en dehors des descriptions proposées par la norme *ISO 19115*. Or, comme nous l'avons fait pour les *concepts thématiques* et *spatiaux*, il est possible de procéder à un inventaire des *concepts temporels* qui permettra différents niveaux de lecture par l'emploi d'échelles *géologiques* (*jurassique*, *crétacé*...), *historiques* (*révolution*, *périodes*, *règne*...), *saisonnnières* (*été*, *hiver*, *solstices*...)...

Il est également possible d'introduire des liens entre les *concepts* de notre *référentiel sémantique* tels que ceux décrits par les *relations d'Allen* (voir Figure 8.5). Ces *relations* permettent d'exprimer par exemples des liens entre *concepts* du type : *avant*, *pendant*, *recouvrement*, *départ*, *fin*...

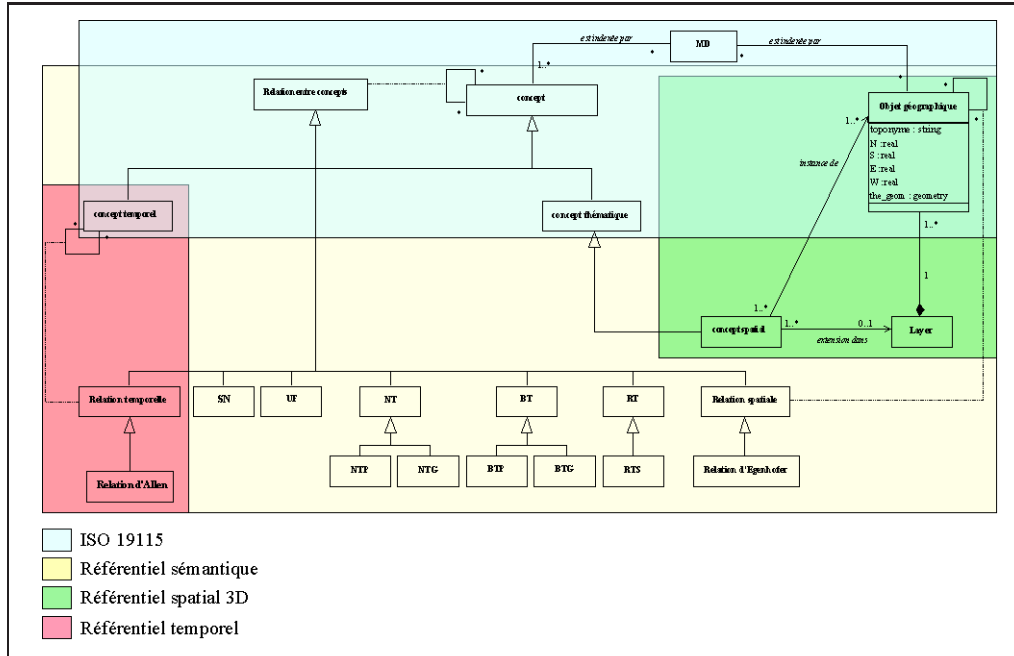


FIG. 8.5 – Position relatives des référentiels sémantique, spatial, temporel et de l'ISO 19115

Grâce à cela, il est envisageable de restituer plus fidèlement la *dynamique* de la *zone côtière* à travers la prise en compte de la *chronologie* des *interactions*. Par exemple, il est possible pour un *scénario* défini d'ordonner chronologiquement les *interactions* et de faire apparaître le lien de causes à effets entre les processus observés. Compte tenu de l'importance de l'approche *état-pression-réponse* dans la GIZC, la prise en compte des aspects précédents paraît donc complémentaire et essentielle pour permettre une expression plus fine de la connaissance du domaine.

L'intégration d'un *référentiel temporel* au sein du *référentiel sémantique* (voir Figure 8.5) permettra d'affiner la qualité des descriptions proposées dans le *service de métadonnées* ou dans la *base de référence terminologique* et, à terme la description de connaissances plus complexes.

Description de connaissances plus complexes

Dans le référentiel sémantique proposé sont possibles :

- Une description fine des *éléments statiques* du *système* (*acteurs*, *ressources* et *outils*) de leurs *relations* et de leurs *propriétés*,
- Une description fine des *interactions* entre ces *éléments* et donc de la *dynamique* du *système*,

Globalement, le *modèle d'ontologie* proposé est pertinent pour exprimer des *connaissances élémentaires* (ou *briques*) sous forme de triplets (*Sujet*, *Verbe*, *Complément*) où *Sujet* et *Com-*

plément peuvent prendre pour valeurs des *concepts* qui spécialisent *Acteur*, *Ressource* ou *Interaction* et *Verbe* prend pour valeur uniquement des *concepts* qui spécialise *Interaction* (voir Figure 8.6). Nous résumons ces propriétés du modèle par la formulation suivante d'une connaissance :

$$\text{Connaissance élémentaire} = \text{Interaction}_i = (A|R|I, I, A|R|I)$$

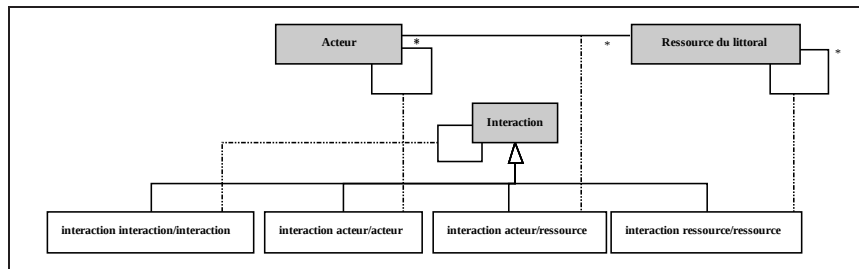


FIG. 8.6 – Modèle d'ontologie proposé à l'issue de ce travail

Cette formalisation est un point essentiel pour ce modèle et doit permettre, à terme, une gestion plus fine de la connaissance. En effet, le modèle respecte donc à la fois les contraintes des langages du *Web sémantique* et l'*approche écosystémique*. Cette simplicité en fait un modèle robuste qui peut, a priori bénéficier de formalisations plus complexes.

Grâce à cela et aux perspectives ouvertes par l'attribution de nouvelles *propriétés* aux *concepts*, nous pensons que le travail réalisé est une base qui permet le passage à la réalisation d'*ontologies formelles*⁶.

Par exemple, à la suite de l'intégration de la *dimension temporelle*, on pourrait proposer une formalisation plus fine de *scénarios* de GIZC (par exemple, le scénario qui rend compte de l'"eutrophisation de la lagune de Thau" ou encore "les conflits d'usage liés aux activités récréatives sur l'étang de Leucate") de la manière suivante :

$$\text{Scénario} = \sum_{i=1}^n \text{Interaction}_i$$

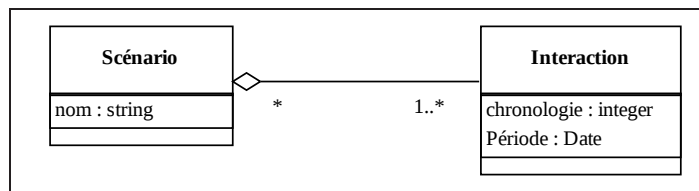


FIG. 8.7 – Formalisation UML possible pour un scénario

La prise en compte de la *dimension temporelle* dans la classe *Interaction* est nécessaire pour restituer la dynamique générale d'un *scénario* et les relations de cause à effet (voir Figure 8.7.

⁶Dans les ontologies formelles, on utilise une représentation dans un formalisme logique. Les catégories sont désignées par des *axiomes* ou des définitions exprimées en un langage logique. Elles sont faites pour effectuer des inférences.

Les premières tentatives de formalisation de connaissances plus complexes (voir partie 5.4.1) ont confirmé les limites du système proposé actuellement à ces fins.

L'introduction de règles en complément des relations portées par le modèle proposé, devrait permettre d'introduire la notion d'*inférence*.

Enfin pour explorer une nouvelle voie en complément de nos analyses et des classifications effectuées avec les *acteurs* du projet, il serait possible d'utiliser des *outils de classification automatique* pour valider les *modèles*, détecter les *incohérences* ainsi que les *conflits potentiels* [19, 67].

A terme, la place du *référentiel sémantique*, de l'*ontologie* et du *service de métadonnées* devraient donc être centrale dans le système d'information visé par *Syscolag* (observatoire, tableau de bord, indicateurs, base de connaissance..., voir détail en partie 2.2.2). On peut supposer que l'*ontologie* sera, à long terme, une base de raisonnement pour les différents composants du système en même temps qu'un aiguilleur pour la navigation de l'utilisateur. En fait, les applications potentielles d'un tel référentiel sont variés. Par exemple, d'après la littérature actuelle, l'émergence des SIG qui utilisent l'approche objet devrait être importante dans les années à venir et on peut alors imaginer que les formalisations réalisées dans le référentiel sémantique pourront nourrir ce type de système.

Pour finir nous faisons référence à la terminologie employée par [70, 71, 55] : lorsqu'une *ontologie* guide un *système d'information* ou un *SIG* alors on peut parler de *SIG conduits par des ontologies* (*Ontology-Driven Geographic Information Systems*) et plus généralement de *Ontology-Driven Information Systems* lorsqu'une *ontologie* joue un rôle central dans la définition de tous les *composants* et aspects d'un *système d'information*.

Synthèse de la Section 8.2.3

- Le mode d'alimentation du *référentiel sémantique* doit prendre en compte les *sources de référence* (pour définir la nature des *concepts*, des *relations*, des *propriétés*) et partager leurs modes de représentations de la connaissance.
- L'intégration des *dimensions temporelle et verticale* est indispensable pour améliorer la qualité du *référentiel sémantique*.
- Un travail de formalisation complémentaire est nécessaire pour élaborer des *ontologies formelles*. Le modèle proposé semble pouvoir supporter cette évolution.
- Par ce biais, des connaissances plus complexes pourront être formalisées, telles que des *scénarios* de *GIZC*.
- L'*ontologie* doit donc prendre une part importante, à terme, pour la réalisation des objectifs de *gestion de l'information et de la connaissance* dans le système d'information visé par *Syscolag*, en orientant le fonctionnement des différents *composants* ainsi que la navigation des utilisateurs.

