



HAL
open science

Vers la Diffusion en Ligne d'Information Géographique sur Mesure

Bénédicte Bucher

► **To cite this version:**

Bénédicte Bucher. Vers la Diffusion en Ligne d'Information Géographique sur Mesure. Web. Université Paris Est Marne la Vallée, 2009. <tel-02474464>

HAL Id: tel-02474464

<https://theses.hal.science/tel-02474464v1>

Submitted on 11 Feb 2020

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.



HAL Authorization



Université Paris-Est Marne-la-Vallée

Habilitation à diriger des recherches
Spécialité : informatique
et sciences de l'information géographique

Vers la Diffusion en Ligne d'Information Géographique sur Mesure

Bénédicte BUCHER
Laboratoire COGIT, Institut Géographique National

Habilitation soutenue le 17 décembre 2009
à l'Institut Géographique National
devant un jury composé de :

Anne RUAS	directrice d'habilitation	IGN-COGIT
Professeur Anne DOUCET	rapporteur	Université Paris VI
Professeur Thérèse LIBOUREL	examinatrice	Université de Montpellier II
Gabriella SALZANO	examinatrice	Université Paris-Est Marne la Vallée
Professeur Christophe CLARAMUNT	rapporteur	IRENav
Professeur Hervé MARTIN	rapporteur	Université de Grenoble I

A ma famille,

TABLE DES MATIERES

INTRODUCTION	7
1 LES INFRASTRUCTURES D'INFORMATION GEOGRAPHIQUE	13
1.1 Les infrastructures de données géographiques	14
1.1.1 Généralités	14
1.1.2 Des infrastructures orienté-produits aux infrastructures orienté- processus	15
1.1.3 Les défis des infrastructures de données géographiques	18
1.1.4 La transformation de schéma	23
1.1.5 Métadonnées et recherche d'information	30
1.1.6 La visualisation de métadonnées géographiques	33
1.1.7 'Collages spatiaux' sur les infrastructures d'information géographique ..	36
1.1.7.1 L'importance de la navigation entre plusieurs niveaux de détails	36
1.1.7.2 Propositions pour l'aide à l'exploration de plusieurs niveaux de détails ...	37
1.2 Le GéoWeb	42
1.2.1 Les moteurs de recherche géographique	42
1.2.1.1 Généralités	42
1.2.1.2 SPIRIT : l'architecture	43
1.2.1.3 SPIRIT : analyse du besoin	46
1.2.1.4 SPIRIT : évaluation	48
1.2.2 Web sémantique	51
1.2.2.1 Langages du Web sémantique	52
1.2.2.2 Linked Data	56
1.2.3 Le projet TerraNumerica	58
1.2.3.1 Glossaire	59
1.2.3.2 Un processus générique d'enrichissement	60
1.2.3.3 Référencement dans les métadonnées	62
1.2.3.4 Mise en cohérence des données	64
1.2.3.5 Création des liens (références internes ou externes)	64
1.2.4 Les projets dédiés à la création collaborative de contenu	66
1.3 Bilan	72
2 LA MISE EN LIGNE DE PROCESSUS ADAPTABLES A UN CONTEXTE	75
2.1 Le cas général	76
2.1.1 Généralités sur la description formelle de processus	76
2.1.2 Modélisation de tâches	78
2.1.2.1 Un modèle dédié au développement collaboratif d'applications de traitement d'image	78
2.1.2.2 TAGE : un modèle de tâches géographiques pour la spécification coopérative d'utilisations de données géographiques	79
2.1.3 Les services	88
2.1.3.1 Généralités sur les services Web	88
2.1.3.2 Les services géographiques	91