Références

- [1] ISO 13250 :2004. *Geographic Information - Web map server interface, ISO 19128*. International Organization for Standardization (ISO), Genève, Suisse, 2004.
- [2] ISO 19115 :2003. *Geographic Information Metadata, ISO 19115*. International Organization for Standardization (ISO), Genève, Suisse, 2003.
- [3] ISO 19128 :2004. *Geographic Information - Web map server interface, ISO 19128*. International Organization for Standardization (ISO), Genève, Suisse, 2004.
- [4] ISO 2788 :1986. *Guidelines for the establishment and development of monolingual thesauri, ISO 2788*. International Organization for Standardization (ISO), Genève, Suisse, 1986.
- [5] ISO 5964 :1985. *Guidelines for the establishment and development of multilingual thesauri, ISO 5964*. International Organization for Standardization (ISO), Genève, Suisse, 1985.
- [6] Agora21. Information et gouvernance. <http://www.agora21.org/information.html>, 2005.
- [7] J. Aitchison, A. Gilchrist, and D. Bawden. *Thesaurus construction and use : a practical manual* (4th ed), 2000.
- [8] Serge Allain, Brigitte Guillaumont, Christophe Le Visage, Lionel Loubersac, and Jacques Populus. Données géographiques de référence en domaine littoral marin. SHOM/IFREMER, Version du 26/04/2000, 2000.
- [9] Serge Allain, Brigitte Guillaumont, Christophe Le Visage, Lionel Loubersac, and Jacques Populus. Données géographiques de référence en domaine littoral. volume 25, pages 67–79, Brest, septembre 1999.
- [10] Youcef Amerouali. *Metadonnées basées sur l’association d’éléments de description de ressources et d’éléments de profil d’utilisateur*. PhD thesis, Université Claude Bernard - Lyon I, Mai 2001.
- [11] Adrian Annen. Geostandards : Opeengeospatial consortium ogc (gml, wfs, wms). In *Journées d’étude Interopérabilité pour l’utilisation généralisée de la Géoinformation*, Standards nationaux et internationaux, Zürich, Suisse, Mars 2005. Swiss Federal Institute of technology.
- [12] Catherine Houstis ans Spyros Lalis and Hariklia Tsalapata. An environmental information metadata architecture for the thetis distributed coastal zone management system, 1999.
- [13] Julien Barde, Jacques Divol, Thérèse Libourel, and Pierre Maurel. Interface adaptable de requêtes pour un service de métadonnées. In *Cinquièmes journées Extraction et Gestion des Connaissances (EGC 2005)*, Méta-données et Adaptabilité pour les Systèmes d’Information sur le Web, Paris, France, Janvier 2005.

- [14] Julien Barde, Jacques Divol, Thérèse Libourel, and Pierre Maurel. Système d'information pour la gestion intégrée des zones côtières. In *SAGEO'2005 International Conference on Spatial Analysis and GEomatics*, Risques, environnement, Avignon, France, Juin 2005. Université d'Avignon (UMR 6012 ESPACE, UMR 5506 LIRMM et UMR 7011 Image et Ville).
- [15] Julien Barde, Thérèse Libourel, and Pierre Maurel. Ontologies et métadonnées pour le partage d'information géographique. *Revue Internationale de Géomatique*, 14(2) :199–217, 2004.
- [16] Julien Barde, Thérèse Libourel, and Pierre Maurel. A metadata service for integrated management of knowledges related to coastal areas. *Multimedia Tools and Applications*, 25(3) :419–429, 2005.
- [17] Julien Barde, Thérèse Libourel, Pierre Maurel, Jean-Christophe Desconnets, Nabila Mazouni, and Lionel Loubersac. A metadata service for managing spatial resources of coastal areas. In *Coastgis 2005*, Data Models, Databases and Metadata, Aberdeen, Scotland, Juillet 2005. University of Aberdeen.
- [18] Julien Barde and Sébastien Roussel. Définition et caractérisation des composantes de la zone côtière : proposition de typologies en région languedoc-roussillon. Le rôle des chercheurs comme acteurs des démarches de gestion intégrée, Paris, France, Mars 2005. Ministère de l'écologie et du développement durable.
- [19] S. Bechhofer, I. Horrocks, C. Goble, and R. Stevens. Oiled : A reasonable ontology editor for the semantic web, 2001.
- [20] GFI Benelux. Guide utilisateur de l'application metawal de la région wallonne. *GFI Benelux*, page 76, January 2004.
- [21] M. Bergeron. Vocabulaire de la géomatique. office de la langue française. 1993.
- [22] Tim Berners-Lee, James Hendler, and Ora Lassila. The semantic web : A new form of web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American*, may 2001.
- [23] Nathalie Bonneric. Intégration des objets géographiques de référence dans la base de données spatiales du programme syscolag, Septembre 2005.
- [24] G. Booch, J. Rumbaugh, and I. Jacobson. The unified modeling language user guide, 1999.
- [25] Gro Harlem Brundtland. *Notre avenir à tous*. Universe Boks, New York, 1989.
- [26] Arion System Catherine. A service infrastructure for e-science : the case of the arion system.
- [27] CCSTI. In *Zones côtières : Quels outils pour quelle gestion ?*, Lorient, 2002.
- [28] CESRB. *Pour une gestion concertée du littoral en Bretagne*. Conseil Economique et Social du Conseil Régional de Bretagne, Section Mer Littoral, 2004.
- [29] L. Choppin. Systèmes d'information géographique et données de référence dans les zones côtières européennes. Technical Report BRGM/RP-51250-FR, BRGM, 2001.
- [30] Biliانا Cicin-Sain and Robert W. Knecht. *Integrated Coastal and Ocean Management : Concepts and Practices ; A Practical Guide to ICM Programs : Summary, Examples of Good Practices in ICM Programs, Prospects and Future Challenges*. ISLAND PRESS, Washington, D.C. & Covelo, Californie (1998), 1998.

- [31] Sébastien Clais. Etude comparative des sgbd spatiaux, Juin 2004.
- [32] Sébastien Clais. Interface cartographique intégrée dans un service de métadonnées, Septembre 2005.
- [33] CoastLearn. Eucc. <http://www.netcoast.nl/coastlearn/website/>, 2005.
- [34] Anita Conti. *Géants des mers chaudes*. 1997.
- [35] Anita Conti. *L'océan, les bêtes et l'homme ou l'ivresse du risque*. éditions Payot & Rivages, 1999, Paris, 1999.
- [36] Anita Conti. *Racleurs d'océans*. Petite bibliothèque Payot / Voyageurs, éditions Payot & Rivages, 1998, Paris, 1999.
- [37] COSIG/KOSIG. Elaboration d'un modèle de métadonnées pour la suisse compatible avec la norme iso/dis 19115. *Projet de document pour avis, Version 1.2 f*, page 52, Décembre 2001.
- [38] COSIG/KOSIG. geocat.ch, catalogue de métadonnées pour les données géographiques. page 6, Août 2002.
- [39] Olivier Courtin. Sig orienté web et standards ouverts. implémentation de l'iso 19128 avec php/postgis et svg/ecmascript, Septembre 2004.
- [40] M. Craglia and K. Evmorfopoulou. Madame : Methods for access to data and metadata in europe. page 46, 2000.
- [41] DCMI. Dublin core metadata initiative. <http://dublincore.org/>, 2005.
- [42] Jacques Denis and Laurence Miossec. Développement d'une méthodologie de qualification de l'espace côtier, application à la façade méditerranéenne. *Rapport interne IFREMER RST, Direction de l'Environnement et de l'aménagement littoral*, page 58, Décembre 1996.
- [43] Jean Christophe Desconnets, Thérèse Libourel, Pierre Maurel, André Miralles, and Michel Passouant. Proposition de structuration des métadonnées en géosciences. *Journées CASSINI'01, Montpellier, France*, Géomatique et Espace Rural :69–82, 2001.
- [44] Jean Christophe Desconnets, Nicolas Moyroud, and Thérèse Libourel. Méthodologie de mise en place d'observatoires virtuels via les métadonnées. Nancy, France, Juin 2003.
- [45] José Devers. Gt littoral, rapport final du cnig. Technical report, Octobre 2002.
- [46] Jacques Divol and Thérèse Libourel. Outils d'aide à la localisation dans le cadre du projet syscolag. Technical report, Mai 2005.
- [47] J.P. Doody. Information required for integrated coastal zone management : Conclusions from the european demonstration programme. *Coastal Management*, (3) :163 173, 2003.
- [48] Jacky Drouet. *L'information géographique en zone côtière : analyse prospective des besoins*. Dea, université de Nantes BRGM, 2002.
- [49] Conservatoire du Littoral. <http://www.conservatoire-du-littoral.fr/>.
- [50] Pascal Dumas. *Caractérisation des littoraux insulaires : approche géographique par télédétection et SIG pour une gestion intégrée. Application en Nouvelle Calédonie*. PhD thesis, Université d'Orléans, 2004.
- [51] Dominique Dupilet. Le règlement des conflits d'usage dans la zone côtière entre pêche professionnelle et autres activités. Technical report, 2001.

- [52] conditions et objectifs Développement durable : définition. <http://www.menv.gouv.qc.ca/developpement/definition.htm#note1>.
- [53] Edward Dwyer, Liz O'Dea, and Valerie Cummins. The marine irish digital atlas : a web portal to coastal and marine data in ireland. page 7, Genova, Italy, July 2003.
- [54] Max J. Egenhofer. Toward the semantic geospatial web, 2002.
- [55] Max J. Egenhofer and Frederico T. Fonseca. Ontology-driven geographic information systems. In . C. Bauzer Medeiros (ed.), editor, *7th ACM Symposium on Advances in Geographic Information Systems*, Kansas City, MO, November 1999.
- [56] Marie-Paule Jacques et Anne-Marie Soubeille. Partage des termes, partage des connaissances ? construire une modélisation unique de plusieurs corpus. 2000.
- [57] Nicolas Chung et Marie-Noëlle Cormenier. Normes et standards de métadonnées. *EDUCNET > Technique, normes et standards > Séminaire Lyon 10/2003*, Septembre 2003.
- [58] Pascal Buléon et Mauro Gaio. Analyse des territoires - des atlas électroniques pour comprendre les territoires. *Technologies Internationales*, 90, Décembre - Janvier 2002-2003.
- [59] ETC-CDS. Gemet : General multilingual environmental thesaurus. <http://www.eionet.eu.int/GEMET>, 2005.
- [60] Massimo Craglia Kalli Evmorfopoulou. Elargir l'accès aux informations du secteur public : Guide des bonnes pratiques. Décembre 2000.
- [61] Sébastien Faisant. Catalogage des données géographiques numériques d'un service sig par les métadonnées. *Mémoire d'ingénieur ESGT 2003*, Septembre 2003.
- [62] Kateryna Falkovych, Marta Sabou, and Heiner Stuckenschmidt. Uml for the semantic web : Transformation-based approaches. In *Knowledge Transformation for the Semantic Web*, pages 92–106. 2003.
- [63] FAO. Terminologie de la fao. <http://www.fao.org/faoterm/index.asp?lang=fr>.
- [64] FAO. Thésaurus agrovoc. <http://www.fao.org/agrovoc/>, Mars 2004.
- [65] FAO. The agricultural ontology server (aos), 2005.
- [66] FAO. Asfa : Aquatic sciences and fisheries abstracts thesaurus. <http://www4.fao.org/asfa/asfa.htm>, 2005.
- [67] A. Farquhar, R. Fikes, and J. Rice. The ontolingua server : a tool for collaborative ontology construction, september 1996.
- [68] FGDC. Shoreline metadata profile of the content standards for digital geospatial metadata. Technical report, Reston, Virginia, Juin 2001.
- [69] FGDC. Federal geographic data committee. <http://www.fgdc.gov/>, 2005.
- [70] Frederico Fonseca, Max J. Egenhofer, Peggy Agouris, and Gilberto Câmara. Using ontologies for integrated geographic information systems. *Transactions in GIS*, 6(3), 2002.
- [71] Frederico T. Fonseca, Max Egenhofer, Clodoveu Davis, and Gilberto Câmara. Semantic granularity in ontology-driven geographic information systems. *AMAI Annals of Mathematics and Artificial Intelligence*, Special Issue on Spatial and Temporal Granularity(36.1-2) :121–151, 2002.
- [72] Andrew U. Frank. Ontology : a consumer's point of view. *On-line paper*, 1997.