2.1.3.3	OPERA : une plate-forme sémantique pour des services géographiques	95
2.1.3.4	MOBIDIC: une plate-forme pour le développement de services adaptables à des contextes mobiles	97
2.1.3.5	WebGen : des services pour faire avancer la recherche au sein d'une communauté	99
2.1.4	Le catalogage et le partage de ressources logicielles	101
2.1.4.1	MDT : un modèle de métadonnées pour le catalogage de ressources logicielles	101
2.1.4.2	Synthèse des enseignements liés à TAGE et MDT	108
2.1.4.3	MFM : Un modèle de métadonnées orienté-services pour le catalogage de ressources logicielles et le déploiement de services web pertinents et utilisables	111
2.1.4.4	Découverte et exploration à distance de traitements implémentés dans GeOxygène fondées sur MFM	117
2.1.4.5	Une version documentaire de MFM dédiée à la constitution d'un vocabulaire de description	122
2.2	L'aide à l'accès à des cartes topographiques sur mesure en ligne	127
2.2.1	Contexte	127
2.2.1.1	Les nouveaux rédacteurs de cartes	127
2.2.1.2	La pertinence des approches centrées utilisateurs pour la conception d'interfaces cartographiques	130
2.2.1.3	Critères d'adaptation d'une carte topographique à un contexte	131
2.2.2	Modèle général des travaux sur la carte à la carte	133
2.2.2.1	Modèle SISSI de spécification d'une carte	133
2.2.2.2	La prise en compte de l'impact émotionnel	136
2.2.3	Des services cartographiques additionnels dans les architectures ISO/OGC de Web mapping	142
2.2.3.1	Approche méthodologique	142
2.2.3.2	Services proposés	145
2.2.3.3	Un répertoire de styles et de couches stylisées pour l'aide à la création de carte	147
2.2.4	Un dialogue pour la création collaborative des styles d'une carte	153
2.2.4.1	Motivation et travail préliminaire	153
2.2.4.2	Une méthode de conception coopérative de légende	156
2.3	Bilan	163
BILAN ET PERSPECTIVES		165
BIBLIOGRAPHIE		171

Introduction

L'information géographique diffusée en ligne peut se présenter sous la forme de d'une carte topographique centrée sur la position courante de l'utilisateur et s'affichant sur un terminal nomade. Elle peut se présenter sous la forme de données sur l'état du trafic routier sur un réseau urbain et le positionnement des bus, données transmises à un programme client qui calcule les temps d'attente à chaque arrêt de bus. Elle peut aussi se présenter sous la forme d'un globe virtuel permettant à un citoyen de prendre conscience de l'ampleur des dégâts causés par un feu de forêt, ou encore sous la forme d'une alerte éditée sur le Web par un citoyen pour signaler une erreur de toponyme dans une carte topographique. Ces exemples visent à montrer que l'information qui nous intéresse est variée de bien des façons: diversité des phénomènes géographiques étudiés, des modes de représentation de ces phénomènes d'intérêt, de profils d'utilisateurs, de contextes d'utilisation, de tâches effectuées sur l'information (stockage, transfert, requête, analyse, correction ou construction). Actuellement, les évolutions techniques permettent d'envisager une diffusion d'information géographique en ligne qui s'adapte à cette diversité d'usages. Nous nous intéressons dans ce mémoire à la diffusion d'information géographique sur mesure. **L'information géographique est une représentation (graphique ou non) de phénomènes situés dans l'espace géographique. Nous voulons la diffuser de façon à supporter le plus facilement possible une activité du destinataire (humain ou logiciel). Nous visons la diffusion sur mesure, c'est-à-dire deux cas de figures :**

- **L'information géographique souhaitée n'existe pas et doit être construite sur mesure pour répondre au besoin. Cela consiste par exemple à intégrer plusieurs jeux, à transformer le schéma ou encore à changer le niveau de détail.**
- **L'information géographique souhaitée existe mais est difficile d'accès et c'est l'interface de consultation qui doit s'adapter le plus possible au besoin. Cela consiste par exemple à associer des taxonomies compréhensibles pour l'utilisateur pour rechercher des objets d'intérêts dans des catalogues de données géographiques.**

Une représentation privilégiée de l'espace géographique est la carte. La diffusion de carte par l'Institut Géographique National se décline en la diffusion de produits en série couvrant les principales échelles des activités d'un pays, la diffusion de produits spéciaux correspondant à un thème spécifique (carte du débarquement, carte des vins du bordelais, ...) et la réalisation de produits à façon sur commande (carte d'un parc

régional par exemple). Une diffusion sur mesure consisterait, par rapport à ces trois activités de l'IGN, à sélectionner le plus facilement possible les produits sur étagère correspondant au besoin (et à ouvrir la carte au bon pli), à faciliter en interne le processus de conception d'une carte adaptée à un thème donné (que ce soit un domaine ou un événement) et à réaliser le plus automatiquement possible une carte sur commande. Il est également important de prendre en compte de nouveaux besoins. La diffusion est maintenant fortement liée à la contribution et diffuser une carte en ligne signifie aussi mettre à disposition de professionnels ou de citoyens une information géographique qui soit éditable ou qui supporte une activité de création de contenu localisé (en tant que données de référence).