- [73] Andrew U. Frank. Consistency constraints in geographic information systems - an ontological view. *IJGIS - Special Issue*, (3465), 2001.
- [74] Aldo Gangemi, Frehiwot Fisseha, Johannes Keizer, Boris Lauser, Jos Lehmann, Anita Liang, Ian Pettman, Margherita Sini, and Marc Taconet. An overview of the fos project : Towards a fishery semantic web, 2004 ?
- [75] Aldo Gangemi, Frehiwot Fisseha, Ian Pettman, Domenico M. Pisanelli, Marc Taconet, and Johannes Keizer. A formal ontological framework for semantic interoperability in the fishery domain, 2002.
- [76] Marc Gilgen. Normes et standards pour données géographiques, Décembre 2000.
- [77] Marc Gilgen. Interopérabilité en matière de géodonnées - État actuel et perspectives dans le canton de vaud. In *Journées d'étude Interopérabilité pour l'utilisation généralisée de la Géoinformation*, Etat de la technique, implémentation, Zürich, Suisse, Mars 2005. Swiss Federal Institute of technology.
- [78] Marc Gilgen. Méta-information de données géoréférencées, Décembre 1999.
- [79] Géoconnexions. Manuel pour les développeurs de l'icdg : produire et publier l'information, les données et les services géographiques. Technical report, february 2004.
- [80] Françoise Gourmelon, Christophe Le Visage, and Marc Robin. *Des SIG spécifiques pour un espace spécifique ?*, pages 21–46. Hermes Science, Juillet 2005.
- [81] Pierre Grabar and Natalia Zweigenbaum. Liens morphologiques et structuration de terminologie. 1999.
- [82] John Graybeal, Luis Bermudez, Philip Bogden, Steven Miller, and Stephanie Watson. Marine metadata interoperability project : Leading to collaboration. Technical report, 2003.
- [83] Pierre Grenon. The formal ontology of spatio-temporal reality and its formalization. *American Association for Artificial Intelligence*, 2002.
- [84] Thomas Robert Gruber. A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2) :199–220, 1993.
- [85] Julie Harris. Mida - a web portal to coastal and marine data in ireland. *Coast Map News*, (5) :1–3, Winter 2004.
- [86] Carsten Heidmann and Rainer Lehfeldt. Information infrastructure for iczm tasks. In *6th International Conference on Hydroinformatics*, Vienna, Austria, 2004. Liong, Phoon and Babovic (eds), World Scientific Publishing Company.
- [87] Yves Hénocque and Jacques Denis. *Des outils et des hommes pour une Gestion Intégrée des Zones Côtières, guide méthodologique*, volume IOC Manuals and Guides No. 42 UNESCO. 2001.
- [88] Gail Hodge. *METADATA MADE SIMPLER*. NISO Press, 2001.
- [89] Catherine Houstis and Spyros Lalis. Arion : An advanced lightweight architecture for accessing scientific collections. 4, Mai 2001.
- [90] Catherine Houstis, Spyros Lalis, Vassilis Christophides, Dimitris Plexousakis, Manolis Vavalis, Marios Pitikakis, Kyriakos Kritikos, Antonis Smardas, and Charalampos Gikas. A service infrastructure for e-science :the case of the arion system. In *CAiSE '02/ WES '02 : Revised Papers from the International Workshop on Web Services, E-Business, and the Semantic Web*, pages 175–187, London, UK, 2002. Springer-Verlag, London, UK.

- [91] Dess IAO. Les objets géographiques du programme syscolag, may 2004.
- [92] IFREMER. Espace et littoral, le zonage comme outil de gestion du littoral méditerranéen, développement d'un cadre conceptuel et application. *Rapport interne IFREMER, Direction de l'Environnement et de l'aménagement littoral*, Septembre 2003.
- [93] IGN. Lexique du serveur éducatif de l'ign et de l'éducation nationale sur l'information géographique. <http://seig.ensg.ign.fr/sommlexi.php3>, 2005.
- [94] Sustainability Institute. Sustainability institute web site. <http://sustainer.org/>, 2005.
- [95] IOC and IGU. Gis for coastal management and coastal education, 2004.
- [96] Joliffe I.P. and C.R. Patman. The coastal zone : The challenge. *Journal of Shoreline Management*, 1(1) :3–36, 1985.
- [97] Dean Jones, Trevor Bench-Capon, and Pepijn Visser1. Methodologies for ontology development. 1998.
- [98] Vipul Kashyap. Information modeling on the web : The role of metadata, semantics and ontologies. *Practical Handbook of Internet Computing CRC Press M. Singh (editor)*, 2001.
- [99] Gilles Kassel, Marie-Hélène Abel, Catherine Barry, P. Boulitreau, C. Irastorza, and S. Perpette. Construction et exploitation d'une ontologie pour la gestion des connaissances d'une équipe de recherche. pages 60–71, 2000.
- [100] Robert Kay, Jacqueline Alder, David Brown, and Paul Houghton. Management cybernetics : A new institutional framework for coastal management. *Coastal Management*, 31 :213–227, 2003.
- [101] Robert Kay and Patrick Christie. An analysis of the impact of the internet on coastal management. *Coastal Management*, (29) :157 181, 2001.
- [102] Wassili Kazakos, Ralf Kramer, and Andreas Schmidt. Coastbase - the virtual european coastal and marine data warehouse, Oct 2000.
- [103] Wassili Kazakos, Andreas Schmidt, and Heiko Paoli. Xml based virtual catalogue module in coastbase. In Lorenz Hilty and Paul Gilgen, editors, *Sustainability in the Information Society. 15th International Symposium Informatics for Environmental Protection*, volume II, pages 513–520, Marburg, Germany, Oct 2001. Metropolis-Verlag.
- [104] Wassili Kazakos, Aleksei Valikov, Andreas Schmidt, and Rainer Lehfeldt. Automation of metadata repository generation with xml schema. In *in proceedings of 16 th International Symposium Environmental Informatics 2002 (Enviro Info)*, Vienna, Austria, 2002.
- [105] Versions Michel Klein. Finding and specifying relations between ontology.
- [106] Ywenn De la Torre, Françoise Gourmelon, and Claude Augris. Evolution morpho-sédimentaire des fonds marins côtiers. base d'information géographique, méthodes et limites. *Revue Internationale de Géomatique*, 12(3) :311–324, Décembre 2002.
- [107] Raouf Latypov. Mise en place d'une base de données référentielle thématique et spatiale dans le cadre du programme syscolag, Septembre 2005.
- [108] Lavoisier, editor. *SIG côtiers*, volume 12. Hermes Science, Décembre 2002.
- [109] Lavoisier, editor. *Les ontologies spatiales*, volume 14. Hermes Science, 2004.
- [110] Lavoisier, editor. *SIG et littoral*. Hermes Science, Juillet 2005.

- [111] Rainer Lehfeldt and Volker Barthel. Information requirements to support integrated coastal zone management. In *3rd GMES Forum*, Athens, Greece, 5-6 June 2003.
- [112] Rainer Lehfeldt and Carsten Heidmann. Nokis : a metadata information system for german coastal regions of the north sea and baltic sea. In *HydroInformatics 2002*, Cardiff, UK, 2002.
- [113] Rainer Lehfeldt and Carsten Heidmann. Information infrastructure for integrated coastal zone management. In *XXX IAHR Congress*, pages 465–472, Thessaloniki, Greece, August 25-29 2003.
- [114] Rainer Lehfeldt and Carsten Heidmann. Metadata application in coastal zone management. In *6 th International Conference of Hydrosience and Engineering (ICHE-2004)*, Brisbane, Australia, May 30-June 2004. Liong, Phoon and Babovic (eds), World Scientific Publishing Company.
- [115] Rainer Lehfeldt, Carsten Heidmann, and M. Piasecki. Metadata in coastal information systems. In *5th International Conference on Hydrosience and Engineering*, Warsaw, 2002.
- [116] Thérèse Libourel, Jean Christophe Desconnets, Pierre Maurel, Nicolas Moyroud, and Michel Passouant. Les métadonnées : pourquoi faire ? In *Géoévénement 2003*, volume in CD-ROM, 2003.
- [117] Thérèse Libourel, Isabelle Mougenot, Jean Sallantin, and Laurent Spéry. Meta-data and biological sequence annotation. In *Third IEEE META-DATA Conference META-DATA '99*, Bethesda, MD, USA, Avril 1999.
- [118] Erick Lopez-Ornelas, Florence Sèdes, and Guy Flouzat. Une approche pour la description et l'interrogation d'images satellitaires à très haute résolution spatiale. In *Cinquièmes journées Extraction et Gestion des Connaissances (EGC 2005)*, Méta-données et Adaptabilité pour les Systèmes d'Information sur le Web, Paris, France, Janvier 2005.
- [119] Jacques Madelaine, Christophe Turbout, and Mauro Gaio. De la gestion à la réutilisation d'information géographique hétérogène multi-sources sur internet. In *Journées Cassini 2002*, 2002.
- [120] Philippe Massonet. La création d'ontologies web sémantique avec protégé-2000. <http://geo-nsdi.er.usgs.gov/talk/thesaurus/outline.html>.
- [121] E. Mata, J.A. Bañares, J. Gutierrez, P.R. Muro-Medrano, and J. Rubio. Semantic disambiguation of thesaurus as a mechanism to facilitate multilingual and thematic interoperability of geographical information catalogues. In *5th AGILE Conference on Geographic Information Science*,., Palma (Balearic Islands, Spain), 25th-27th Avril 2002.
- [122] Pierre Maurel, Thérèse Libourel, and Julien Barde. Building spatial data infrastructures based on spatial cognition : an example for integrated coastal zone management. *Environment and Planning B*, -(-) :en cours, 2005.
- [123] Donella H. Meadows, Dennis L. Meadows, and Jorgen Randers. *Beyond the limits*. Chelsea Green Publishing Company, 1992.
- [124] Donella H. Meadows, Dennis L. Meadows, Jorgen Randers, and W.W. Behrens. *The limits to Growth*. Universe Boks, New York, 1972.
- [125] MedWet. Mediterranean wetlands initiative. <http://www.medwet.org/>, 2005.

- [126] Keiran Millard, Rob Atkinson, Andrew Woolf, Roy Lowry, Pieter Haaring, Francisco Hernandez, Torill Hamre, Dan Pillich, Holger Bothien, Ward Vanden Berghe, Ian Johnson, Brian Matthews, Hans Dahlin, Nic Flemming, Vladimir Vladymyrov, Bev Meckenzie, Alan Edwards, Willem van der Hoeven, Quillon Harphen, and Gil Ross. Using xml technology for marine data exchange. a position paper of the marinexml initiative. Technical report, Mars 2005.
- [127] Théodore Monod. *Et si l'aventure humaine devait échouer*. Livre de poche, 2002.
- [128] F. Muller, Jean Paul Donnay, K. De Cauwer, and L. Schwind. Design of oceanographic database. Geomatics and coastal environment, pages 124–133, Brest, France, 2000. Actes de colloques Ifremer / Shom.
- [129] Leitzelman Mylène and Henri Dou. Typology of information systems ; essai de typologie des systèmes d'informations. *International Journal of Information Sciences for Decision Making*, 2 :55–73, 1998.
- [130] Noy Fridman N. and D.L. Mc Guinness. Ontology development 101 : A guide to creating your first ontology, technical report, 2001.
- [131] Douglas D. Nebert. Developing spatial data infrastructures : The sdi cookbook. page 46, January 2004.
- [132] Irene Niesing and Hugo Lucius. Coastbase as a tool for integrated coastal zone management. In Ed. EUROCOAST / EUCC, editor, *Littoral 2002, The Changing Coast.*, Porto, Portugal, 2002.
- [133] J. Nogueras-Iso, F.J. Zarazaga-Soria, J. Lacasta, R. Béjar, and P.R. Muro-Medrano. Metadata standard interoperability : application in the geographic information domain. *Computers, Environment and Urban Systems*, 28 :611–634, 2004.
- [134] J. Nogueras-Iso, F.J. Zarazaga-Soria, J. Lacasta, R. Béjar, and P.R. Muro-Medrano. Ogc catalog services : a keyelement for the development of spatial data infrastructures. *Computers & Geosciences*, 31 :199–209, 2005.
- [135] NOKIS. North sea and baltic sea information system. <http://nokis.baw.de/>, 2005.
- [136] Natalya F. Noy, Michael Sintek, Stefan Decker, Monica Crubezy, Ray W. Ferguson, and Mark A. Musen. Creating semantic web contents with protege-2000. *IEEE Intelligent Systems*, 2(16) :60–71, 2001.
- [137] Liz O'Dea, Valerie Cummins, and Edward Dwyer. Developing an informational web portal for coastal data in ireland : Data issues in the marine irish digital atlas. pages 32–40, London, England, 18 March 2004.
- [138] Liz O'Dea, Edward Dwyer, Valerie Cummins, Didac Perales I Gimenez, and Declan Dunne. Harmonising marine information exchange in ireland. Aberdeen, Scotland, July 2005.
- [139] Report of the United Nations Conference on Environment and Development. United nations. <http://www.un.org/documents/ga/conf151/aconf15126-1annex1.htm>.
- [140] Ontopia. Omnigator knowledge browser. <http://www.ontopia.net/omnigator/>, 2005.
- [141] Nicholas M. Patrikalakis, Stephen L. Abrams, J. G. Bellingham, W. Cho, K. P. Miha-netzis, Allan R. Robinson, Henrik Schmidt, and Pubudu C. H. Wariyapola. The digital ocean. In *Computer Graphics International*, pages 45–54, 2000.

- [142] Domenico M. Pisanelli, Geri Steve, and Aldo Gangemi. Ontologies and information systems : the marriage of the century ? In *Lyee workshop*, Paris, 2002.
- [143] PNUE-PAM-PAP. *Cadre conceptuel et directives pour la gestion intégrée du littoral et des bassins fluviaux*. Split, Croatie, 1999.
- [144] POSEIDON. A distributed information system for ocean processes. <http://czms.mit.edu/poseidon/new1/>, 2004 2004.
- [145] Roland Prelaz-Doux. *Système d'information et gestion du territoire*. Lausanne : PPUR, 1995.
- [146] Les principes du développement durable. Ministère du développement durable, de l'environnement et des parcs du québec. <http://www.menv.gouv.qc.ca/developpement/principe.htm>, 2005.
- [147] Centre Méditerranéen d'Information en Matière de GIZC Programme d'Actions Prioritaires. http://www.pap-medclearinghouse.org/fra/toolbox_start.asp. .
- [148] D.V. Pullar and Max J. Egenhofer. Towards formal definition of topological relations among spatial objects. In *3RD International Symposium on Spatial Data Handling*, pages 225–241, Sydney, Australia, 1988.
- [149] Office québécois de la langue française. Grand dictionnaire terminologique. <http://w3.granddictionnaire.com/>, 2005.
- [150] Pôle relais Lagune Méditerranée. Thésaurus du pôle relais lagune méditerranée. <http://pole-lagunes.org/>, 2005.
- [151] Hélène Rey-Valette. Interactions société/lagune : Enjeux et situation de référence dans le cas de la lagune de thau et de son bassin versant. page 23, 2004.
- [152] Hélène Rey-Valette, J. Brisard, and A. d'Artigues. Situation de référence des activités anthropiques et des usages au niveau de la lagune de thau et de son bassin. 2004.
- [153] Hélène Rey-Valette and Sébastien Roussel. Analyse de la capacité des territoires littoraux à partir de la définition d'une notion de marge de manoeuvre ou de capacité territoriale. page 18, Paris, 2005.
- [154] Marc Robin, Véronique Morsetti, Dominique Pichereau, Pascal Babillot, Sébastien Colas, Caroline Texier, and Angélique Thuillier. *L'information géographique de référence en zone côtière : la situation française*, pages 49–82. Hermes Science, Juillet 2005.
- [155] J. Rondeux and François Leruth. Guide d'utilisation de la base de méta-données metater. *Conférence Permanente du Développement Territorial C.P.D.T., cellule bases de données, Volet données*, page 73, Octobre 2001.
- [156] J. Rondeux and François Leruth. Les métadonnées cartographiques en région wallone. *Conférence Permanente du Développement Territorial C.P.D.T., cellule bases de données, colloque S.I.G.*, page 11, Novembre 2001.
- [157] Guillaume Rousse. *Vers une ontologie de la systématique Les objets-metier du domaine*. PhD thesis, universite Pierre et Marie Curie Paris 6, 2003.
- [158] François Rousselot and Pierre Frath. Terminologie et intelligence artificielle. 2000.
- [159] P.C. Smits, U. Düren, O. Ostensen, L. Murre, M. Gould, U. Sandgren, M., Marinelli, K. Murray, E. Pross, A. Wirthmann, F. Salgé, and M. Konecny. Infrastructure for spatial information in europe. page 64, 2002-10-03.

- [160] Genoveva Vargas Solar and Anne Doucet. Médiation de données : solutions et problèmes ouverts. In *Actes des deuxièmes assises nationales du GdR I3*, Modèles de données et conception des systèmes d'information, Nancy, Décembre 2002.
- [161] Michela Spagnuolo and Simone Marini. User/scientific collection/system requirements, 2001.
- [162] Christos Stefanakos, Theodoros Gerostathis, Gerasimos Athanasoulis, Catherine Houstis, and Emmanuel Vavalis. Building ontologies of environmental applications for a digital library of scientific collections. In *EnviroInfo*, Vienna, Autriche, 2002.
- [163] L. Stoimenov and S. Dordevic-Kajan. Realization of gis semantic interoperability in local community environment. pages 73–80, Lyon, France, April 20-23 2003.
- [164] SYSCOLAG. *Convention cadre Programme Systèmes Côtiers et Lagunaires*. 2003.
- [165] THETIS. A data management and data visualization system for coastal zone management of the mediterranean sea. <http://www.cacr.caltech.edu/ewus/documents/houstis.html>, 2005.
- [166] Mathieu Le Tixerant. Représentation logique et spatiale de la réglementation des activités humaines en zone côtière. *Revue Internationale de Géomatique*, 12(3) :325–335, Décembre 2002.
- [167] Matthieu Le Tixerant, Mathias Rouan, Françoise Gourmelon, and François Cuq. Simulation of human activities dynamics (dahu) in coastal seas. Genova, Italy, July 2003.
- [168] Roland Trique, Frédéric de Solliers, and Nat Makarévitch. Dictionnaire d'informatique francophone. <http://www.linux-france.org/prj/jargonf/>, 2005.
- [169] Christophe Turbout, Mauro Gaio, and Jacques Madelaine. Placides : Plateforme d'aide à la construction d'information dynamique Électronique spatialisée. In *H2PTM'01 : nouvelles écritures, nouveaux langages*, Valenciennes, France, Octobre 2001. Hermès Sciences.
- [170] UNESCO. *Guide méthodologique d'aide à la gestion intégrée de la zone côtière*, volume IOC Manuals and Guides No. 36 UNESCO. juillet 1997.
- [171] UNESCO. *A Reference Guide on the Use of Indicators for Integrated Coastal Management- ICAM Dossier 1*. IOC Manuals and Guides No. 45 UNESCO, 2003.
- [172] USGS. Organizing usgs information with consistent vocabulary. <http://geo-nsdi.er.usgs.gov/talk/thesaurus/outline.html>, 2005.
- [173] Romain Vallée. Sig et sgbd : Etude sur l'interconnexion et l'interopérabilité de oracle 9i et de postgresql, Juillet 2004.
- [174] Carolina Pérez Valverde. Distance iczm training for central and eastern europe. In Ed. EUROCOAST / EUCC, editor, *Littoral 2002, The Changing Coast.*, Porto, Portugal, 2002.
- [175] Cyril Vidal. Théorie et pratique du web sémantique, may 2002.
- [176] World Wide Web Consortium (W3C). Web ontology language (owl) use cases and requirements. <http://www.w3.org/TR/webont-req/>, Février 2004.
- [177] P. C. H. Wariyapola, Nicholas M. Patrikalakis, S. L. Abrams, P. Elisseeff, A. R. Robinson, H. Schmidt, and K. Streitlien. Ontology and metadata creation for the poseidon distributed coastal zone management system. In *Advances in Digital Libraries*, pages 180–189, 1999.