Un premier aspect qu'il me semble important de souligner dans cette introduction est que **la diffusion en ligne d'information géographique est en partie un enjeu public**. S'assurer que l'information à laquelle doit légitimement avoir accès un citoyen ou expert d'un domaine lui est délivrée de façon adaptée à l'usage qu'il doit en faire est un enjeu public. La notion de Digital Earth a été introduite par Al Gore, alors vice-président des Etats-Unis, pour désigner une représentation numérique de notre terre qui puisse répondre aux requêtes de tous concernant des phénomènes localisés dans l'espace et le temps. Un domaine clé dans lequel cette notion a été beaucoup reprise est celui du **développement durable** : la compréhension des phénomènes, le contrôle des politiques et la modification des pratiques passent par un meilleur accès à une meilleure information. Ce domaine a fortement dynamisé la diffusion d'information géographique légale. En Europe, la directive INSPIRE, portée par la Direction de l'Environnement de l'Union Européenne, demande la mise en place d'infrastructures d'information géographique pour faciliter l'accès aux données légales localisées à des fins de contrôle des politiques communautaires environnementales. Les sous-domaines prioritaires de cette Direction ont tous un fort aspect localisé : le changement climatique, la nature et la biodiversité, l'environnement et la santé, ainsi que la qualité de la vie, les ressources naturelles et les déchets. Un autre enjeu public est la **gestion du risque**, domaine très demandeur d'information géographique. Au niveau international, après l'ouragan Katrina, une charte internationale 'Espace et Catastrophes Majeures' a été signée par plusieurs agences spatiales pour fournir des données les plus actuelles possibles aux organismes chargés des secours en cas de crise (<http://www.disasterschartercatalog.org>). Par ailleurs un nouvel enjeu est la gouvernance des territoires et la démocratie participatives. L'évolution des technologies de communication de l'information Web ont permis de repenser les relations entre le citoyen et l'Etat ainsi qu'entre organismes de l'Etat. La **cyber-administration** vise à faciliter la communication, l'accès à certaines informations administratives et l'accomplissement de certaines tâches (déclaration d'impôts). La cyber-administration peut souvent s'appuyer sur l'information géographique. La Direction Générale pour la Modernisation de l'Etat a ainsi œuvré pour la mise en place d'un portail national facilitant l'accès en ligne –pour des services d'états puis des citoyens- à des données géographiques publiques: le Géoportail. D'autres portails peuvent également se concentrer sur des zones géographiques comme les régions. En Angleterre, une

étude¹ a montré que les citoyens attendaient de nouveaux services en ligne de leur administration et les exemples de sites indiqués par les citoyens sondés s'appuient majoritairement sur des données localisées. Les sites Web informant les citoyens sur des projets (comme la réforme de l'enseignement ou l'aménagement du quartier des Halles) peuvent aussi servir à recueillir des contributions de citoyens et à susciter un débat collectif –chaque citoyen peut lire la contribution d'un autre et y réagir -. (Rinner et al. 2008) ont par exemple conduit une étude concluante sur l'apport d'outils web cartographiques participatifs pour l'aménagement d'un campus. C'est un domaine de recherche actif, désigné par le terme Public Participatory GIS (PPGIS) proposé dans l'ouvrage fondateur de (Jankowski et Nyerges 2001).