- [178] The Marine Metadata Interoperability web site. Marine metadata proposal, marine metadata wiki. *[http ://gustnado.mbari.org/marinemetadawiki/MipProposal](http://gustnado.mbari.org/marinemetadawiki/MipProposal)*, 2005.
- [179] L’encyclopédie libre Wikipédia. *[http ://fr.wikipedia.org/wiki/](http://fr.wikipedia.org/wiki/)*.

Table des figures

2.1	Redéfinir le rapport entre l'homme et les ressources naturelles (Photo : Anita Conti)	25
2.2	Zone côtière : continuum terre-mer	31
2.3	"La ronde des acteurs" dans la GIZC d'après [87]	32
2.4	"Boîte à outils" pour la GIZC d'après [147]	37
2.5	Place du SI dans le processus de GIZC d'après [87]	38
2.6	Une zone côtière particulière : le littoral du Languedoc Roussillon	42
2.7	Position de la thèse dans le programme Syscolag	45
2.8	Les verrous analysés par le MMI [82]	48
2.9	Le SI FIGIS	50
2.10	L'architecture du SI POSEIDON [141]	52
2.11	Le SI NOKIS	53
2.12	Le SI du MIDA	54
2.13	Le SI COASTBASE	56
3.1	Les trois niveaux de l'infrastructure Syscolag [16]	68
3.2	Synthèse des diagramme des cas d'utilisation des différents acteurs du service de métadonnées	71
4.1	Vue simplifiée de l'organisation générale d'une norme	82
4.2	Les différentes rubriques de l' <i>ISO 19115</i> [16, 2]	88
4.3	Détail du diagramme de classes <i>UML</i> pour la rubrique <i>Spatial representation</i> de l' <i>ISO 19115</i> [2]	90
4.4	Profil de l' <i>ISO 19115</i> [2]	92
4.5	tableau de correspondances entre éléments de métadonnées du <i>DC</i> et de l' <i>ISO 19115</i> [112]	93
4.6	Exemples de fichiers de métadonnées dans Arccatalog	95
4.7	Interface graphique utilisateur de recherche de métadonnées avec Geonetwork	98
4.8	Les différents types d'éléments de métadonnées vus par les partenaires de <i>Syscolag</i>	102
4.9	Importance particulière des descripteurs relatifs aux mots clés et à l'étendue géographique	104
4.10	Proposition de typologie de ressources informationnelles pouvant faire l'objet d'un gabarit	105
4.11	Approche générique d'une norme de métadonnées	107
4.12	Adaptation pour Syscolag du modèle générique multi-norme en modèle mixte qui distingue le mode de valuation selon la nature des éléments de métadonnées	109

4.13	Extrait de <i>MPD</i> conçu selon une approche mixte pour l'implémentation en SQL	112
5.1	Les différentes sortes de concepts pour renseigner l'élément <i>mot clé</i> de l'ISO 19115	120
5.2	Les différentes relations standard d'un glossaire	122
5.3	Illustration d'interfaces graphiques du glossaire <i>UNESCO</i>	123
5.4	Les différentes relations standards d'un thésaurus	125
5.5	Echantillon du thésaurus <i>ASFA</i> sous format texte tabulé	127
5.6	Illustration d'interfaces graphiques du thésaurus <i>ASFA</i> [66]	127
5.7	Illustration du thésaurus <i>AGROVOC</i> [64]	128
5.8	Illustration du thésaurus <i>GEMET</i> de l'Union Européenne [59]	129
5.9	Illustration du thésaurus du pôle relais lagunes méditerranéennes [150]	130
5.10	Exemple d'extension de la norme <i>ISO 19115</i> pour la mise en place d'une <i>taxonomie</i> pour décrire un nouveau type d'élément <i>mot clé</i> [2]	131
5.11	Les différentes relations standard d'une ontologie	134
5.12	Structure et éléments de l'ontologie [75]	137
5.13	Catégories générales des domaines de l'océanographie physique et acoustique dans l'ontologie <i>POSEIDON</i> [177]	138
5.14	Exemple d'interface Web pour la construction collaborative de l'ontologie <i>POSEIDON</i> [141]	139
5.15	Exemple d'ontologie <i>ARION</i> pour le domaine de la houle océanique [162]	140
5.16	Complémentarité des différents outils de contrôle du vocabulaire	140
5.17	Les différents langages du <i>Web sémantique</i> 5.17(a), la formalisation élémentaire de la connaissance employée 5.17(b) et un extrait de l'ontologie <i>FOS</i> au format <i>OWL</i> 5.17(c)	143
5.18	Premiers travaux de représentations des concepts et de leurs relations dans le domaine de la <i>GIZC</i>	145
5.19	Evolution des modèles validés avec les partenaires	146
5.20	La vision binaire de l'approche écosystémique opposant <i>sphères humaine</i> et <i>naturelle</i> 5.20(a) et un premier travail de catégorisation et de spatialisation des <i>éléments</i> et <i>activités</i> liés à au <i>développement humain</i> sur la zone côtière 5.20(b)	148
5.21	Approche écosystémique pour la structuration des composantes principales de la zone côtière [143]	149
5.22	Diagramme finalisé du modèle de l'ontologie [18]	150
5.23	Spatialisation des concepts du modèle d'ontologie	152
5.24	Complémentarité entre normes ISO et vocabulaires contrôlés [15]	154
5.25	Traduction des recommandations de la <i>GIZC</i> 5.25(a) en typologies d'acteurs (individus ou organismes) utilisant le formalisme UML : 5.25(b) et 5.25(c)	157
5.26	Utilisation des recommandations de l'ISO 19115 5.26(a) pour l'attribution de propriétés aux concepts acteur 5.26(b) et organisme 5.26(c) de notre ontologie	158
5.27	Exemples extraits de typologies de ressources naturelles 5.27(a) ou artificielles 5.27(b) d'intérêt pour la <i>GIZC</i>	159
5.28	Diagramme de classes traduisant le point de vue des <i>géologues</i> (variété de concepts dérivant du concept <i>terre</i>) [91]	161
5.29	Utilisation des sources de référence pour l'inventaires des activités et usages sur la zone côtière	163

5.30	Utilisation de sources de référence 5.30(a) pour alimenter la typologie d'interactions de notre ontologie 5.30(b)	165
5.31	Utilisation de sources de référence 5.30(a) pour alimenter la typologie d'interactions de notre ontologie 5.30(b)	167
5.32	Diagramme des classes de la typologie des outils impliqués dans la GIZC	168
5.33	Exemple de scénario sur la " <i>pêche au merlu</i> " formalisé à l'aide de notre <i>modèle d'ontologie</i>	169
5.34	Synthèse sur le lien du modèle conceptuel du référentiel sémantique avec le service de métadonnées 5.34(a) et la nature des concepts qui l'alimentent pour le domaine de la GIZC 5.34(b)	170
5.35	Modèle physique de données du référentiel sémantique implémenté dans <i>Postgres</i> [107]	172
6.1	Les différentes rubriques de l'ISO 19115 d'après [2]	178
6.2	Illustration des trois types de descripteurs géographiques de l'ISO 19115	179
6.3	Structure graphique et tabulaire de l'information géographique dans un SIG-outil	182
6.4	Lien entre concepts thématiques et spatiaux	182
6.5	Modèle conceptuel des référentiels thématiques et spatiaux	183
6.6	Ensemble minimal de relations topologiques entre deux objets d'après [118] . .	184
6.7	Synthèse des relations entre l'ISO 19115 et les référentiels sémantique et spatial	185
6.8	Evolution de l'ontologie en <i>ontologie spatiale</i>	185
6.9	Modèle physique simplifié de données issu du modèle conceptuel	187
6.10	Structure graphique et tabulaire de l'information géographique	189
6.11	Duplication des géométries et toponymes des enregistrements des tables <i>Layer_\$concept</i> ayant servi à la description spatiale d'une ressource	190
6.12	Exemple de <i>requêtes spatiales</i> possibles pour consulter les fiches de métadonnées disponibles sur l' <i>étang de Thau</i>	191
6.13	Exemple de contenu de la méta-table <i>geometry_columns</i> dans <i>Postgis</i> : exemples d'enregistrements	194
6.14	Exemple de <i>table spatiale</i> département du type <i>Layer_\$concept</i> générée par l'utilitaire d'import <i>shp2pgsql</i> dans une base <i>Postgis</i>	194
6.15	L'approche écosystémique au service de l'inventaire de l'IG pertinente	201
6.16	Exemples de descriptions de <i>concepts spatiaux</i> en mode <i>textuel</i> (6.16(a)) ou <i>graphique</i> (6.16(b)) pouvant être utilisés pour cartographier la zone côtière (6.16(c)) et perspectives 3D (6.16(d))	203
6.17	Exemple de d'analyse et de formalisation <i>UML</i> d'un entretien conduit auprès de l' <i>IFREMER</i>	205
6.18	Le modèle d'ontologie à quatre classes	207
6.19	Utilisation des liens entre <i>concepts thématiques</i> et <i>spatiaux</i> pour générer <i>dynamiquement</i> une <i>structuration thématique</i> des <i>concepts spatiaux</i> et <i>toponymes</i> du <i>référentiel spatial</i>	209
7.1	Le choix de <i>PHP</i> comme langage de script pour notre portail Web [179]	215
7.2	Formulaire (<i>X</i>) <i>HTML</i> généré dynamiquement par <i>PHP</i> pour un <i>gabarit</i> défini dans la base de métadonnées	217
7.3	Exemple de <i>GUI</i> pressentie pour renseigner les mots clés	218
7.4	Exemples d'outils de visualisation de réseau sémantique	219