Un autre aspect à souligner est le suivant: **les évolutions technologiques de l'information géographique numérique et des réseaux de communication bouleversent les relations entre les défis techniques et les acteurs susceptibles de les relever**. Une de ces évolutions est **le Web**. Alors qu'une démarche phare en géomatique est celle de l'Open Geospatial Consortium, le grand public retient plutôt comme avancée majeure dans ce domaine le produit Google Earth qui ne s'appuie pas sur les spécifications OGC mais sur les techniques et paradigmes du Web. Mike Jackson, professeur de l'Université de Nottingham et fondateur du constructeur de logiciel SIG Laserscan (actuel OneSpatial), va jusqu'à dire que les vendeurs traditionnels de logiciels SIG ont 'raté' le virage du Web et laissé ainsi la place libre pour des acteurs comme map24 et Google. Le **Web 2.0** a facilité l'apparition de nouveaux modes de production, illustrés par la communauté OpenStreetMap. Actuellement, l'essor du **Web sémantique** correspond à une architecture distribuée de contenu différente de celle proposée dans les modèles en vogue en géomatique. Les **terminaux nomades et les capteurs** sont également une évolution importante. « Location-based mobile services delivered to your phone or Blackberry are predicted to be the next "next SMS" in terms of impact on social networking » Andrew Coote, Assessing the UK Market for GI dans GIS Professionnal, février 2009. Ce domaine rejoint peu à peu celui de **l'informatique ubiquitaire**. Il s'agit d'une évolution des architectures distribuées dans laquelle des programmes sont accessibles en tout lieu et en tout temps. Elle est dynamisée par l'essor des réseaux de positionnement et des réseaux de communication. Dans un tel contexte, il est important d'alléger la charge cognitive de l'utilisateur pour lui permettre de se concentrer sur son activité courante (comme conduire). Une autre évolution, de pratique celle-là, est **le libre**. Un nouveau type d'acteur se positionne sur le domaine de la diffusion dont le modèle économique est souvent la 'gratuité' : les communautés du libre. Il s'agit des contributeurs à des logiciels libres (Postgresql, MapServer, OpenServer, ..) et à des données libres (Un Point C'est Tout, OpenStreetMap). Cet

¹ <http://www.showusabetterway.co.uk/> ...Tell us what you'd build with public information and we could help fund your idea!

acteur entretient des liens complexes avec les précédents : les administrations utilisent et contribuent au libre, les entreprises privées aussi.

Je viens de souligner deux aspects de la diffusion en ligne d'information géographique : le fait que ce soit en partie un enjeu public et le fait que ce domaine subit de fortes évolutions technologiques qui changent le positionnement des acteurs. Dans ce contexte, **une tâche difficile et nécessaire est de tenter de cerner les rôles des uns et des autres** et de savoir par exemple d'une activité si elle est du ressort du service public ou d'une entreprise à but lucratif, si elle doit être relevée par un expert en géomatique ou par un expert d'un autre domaine. Répondre aux attentes de plus en plus exigeantes et impérieuses des utilisateurs de réseaux d'information et de terminaux nomades est devenue une industrie prospère là où le besoin et les moyens actuels se rapprochent suffisamment. Les vendeurs de logiciels spécifiquement dédiés à la manipulation de données géographiques ont proposé des solutions pour construire des serveurs de cartes ou de données, souvent des extensions Web de leurs produits desktop. D'autres entreprises privées, venant de l'informatique (Microsoft) et du Web (Google) ont investi ce domaine, parfois plus rapidement. L'importance de ce marché est telle qu'un consortium a pu se créer autour de sa dynamisation : l'Open Geospatial Consortium. La place du service public doit être déterminée en fonction de ce marché comme le dit (Didier 90) : 'la valeur d'une carte, et plus généralement d'un produit ou d'un système d'information géographique nouveau est l'ensemble des dispositions à payer des utilisateurs. L'Etat prend utilement en charge le ou les actions dont la valeur est supérieure au coût, et que l'initiative privée n'est pas en mesure d'assurer.'

A l'IGN, la problématique de la diffusion en ligne était traitée il y a dix ans dans un projet dédié à la diffusion en ligne en interne (projet Diffusion des Données Numériques dirigé par Didier Richard). Puis le groupe GAQ (Groupe pour l'Acquisition de la Qualité) a beaucoup œuvré pour organiser l'acquisition des métadonnées, qui sont des données nécessaires à la diffusion. Par ailleurs, au Service de la Recherche de l'IGN, une action de recherche du laboratoire COGIT sur le thème de l'interopérabilité était créée en 1997 sous la responsabilité de François Chirié. Cette action participe à des projets européens comme Commuter sur la mise à jour distante de bases de données distribuées ou Geoserver, un catalogue de jeux de données géographiques provenant de divers pays. Au sein de cette action, des travaux s'attachent aussi à l'accès utilisateurs à des données géographiques. Ma thèse propose de coupler l'accès aux données avec un accès à des traitements pour faciliter l'expression du besoin (Bucher 2002). Cette proposition veut permettre à l'utilisateur de spécifier un besoin d'application et non un besoin de données. Elle s'appuie sur des techniques de représentation des connaissances mais ne prend pas en compte les possibilités actuelles des services Web. Par la suite, j'ai dirigé plusieurs actions de recherche successives sur le thème de la consultation. Ces actions ont cherché à conduire des propositions en étant suffisamment indépendant et complémentaire à la fois des autres travaux de l'IGN sur ce domaine. Au fil du temps, nous avons ainsi identifié des 'niches durables' de recherche.

Des étapes décisives dans la diffusion en ligne d'information géographique ont été Google Earth puis, en France, la mise en place du **Géoportail** en 2006. Le Géoportail a été commandé par la Direction Générale de la Modernisation de l'Etat en vue de fournir à la France un portail national pour l'accès aux données localisées et leur visualisation. Sa maîtrise d'œuvre est assurée par l'IGN et le BRGM (Bureau des Recherches Géologiques et Minières). Aujourd'hui, il semble que l'IGN peut jouer un rôle important dans l'émergence d'une infrastructure d'information où le spatial joue son rôle d'intégrateur et de facilitateur de découverte et d'analyse. Pour que cette infrastructure fonctionne, des **verrous scientifiques** devront être levés par la communauté géomaticienne dans son ensemble. On peut espérer que la décision prise par le Directeur Général de l'IGN de permettre à des chercheurs de laboratoires publics d'accéder gratuitement à des données géographiques de référence pour leurs travaux accélère les progrès de cette communauté, progrès nécessaires pour répondre aux défis du développement durable. Dans ce domaine, un acteur essentiel est également le **Groupement de Recherche pluridisciplinaire MAGIS** dirigé par Christophe Claramunt, directeur de l'Institut de Recherche de l'Ecole Navale, et Thierry Joliveau, professeur à l'Université de Saint Etienne. Ce groupement succède à Cassini et SIGMA et réunit des laboratoires en géographie et en informatique.

Ce mémoire est rédigé en vue d'obtenir l'habilitation à diriger des recherches en sciences de l'information géographique. Il traite de la problématique de la diffusion d'information géographique en ligne. Une première partie présente la notion d'infrastructure d'information géographique. C'est un domaine qui est devenu de plus en plus vaste au fil des ans avec par exemple la naissance d'une communauté de recherche autour des infrastructures de données géographiques (Spatial Data Infrastructures) et la naissance d'une communauté sur la recherche d'information géographique (Geographical Information Retrieval). Etant responsable des travaux de recherche sur l'accès en ligne à l'information géographique, j'ai tenté d'explorer ce domaine en largeur grâce à des collaborations, à des travaux de stage et à une veille scientifique. Cependant cette première partie du mémoire est loin de couvrir cette problématique dans son ensemble. Ainsi, j'ai fait le choix de ne pas détailler certains aspects techniques (comme les architectures internet, services localisés, modèles standard ISO/OGC). De plus, les problématiques d'intégration et de représentation multiple sont du ressort d'une autre action de recherche du COGIT dirigée par Sébastien Mustière. La partie suivante présente notre principale contribution au domaine de la diffusion, qui correspond à un cas particulièrement important pour l'Institut Géographique National : la diffusion de processus complexes sur mesure. Ce sujet a été exploré en profondeur grâce à des travaux de thèses successifs que j'ai co-encadrés et à mes travaux de recherche. L'objectif général est de mettre en ligne des processus de manipulation de données géographiques pour qu'ils puissent être invoqués dans des contextes dans lesquels ils sont pertinents. Il y a quelques années, on lisait couramment que la manipulation de données géographiques nécessitait des outils complexes et

coûteux et une forte expertise. C'est toujours vrai sauf que ces outils et cette expertise peuvent être hébergés sur un serveur et exploitées par des utilisateurs sans qu'ils doivent les posséder. Un processus spécifique que nous avons particulièrement étudié est la rédaction d'une carte sur mesure.

1 Les infrastructures d'information géographique

Cette partie présente les principales solutions pour diffuser l'information géographique sur le Web.