7.5	Utilisation du langage SVG pour des représentations graphiques en mode <i>vecteur</i>	220
7.6	Cas d'utilisation associés au <i>SMD</i>	222
7.7	Cas d'utilisation associés à la saisie en mode texte dans le <i>SMD</i> des mots clés et toponymes contenus dans le référentiel sémantique	222
7.8	Exemple de GUI pressenti pour renseigner les mots clés avec des <i>concepts thématiques</i> ou <i>spatiaux</i> et des <i>toponymes</i>	223
7.9	Cas d'utilisation associés à la saisie des descripteurs géographiques avec l'interface cartographique dans le <i>SMD</i>	224
7.10	Description géographique d'une ressource dans l' <i>ISO 19115</i>	224
7.11	Description géographique d'une ressource dans l' <i>ISO 19115</i>	225
7.12	Insertion des différents <i>composants</i> de <i>contrôle</i> ou d' <i>aide</i> à la <i>valuation</i> des <i>éléments de métadonnées</i> essentiels pour la GIZC dans les <i>formulaires HTML</i> associés aux <i>gabarits de saisie</i> du <i>SMD</i>	227
7.13	Cas d'utilisation associés à la consultation du <i>SMD</i>	228
7.14	Exemple d' <i>GUI</i> pressenti pour les recherches multicritères pour la consultation de métadonnées	228
7.15	Cas d'utilisation associés à la consultation des résultats d'une requête	229
7.16	Maquette 7.16(a) et prototype 7.16(b) d'interfaces graphiques de réponses à une requête de consultation du <i>SMD</i>	230
7.17	Cas d'utilisation associé à la génération d'un nouveau <i>profil</i> ou <i>gabarit</i>	231
7.18	Outil d'édition de gabarits	232
7.19	Les différents cas d'utilisation liés à la visualisation du contenu du <i>référentiel sémantique</i>	233
7.20	Les différents cas d'utilisation liés à la visualisation du contenu du <i>référentiel sémantique</i> hors représentations cartographiques	234
7.21	Visualisation du <i>référentiel sémantique</i> sous la forme d'une <i>base terminologique</i>	235
7.22	Visualisation du <i>référentiel sémantique</i> sous la forme d'un <i>réseau sémantique</i>	235
7.23	Exemple d'approche mixte entre <i>base terminologique</i> et <i>réseau sémantique</i>	236
7.24	Consultation du contenu du <i>référentiel spatial</i> à l'aide d'un <i>atlas cartographique</i>	237
7.25	<i>Structuration thématique</i> des <i>Layers</i> et <i>visualisation</i> de leur contenu par des <i>interfaces cartographiques</i> sur le Web	238
7.26	Extension de l'utilisation de l' <i>interface cartographique</i> du <i>SMD</i> en <i>atlas</i> indépendant	239
7.27	Passage d'une application à l'autre par liens hypertextes	240
8.1	Exemple de connexion à une base <i>Postgis</i> avec le logiciel libre <i>Quantum GIS</i>	251
8.2	Sites de diffusion de la connaissance sur le Web vers différents acteurs : scientifiques 8.2(a) ou pêche de loisirs 8.2(b) et formalisation de ces connaissances en vue d'une d'intégration dans notre ontologie 8.2(c) et 8.2(d)	254
8.3	Exemple d'utilisation de sources de référence du domaine (voir Figure 8.2) pour l'attribution de propriétés aux classes de notre ontologie	255
8.4	Intégration de la dimension verticale et temporelle sous la forme d'attributs de la classe <i>Objet géographique</i> de l' <i>ISO 19115</i>	256
8.5	Position relatives des référentiels sémantique, spatial, temporel et de l' <i>ISO 19115</i>	257
8.6	Modèle d'ontologie proposé à l'issue de ce travail	258
8.7	Formalisation UML possible pour un scénario	258

Glossaire

analyse spatiale Raisonnement qui permet de déduire les caractéristiques d'un phénomène en faisant intervenir des données géographiques.

L'analyse spatiale exploite l'information topologique et l'information métrique.

Les systèmes de gestion de base de données géographiques offrent des fonctions d'analyse spatiale utilisant des opérateurs dont au moins un est spatial. Le résultat de ces traitements est utilisé par l'utilisateur pour effectuer l'analyse spatiale [149]. Terme anglais : spatial analysis.

page 254

annuaire électronique Annuaire organisé en base de données et interrogeable à partir d'un micro-ordinateur. Note(s) : Les annuaires électroniques peuvent être des listes de personnes ou de services. Les termes autres que " annuaire électronique " font référence à un annuaire téléphonique traditionnel, transféré sur une base de données et interrogeable à partir d'un Minitel (en France) ou d'un micro-ordinateur. C'est par extension que "annuaire électronique" désigne toutes les formes d'annuaires organisés en bases de données. Quasi-synonyme(s) : annuaire informatisé, annuaire informatique, annuaire téléphonique électronique, répertoire électronique, répertoire téléphonique électronique [149]. En anglais : electronic directory.

page 157

application Web Application pouvant tourner sur un serveur Web. Note(s) : une application Web n'a pas été nécessairement développée pour être utilisée dans le réseau Internet. Cependant, son utilisation avec d'autres applications complémentaires dans une architecture client-serveur la rend apte à cela [149].

page 99

base de connaissance Élément d'un système expert contenant des informations représentant les connaissances acquises dans un domaine particulier [149]. En anglais : Knowledge Base.

page 65

base de données géographiques Base de données constituée de données géographiques. C'est une des parties les plus importantes du système d'information géographique (SIG). Elle se présente généralement comme un ensemble de sujets thématiques qui se superposent à une carte géographique numérisée.

Même si on peut associer le mot spatial à des systèmes de référence autres que géodésiques, c'est en contexte géomatique qu'on l'utilise le plus souvent [149].

Terme anglais : geographic information system database. Synonymes : base de données à référence spatiale, base de données spatiales, base de données géospatiales, base de

données localisées. Synonymes anglais : GIS database, geographic database, geospatial database, spatial database. Terme non retenu : BD spatiales. Termes non retenus : geographic information database, geo-referenced database, GRDB.

page 66

cadre de référence Ensemble d'informations significatives accumulées au cours de la vie et auxquelles un individu se réfère pour interpréter toute donnée ou situation nouvelle. Système de valeurs et d'attitudes, généralement implicite, né de l'expérience vécue du sujet ou du groupe. Note(s) : ce système exerce une influence profonde dans les domaines cognitif, affectif et conatif [149].

page 66

cohérence des données Propriété selon laquelle les données contenues dans la base de données reflètent bien la réalité qu'elles représentent.

Quand la cohérence des données est respectée, il n'y a pas de contradiction entre les données et elles ne créent pas de confusion pour ceux qui les utilisent. La cohérence des données inclut la validité des données et l'intégrité des données. En contexte, la forme abrégée "cohérence" est fréquemment utilisée pour désigner la cohérence des données [149].

page 136

composant Logiciel, programme ou élément d'un logiciel ou d'un programme qui constitue un module indépendant utilisé comme élément d'un système plus complexe et qui est spécialement conçu pour fonctionner sans problèmes avec d'autres logiciels ou programmes [149]. Élément d'un système. On parle de "composant électronique" pour désigner les puces, et de "composant logiciel" pour désigner les petits bouts de programme réutilisables une fois sur deux (objet) [168].

page 220

concept Élément de la pensée, exprimé en général par un terme ou par un symbole littéral ou autre. En informatique, à un concept donné on peut associer son extension c'est-à-dire la classe des entités qui tombent sous le concept.

Dans le vocabulaire de l'intelligence artificielle, le mot concept tend à être remplacé par "objet" [149].

page 116

connaissance Ensemble de faits, événements, règles d'inférence et heuristiques permettant à un programme de fonctionner intelligemment. On distingue traditionnellement, en intelligence artificielle, la connaissance et le raisonnement. Bien que cette distinction ne soit certainement pas si nette dans le cerveau humain, la séparation a été très fructueuse pour la mise en juvre de programmes "intelligents". Les questions liées à la connaissance sont celles de son acquisition (apprentissage), de sa représentation et de son utilisation (raisonnement) [149].

page 134

correspondance Opération qui consiste à vérifier si deux éléments de données sont identiques. Matching en anglais [149].

page 86

Document Type Definition Le Document Type Definition (DTD), ou Définition de Type de Document, est un document permettant de décrire un modèle de document SGML

ou XML. Une DTD ne décrit cependant que la structure du document (hiérarchie des champs, paramètres, type des données...) et non, par exemple, les valeurs autorisées des champs ou paramètres, ce en quoi elle se distingue de Schéma XML, Relax NG et Schematron. De plus, la norme DTD fait appel à une syntaxe spécifique distincte de XML. Une DTD n'est donc pas un document XML [179].

page 253

domaine Cadre délimitant un champ d'intervention, dans lequel sont appliquées des règles de gestion et de fonctionnement, ou des actions [149].

page 64

donnée Dans les technologies de l'information (TI), une donnée est une description élémentaire, souvent codée, d'une chose, d'une transaction d'affaire, d'un événement, etc. Elle est conservée et classée sous forme numérique, alphabétique, d'image, de son, ... Dans les grands services informatiques, les données doivent être répertoriées et organisées de façon à ce qu'elles soient aisément trouvées et manipulées par tous les utilisateurs potentiels et par la communauté de développeurs. Très souvent, il existe une ou plusieurs personnes chargée(s) d'organiser les données du système d'information de l'entreprise : ce sont les Administrateurs De Données (ADD) [179].

page 35

entité géographique Phénomène repérable au sol, par l'intermédiaire des sens ou non, et pouvant être représenté sur une carte. Note : L'entité géographique est au stade de notion, alors que l'objet géographique en est sa représentation [149].

page 184

extension voir intension [179].

page 186

glossaire Répertoire des termes tirés d'un **corpus** pour leur difficulté de compréhension et pour lesquels il est donné un **synonyme** connu ou une explication [149].

page 121

Graphical User Interface Une interface conviviale pour utiliser les programmes : environnement graphique, intégration de la souris, etc. Le GUI est à l'origine du succès actuel d'Internet. <http://www.dicofr.com>. ou Interface utilisateur basée sur l'emploi d'éléments graphiques tels que les fenêtres, les icônes et les menus, qui visent la simplicité d'emploi et qui créent un environnement de travail convivial. Le terme interface utilisateur est plus englobant que le terme interface utilisateur graphique, mais il est très souvent employé dans le sens de ce dernier, car c'est ce type d'interface utilisateur qui est le plus répandu actuellement [149].

page 218

indexation Caractérisation du contenu d'un document par des descripteurs de langage contrôlé. Toute caractérisation d'un document donné au moyen de mots empruntés à un vocabulaire généralement limité, pour en décrire la source (auteur, date, références, etc.) ou le contenu (sujet traité) [149].

page 102

inférence Opération qui permet de produire de nouvelles propositions à partir d'autres propositions tenues pour acquises [149].

page 142

information Élément de connaissance concernant un phénomène et qui, pris dans un contexte déterminé, a une signification particulière [149].

page 35

information géographique L'information géographique peut être définie comme l'ensemble de la description d'un objet et de sa position géographique à la surface de la Terre [179] ou comme de l'information déduite à partir de données géographiques [149]. L'information géographique peut être définie comme l'ensemble de la description d'un objet et de sa position géographique à la surface de la Terre [179]. Synonyme(s) : information à référence spatiale, information spatiale, information géospatiale, information localisée.

page 37

infrastructure de données spatiales ou SDI. L'infrastructure mondiale de données spatiales (Global Spatial Data Infrastructure) décrit ainsi les infrastructures : une SDI fournit une base pour la découverte, l'évaluation et l'application des données spatiales par des utilisateurs et des fournisseurs à tous les paliers gouvernementaux, aux secteurs commercial et sans but lucratif, aux milieux universitaires et par des citoyens en général.

page 197

infrastructure informatique Ensemble de toutes les ressources informatiques sur lesquelles reposent l'organisation de l'information et la communication des données à l'intérieur d'une entreprise, d'un organisme, d'un territoire ou d'un pays [149]. Synonyme(s) : infrastructure d'information, infostructure.

page 217

intégration En informatique : assemblage d'un système informatique à partir de différents composants. En écologie : opération ou processus par lequel des éléments, des ensembles, des individus ou des collectivités s'assemblent pour former un système présentant de nouvelles propriétés [149].

page 29

intension L'intension est un concept logique qui s'oppose à l'extension. Toute classe d'éléments peut être définie en extension (en nommant ou en désignant chaque individu qui en fait partie) ou en intension, par une description (spécification d'un certain nombre de prédicats) qui définit la classe. Par exemple, la classe des rois de France peut être désignée extensionnellement en donnant une liste de noms, ou intensionnellement par le concept "roi de France" (c'est-à-dire le prédicat, la propriété "être un roi de France") [179].

page 82

interopérabilité L'interopérabilité est le fait que plusieurs systèmes, qu'ils soient identiques ou radicalement différents, puissent communiquer sans ambiguïté et opérer (travailler) ensemble. En informatique, l'interopérabilité est une notion absolument cruciale pour le réseau téléphonique mondial et Internet, qui par essence sont des réseaux hétérogènes, où se croisent des matériels divers et variés, faisant tourner une panoplie encore plus vaste de logiciels [179].

L'interopérabilité nécessite que les communications obéissent à des normes, clairement établies et univoques. Si la norme est correctement écrite, deux systèmes qui satisfont aux exigences doivent dialoguer ensemble sans souci tout en évoluant librement tant qu'ils respectent la norme définissant leurs interfaces [179].

page 78

médiateur Dans le processus d'alimentation d'un entrepôt de données, logiciel spécialisé qui fusionne les données en provenance de sources multiples et diverses (bases de données internes et externes, fichiers, réseau Internet, etc.), après les avoir filtrées, transformées et réorganisées, et ce, de manière à en faire un ensemble homogène. Note : Les données traitées par le médiateur sont fournies par l'adaptateur de source [149]. Anglais : mediator.

page 77

médiation Le problème de la médiation peut être résumé de la manière suivante : étant donné un ensemble de sources (niveau local) et d'applications, il s'agit de construire une infrastructure intermédiaire facilitant l'accès (expression, traitement et optimisation de requêtes) et la manipulation (mise à jour, gestion de la cohérence) des données (niveau global). La médiation des données est guidée par trois aspects : la distribution et l'hétérogénéité des données, l'autonomie des sources, et l'interopérabilité des systèmes qui gèrent les bases [160].

page 79

méta-modèle Un métamodèle est une description très formelle de tous les concepts d'un langage. Il limite les ambiguïtés et encourage la construction d'outils (uml.free.fr)

page 91

métadonnées données sur les données ou données qui renseignent sur certaines données et qui permettent ainsi leur utilisation pertinente [21].

page 48

modèle conceptuel de données Modèle de données conçu indépendamment de l'équipement et des logiciels disponibles pour la gestion des données. synonyme(s) : modèle conceptuel, quasi-synonyme(s) : schéma conceptuel, abréviation(s) : MCD. Le modèle conceptuel de données regroupe les entités sur lesquelles portent les données, les relations que lient les entités entre elles et, enfin, les attributs des entités et des relations. On trouve trois niveaux d'abstraction de modèles de données : le modèle conceptuel de données, qui présente les informations telles que les voit l'entreprise, sans tenir compte de la technologie informatique disponible, le modèle logique de données, qui va plus loin en ne présentant que des entités normalisées et, enfin, le modèle physique de données, qui est un modèle logique adapté à une plateforme informatique précise [149].

page 109

modèle physique de données Modèle logique de données adapté spécifiquement à un système de gestion de base de données précis, pour une plate-forme déterminée. Par exemple, *Postgres*, *Oracle*, *Sybase* et *Informix* sont des exemples de plates-formes utilisables pour un modèle physique de données [149].

page 109

norme est un ensemble de règles de conformité plus ou moins optionnelles, édictées par un organisme de normalisation au niveau national ou international [57]. Il ne faut pas faire la confusion entre norme et, un standard (voir standard) qui désigne ce qui est produit habituellement par un producteur et ne dépend que de lui. Cette confusion vient de l'anglais qui n'a qu'un seul mot et pour qui standard signifie aussi norme. La norme ou la recommandation est établie par un organisme indépendant qui limite les modifications unilatérales. On comprend donc qu'il est inopportun de définir une interopérabilité à partir d'un standard non ouvert. Par exemple, la norme peut définir

les séquences d'informations ou de commandes qu'un système doit envoyer et, comment son correspondant doit y répondre [179].

page 82

Objet géographique Phénomène modélisé à des fins de représentation cartographique. Note : L'objet géographique est représenté par le point, la ligne et la surface, ou par une combinaison de ceux-ci. Terme anglais : geographic object. Synonyme anglais : geographic feature [149].

page 156

ontologie Ensemble d'informations dans lequel sont définis les concepts utilisés dans un langage donné et qui décrit les relations logiques qu'ils entretiennent entre eux [84].

page 47

Open Gis Consortium l'OGC est un consortium international regroupant plus de deux cent membres dans le monde entier. Il se consacre principalement à l'élaboration de solutions techniques visant à l'interopérabilité de systèmes d'information géographique. L'OGC a développé des accords de collaboration avec le TC211 de l'ISO mais également avec le consortium W3 ou le groupe de gestion des objets (Object Management Group, OMG 6). La norme sur les métadonnées est élaborée en coopération avec le TC 211 de l'ISO et la version préliminaire de la norme ISO/DIS 19115 a déjà été reprise en l'état comme "OGC Abstract Specification" [37].

page 47

Open Source Définition particulière du logiciel libre, mise au point en 1998 par Eric Raymond, cherchant à adapter le principe à l'entreprise. Elle comporte neuf points (pour le moment) : la libre redistribution, la mise à disposition du code source, la possibilité de distribuer ses travaux dérivés, le respect du code source originel, l'absence de discrimination envers des personnes, l'absence de limitation sur le domaine d'application du logiciel, la distribution de la licence et sa non-spécificité à un produit, et enfin le fait qu'elle ne contamine pas le travail des autres [168].

page 53

polysémie L'évolution du langage (due elle-même au fait qu'il faut bien décrire soit un monde qui évolue, soit un monde dont au moins notre connaissance évolue) conduit à utiliser parfois un mot dans un nouveau sens, le plus souvent par extension de sens. On parlera par exemple d'une feuille de papier ou du pied d'un arbre, par analogie avec une feuille d'arbre ou avec le pied d'un animal.

Lorsqu'un mot se trouve de facto avoir deux sens différents, on le qualifie de polysémique. C'est le cas d'une très grande majorité des mots courants du dictionnaire [179]. Il arrive même qu'un mot désigne ainsi à la fois une chose et son contraire : ainsi en est-il des mots :

- hôte, désignant selon le contexte celui qui reçoit ou celui qui est reçu,
- amateur, désignant selon le contexte une personne avertie ou ignorante.

page 116

portail Web Site Web dont la page d'accueil propose, en plus d'un moteur de recherche, des hyperliens avec une foule d'informations et de services utiles et attrayants, qui est conçu pour guider les internautes et faciliter leur accès au réseau. Synonyme(s) : site portail, portail Internet, portail Web. Quasi-synonyme : porte d'entrée [149].

page 52

programmation orientée objet Activité qui consiste à produire un programme informatique constitué d'un ensemble de sous-programmes (les objets) possédant chacun une identité et des propriétés qui lui sont propres, et qui peuvent communiquer et collaborer entre eux [149].

page 252

référentiel Ensemble structuré d'informations, utilisé pour l'exécution d'un logiciel, et constituant un cadre commun à plusieurs applications [168].

page 37

réseau sémantique Représentation des connaissances en vertu de laquelle les objets et les valeurs sont formalisés sous forme de noeuds, qui sont connectés aux arcs ou liens indiquant les relations entre divers noeuds. En général, les noeuds représentent des concepts et les arcs des relations entre les concepts. Les réseaux sémantiques présentent des difficultés, mais ils sont irremplaçables pour représenter des faits complexes ayant beaucoup de relations entre eux. Introduits par les psychologues comme modèle de la mémoire associative humaine, les réseaux sémantiques correspondent à un besoin de classification des concepts et d'établissement de modèle du monde. Aujourd'hui, les réseaux sémantiques servent de supports à d'autres formalismes : les systèmes de classification utilisés dans les systèmes experts [149]. En anglais : semantic network.

page 222

raster format RIFF (Raster Image File Format), variante plus compacte du format d'image TIFF, développée et employée par Letraset pour le stockage et l'exploitation des images numérisées, en particulier des images couleur. [149]

page 89

rectangle englobant Rectangle imaginaire qui entoure des éléments d'une image tout en respectant leurs dimensions. Synonymes :cadre de contour, boîte englobante. Anglais : bounding box [149].