Son titre se rapproche de l'expression 'infrastructure de données géographiques' qui est une appellation consacrée en géomatique pour désigner les architectures de médiation dédiées à faciliter l'accès d'utilisateurs à des données géographiques provenant de sources variées. Ces solutions sont présentées en section 1.1.

Si le titre se différencie du terme consacré c'est que nous étudierons aussi en section 1.2 des solutions provenant d'une autre communauté, celle du Web dont les objectifs visent à favoriser la diffusion d'information, et en particulier l'information géographique. Ces objectifs rejoignent donc ceux des infrastructures de données géographiques.

1.1 Les infrastructures de données géographiques

1.1.1 Généralités

Rappelons qu'**une application utilisant des données géographiques réutilise le plus souvent des données existantes**. Il faut donc à son concepteur identifier les données d'intérêt, les acquérir et éventuellement les intégrer. Le concepteur rencontre plusieurs obstacles. Certains obstacles sont d'ordre financier. Il est entendu que le coût d'acquisition de données géographiques dépasse le prix que peuvent payer la plupart des utilisateurs et qu'il est donc nécessaire que ce coût soit absorbé par un intermédiaire (Didier 1990). Cela peut être une entreprise privée, comme Google, dont le modèle économique s'accommode de l'acquisition de ressources qu'il rediffuse gratuitement. Cela peut être l'Etat, comme c'est le cas pour les données de l'USGS (United State Geodetic Survey). Dans de nombreux cas, comme en France, l'Etat finance en partie le coût d'acquisition de données géographiques et demande à l'organisme producteur de récupérer lui-même un reste à payer, non négligeable. D'autres obstacles sont des obstacles techniques liés en particulier à l'intégration de données : mise en cohérence de données provenant de diverses sources ou adaptation de données à l'application.

L'effort de résolution des obstacles mentionnés ci-avant peut être mutualisé. Plusieurs utilisateurs de données peuvent se regrouper, au sein d'une *'table ronde'*, pour acquérir ensemble des données qu'ils se partagent ensuite. Plusieurs producteurs de données peuvent mettre en place, ensemble, des mécanismes d'intégration ou de co-visualisation qui facilitent l'utilisation conjointe de leurs données respectives. **Le terme 'infrastructure de données géographique' désigne de façon générale des architectures de médiation entre des utilisateurs et des données géographiques fournies par des sources diverses**, comme illustré sur la Figure 1. Les principes clés de ces architectures reprennent les propositions du domaine des bases de données qui ne seront pas exposées ici (Gardarin 2000)(Pucheral 2001)(Doucet et Gançarski 2001)(Özsu and Valduriez 1999).

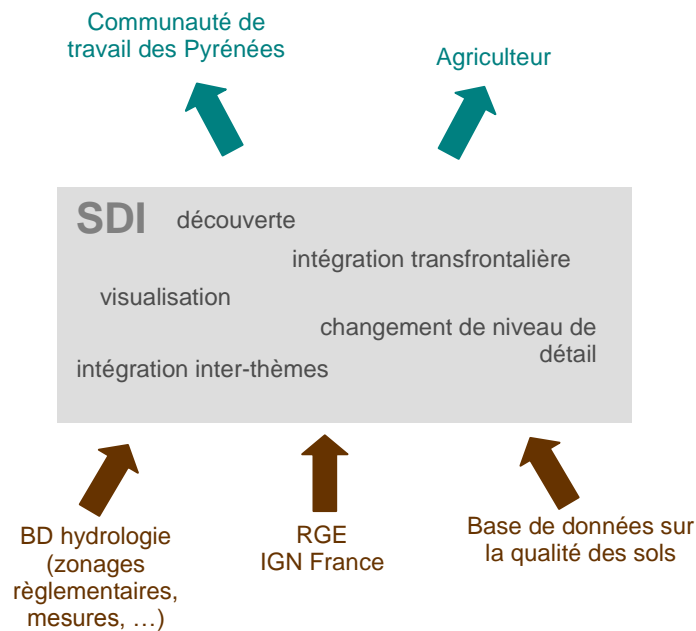


Figure 1. Une infrastructure de données géographiques (SDI) est une solution de médiation entre des utilisateurs ayant besoin de données géographiques et des données géographiques fournies par des sources diverses.

1.1.2 Des infrastructures orienté-produits aux infrastructures orienté-processus

(Masser 1999) distingue deux générations de SDI :

- la première génération est centrée sur l'amélioration de la diffusion de produits existants.
- la deuxième génération est centrée sur l'amélioration des processus de décision que la SDI vise à supporter.

La première génération des SDI date des années 80. Elle correspond au besoin de faciliter le développement économique, la gestion de l'environnement et la gouvernance des territoires en facilitant l'intégration de données existantes (Masser 1998). En Europe, un précurseur de telles SDI est l'ensemble des spécifications de produits continus et harmonisés sur l'Europe proposées par l'association Eurogeographics, Association (loi 1901) qui regroupe 52 organismes nationaux de cartographie ou cadastre de 43 pays européens. Les spécifications Eurogeographics sont par exemple celles des produits EuroGeonames, EuroGlobalMap, EuroBoundaries. Comme le soulignent (Williamson et

al. 2003) ces SDI visaient à éviter la duplication des données, à amplifier l'usage des données existantes et à dynamiser le marché de l'information géographique et ainsi que la productivité de cette industrie. Cela explique que ces solutions aient toutes privilégié une approche consistant à spécifier a priori les spécifications d'un ou plusieurs produits qui puissent être obtenus à partir des diverses sources. La plupart des auteurs s'accordent, pour cette raison, à qualifier ces premières SDI d'**orientées produit**. Cette approche est légitime dès lors qu'une SDI s'appuie sur des produits existants. Le principe est que chaque fournisseur héberge et maintient ses données. La donnée servie par la SDI est créée plus ou moins dynamiquement en exploitant ces diverses ressources. Par exemple, le projet European Spatial Data Infrastructure with a Best Practice Network (ESDIN, <http://www.esdin.eu>) vise à apporter davantage d'automatisation dans le processus d'alimentation des produits fédérés Eurogeographics par les organismes nationaux. Des défis techniques auxquels ESDIN cherche à répondre sont la transformation automatique de schéma, le recalage automatique des données aux frontières (limites administratives et réseaux).

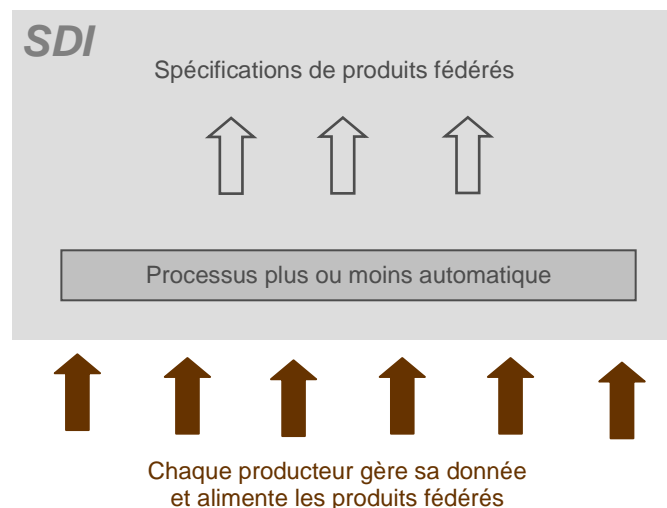


Figure 2. Une SDI de première génération, 'orientée-produits'.

Depuis 2007, la **directive INSPIRE** 2007/2/CE vise à mettre en place une infrastructure de données localisées destinée au contrôle des politiques environnementales communautaires. Il s'agit d'une proposition techniquement rattachée aux SDI de première génération. Des spécifications de produits fédérés INSPIRE sont en cours de définition. Les travaux actuels portent sur les spécifications des services nécessaires à l'infrastructure INSPIRE : services de catalogages, d'intégration (y compris transformation de schéma, généralisation, recalage) et de visualisation.