page 183

requête Sélection des objets dans une base de données répondant à un caractère commun : descriptif (requête attributaire) ou géométrique (requête spatiale) [93].

page 199

ressources informationnelles Ressource utilisée par une organisation, dans le cadre direct de ses activités de traitement de l'information, pour mener à bien sa mission, pour la prise de décision, ou encore pour la résolution de problèmes. Une ressource informationnelle peut être une ressource humaine, matérielle ou financière directement affectée à la gestion, à l'acquisition, au développement, à l'entretien, à l'exploitation, à l'accès, à l'utilisation, à la protection, à la conservation et à la destruction des éléments d'information. Une ressource peut donc être une personne, un fichier ou le système informatique lui-même. L'expression ressource informationnelle est très souvent utilisée au pluriel (ressources informationnelles). Elle désigne alors un ensemble de ressources qui peuvent être répertoriées dans l'actif informationnel de l'organisation [149].

page 15

sémantique Ensemble des relations entre les caractères, ou groupes de caractères, et leur signification, indépendamment de la façon de les employer ou de les interpréter. Note(s) : Si, en linguistique, la sémantique porte sur l'étude du sens à partir de la combinaison

des mots, en intelligence artificielle, elle porte sur la capacité d'un réseau à représenter de la manière la plus humaine possible des relations entre des objets, des idées ou des situations. Par extension, la sémantique désigne également la signification d'un ensemble de mots dans un texte [149].

page 116

scénario Technique de représentation des connaissances sous la forme de situations préétablies [149].

page 55

Scalable Vector Graphics Grammaire XML définie par le W3C pour permettre la descriptions d'images 2D vectorielles en utilisant XML, les CSS, le PNG et le JPEG [168].

page 222

standard est un ensemble de recommandations émanant d'un groupe représentatif d'utilisateurs réunis autour d'un forum, comme l'IETF (Internet Engineering Task Force), le W3C (World Wide Web Consortium), le Dublin Core [57].

page 82

synonymie La synonymie est un rapport de proximité sémantique entre des mots d'une même langue. Des termes liés par synonymie sont des synonymes. Il existe un postulat qui explique qu'il ne peut pas exister de synonymes parfaits, auquel cas la langue ferait disparaître l'un des termes, devenant alors inutile. Dans les faits, il y a toujours une différence, aussi minime soit elle, entre des synonymes, qu'elle soit liée au signifié lui-même, aux connotations véhiculées (nuance méliorative, péjorative, laudative, etc.), au registre de langue ou encore au contexte d'emploi des mots. En règle générale, le lien de proximité réside dans la seule dénotation des termes liés [179].

page 116

système d'information L'ensemble des ordinateurs, des logiciels et tous les éléments du réseau d'un organisme tel une entreprise, i.e. tout ce qui lui permet de manipuler automatiquement de l'information [168].

page 35

Système d'information géographique un SIG permet de gérer des données alphanumériques spatialement localisées. En France, dans son acception courante, le terme fait référence aux outils logiciels. Cependant, le concept englobe l'ensemble constitué par les logiciels, les données, le matériel et les savoir-faires liés à l'utilisation de ces derniers. On peut aussi parler de Système d'information à référence spatiale (SIRS) pour les données et leur structuration. Le rôle du système d'information est de proposer une représentation plus ou moins réaliste de l'environnement spatial en se basant sur des primitives géographiques telles que des points, des arcs, des polygones (vecteurs) ou des maillages (raster). A ces primitives sont associées des informations qualitatives telles que la nature (route, voie ferrée, forêt...) ou toute autre information contextuelle [149].

page 36

télédétection la télédétection désigne, au sens large, la mesure ou l'acquisition d'informations par n'importe quel type d'instrument sur un objet ou un phénomène, par l'intermédiaire d'un instrument de mesure n'ayant pas de contact avec l'objet étudié (par exemple, d'un avion, d'un engin spatial, d'un satellite ou encore d'un bateau). On fait souvent appel à des instruments tels qu'appareils photographiques, laser, radar, sonar ou gravimètre.

La télédétection spatiale, dans le domaine de l'astronautique, est l'ensemble des connaissances et des techniques utilisées pour déterminer les caractéristiques de la surface et de l'atmosphère de la Terre ou d'une autre planète, par des mesures effectuées à partir d'un engin spatial évoluant à distance convenable de cette dernière (en anglais *remote sensing from space*) [149].

page 36

terme Unité signifiante constituée d'un mot (terme simple) ou de plusieurs mots (terme complexe) et qui désigne une notion de façon univoque à l'intérieur d'un domaine. Mot appartenant à un vocabulaire spécial notamment au vocabulaire scientifique [149].

page 117

thésaurus Vocabulaire normalisé et structuré de façon à éliminer les ambiguïtés du langage naturel utilisé pour l'analyse du contenu informatif des documents et la recherche automatisée de ces documents [149].

page 121

topologie Branche des mathématiques traitant des relations de voisinage qui s'établissent entre des figures géométriques, et qui ne sont pas altérées par la déformation des figures. Note(s) : L'application des principes de la topologie en géomatique assure la cohérence des données géométriques et facilite leur utilisation parce qu'elle élimine la redondance des points, lorsqu'il s'agit de former des lignes, et la redondance des lignes, lorsqu'il s'agit de former des polygones [149].

page 208

Unified Modeling Language UML (Unified Modeling Language) est un langage de description normé, c'est une formalisation très aboutie et non-propriétaire de la modélisation objet utilisée en génie logiciel. Il est l'accomplissement de la fusion des précédents langages de modélisation objet Booch, OMT, OOSE et Classe-Relation. Principalement issu des travaux de Grady Booch, James Rumbaugh, Ivar Jacobson et Philippe Desfray, UML est un standard défini par l'OMG (Object Management Group) [179].

page 91

vecteur Dans une représentation graphique, segment tracé dans une certaine direction, depuis un point de départ jusqu'à un point d'arrivée. Note(s) : Dans certains programmes graphiques, les vecteurs sont utilisés à la place des groupes de points (sur le papier) ou des pixels (à l'écran). Par opposition à raster [149].

page 89

vocabulaire contrôlé Dans un domaine préalablement défini (d'ordre scientifique, technique, professionnel ou autre, et en général pour une langue donnée), choix de termes sélectionnés, classés et indexés en vue de faciliter l'indexation, le stockage et la recherche des publications traitant des concepts apparentés à ces termes. Comme exemple de langage contrôlé, on peut mentionner le thésaurus. Synonyme(s) : vocabulaire normalisé [149].

page 64

vue Résultat d'une requête. Une vue ne permet selon certains que la consultation des informations, et pas leur modification, tandis que d'autres sources disent qu'une vue est un accès en lecture et écriture à une BD, mais sans connaître la structure interne de la base [168].

page 230

Web sémantique Le Web sémantique est une extension du World Wide Web permettant de publier, de consulter et, tout particulièrement, d'automatiser le traitement de connaissances précisément formalisées [179].

page 19

Web Service Service Web en français : services accessible via le web. Un service est composé d'un ensemble d'opérations (donner les cours de la bourse, exécuter une requête...), auxquelles on accède par HTTP(S), SMTP/POP/IMAP ou n'importe quelle technique web [168].

page 129

Mutualisation de données et de connaissances pour la Gestion Intégrée des Zones Côtières. Application au projet Syscolag

Résumé

Cette thèse s'inscrit dans le cadre du programme régional de recherche pluridisciplinaire Syscolag sur la gestion intégrée de la zone côtière (GIZC). Elle étudie la mise en place de méthodes génériques pour optimiser la gestion de l'information et de la connaissance dans les dispositifs de GIZC. Compte tenu de l'hétérogénéité des ressources informationnelles nécessaires et réparties chez des acteurs variés et de l'importance de l'information géographique dans ce domaine, nous proposons, en réponse à cette problématique, une solution qui s'appuie sur l'utilisation d'un service de métadonnées pour décrire et localiser l'information existante et d'un référentiel sémantique pour intégrer et partager la connaissance experte. Ces outils sont accessibles sur un portail Web. Le premier implémente la norme ISO 19115 (relative à la gestion de métadonnées pour l'information géographique), le second s'appuie sur un modèle d'ontologie a priori qui structure l'inventaire des concepts du domaine et exprime la connaissance qui leur est associée. Les concepts spatiaux possèdent des propriétés géométriques qui permettent leurs représentations cartographiques géoréférencées et des relations spatiales normalisées d'après les travaux de l'Open Gis Consortium. Le référentiel sémantique sert pour le contrôle de la valuation d'éléments clés du service de métadonnées, en particulier les éléments de descriptions thématique et spatiale (avec une interface cartographique). Le gain en qualité d'indexation améliore la localisation d'information. Le détail du référentiel est consultable sous la forme d'une base terminologique, d'un réseau sémantique et d'un atlas cartographique (basé la norme relative aux Web Mapping Service de l'OGC) dans le cas des concepts spatiaux. De tels systèmes distribués sont capables d'interopérer et de partager les métadonnées, l'information géographique, ou les concepts qu'ils hébergent.

Mots-clés : métadonnées, partage de l'information, partage de la connaissance, ontologie a priori, information géographique, gestion intégrée de la zone côtière, SIG.

Data and knowledge sharing for the Integrated Coastal Zone Management. Application to the Syscolag's Research Program.

Abstract

This Phd is a part of the Syscolag's research program and studies concepts and tools to set up in order to mutualize data and knowledge to serve integrated coastal zone management (ICZM). The main problems consist in the variety of information resources distributed by the different stakeholders as well as in the multidisciplinary framework which involves the setting up of a common vocabulary between the partners. This work propose a solution to minimize syntax and semantic interoperabilities problems with an additional difficulty involved by the spatial dimension of information concerned. The goal of this PhD is the setting up of generic methods to improve information and knowledge management in the ICZM context. The proposed solution consists in a mutualization infrastructure which core is made of a metadata management service to describe and locate existing information as well as a semantic referential to integrate and share expert knowledge. For the metadata service, the ISO 19115 standard (adapted to spatial information description) has been chosen. A specific profile of the ISO 19115 has been implemented for the specific Syscolag's context and has been used differently among the different kind of information resources. The thematic referential comes from an a priori ontology model which structures the concepts inventory of this domain and restitutes the knowledge which is related to. This model has been created by an ecosystemic approach. Spatial dimension has been used at the concepts and relations level as well as for the maps representation (OGC recommendations compliant). The whole solution can be accessed by a Web portal.

Keywords : metadata, knowledge sharing, information sharing, ontologies, geographic information, Integrated Coastal Zone Management, GIS.