Un autre exemple de SDI de première génération est le projet Gimodig dédié à la visualisation d'information topographique transfrontalière (information générée à partir des sources des agences cartographiques nationales) sur un téléphone portable. L'architecture GiMoDig est décrite par (Lehto 2003) et reprise de façon schématique sur la Figure 3. C'est une fédération de schémas (Illert and Afflerbach 2003). Le schéma global, GiMoDig.xsd, est écrit en format GML. Chaque agence de cartographie définit un schéma local dans lequel elle fournit ses données à un médiateur GiMoDig. Ces schémas locaux sont différents du schéma global mais correspondent à des vues de leurs données plus simples à maintenir pour chaque agence. Du point de vue du médiateur, le fait d'introduire des schémas locaux simplifie par la suite l'intégration des données dans la vue globale, intégration faite à la volée. L'intégration repose sur des services de transformation de coordonnées et des services de transformation de schéma. Par la suite les données GiMoDig sont traitées pour être envoyées sur des téléphones mobiles, par des services de généralisation. Une dernière couche est celle des services de requête et de transfert vers les clients nomades.

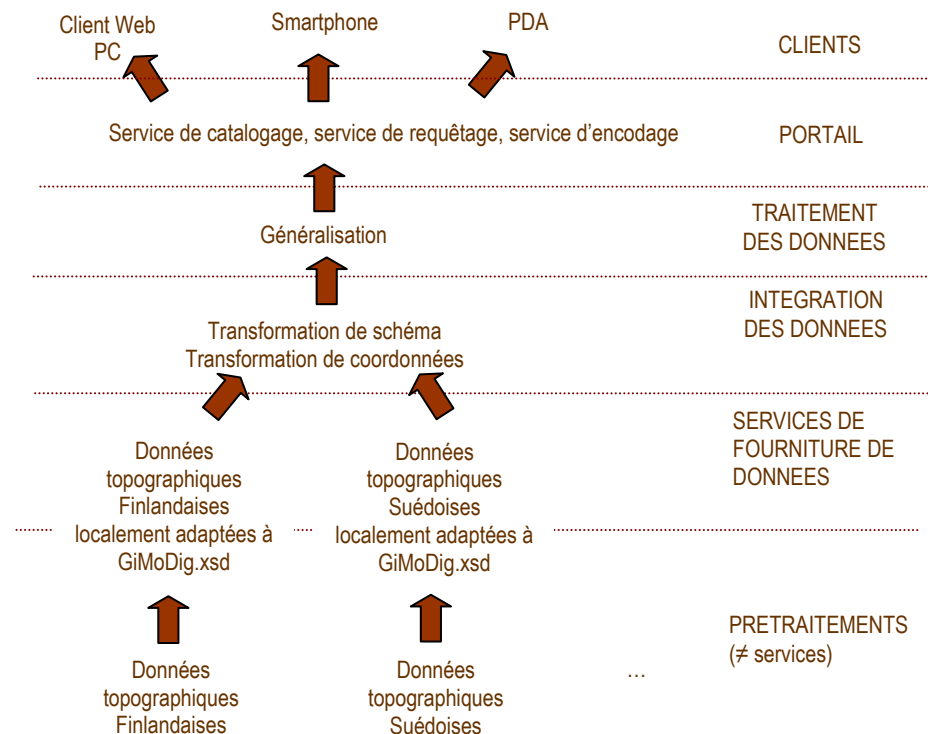


Figure 3. Architecture GiMoDig.

Un des principaux apports des travaux sur les SDI de première génération est d'avoir permis de prendre du recul sur ce besoin de médiation et de mieux en appréhender les

défis. Ainsi, la deuxième génération des SDI est marquée par l'apparition d'une communauté internationale dédiée à cette problématique où se font les échanges d'expériences et où se construisent des nouveaux modèles qui itèrent sur l'expérience des SDI de première génération. La spécificité de la 2^e génération consiste à adopter un modèle '**orienté processus**' pour compléter le modèle 'orienté produit' préconisé dans les approches de 1^{ère} génération. Au lieu (ou en plus) de s'attacher à la qualité des produits fournis par la SDI, il est préconisé d'étudier la qualité des processus de prise de décision qui s'appuient sur la SDI. Le projet européen Humboldt s'attache par exemple à étudier la réduction des coûts que permet une SDI sur un processus final de prise de décision. La section suivante énumère les principaux défis que doivent relever les SDI actuelles.

1.1.3 Les défis des infrastructures de données géographiques

Les défis que doivent relever les SDI sont nombreux. Certains datent des SDI de 1^{ière} génération et d'autres sont apparus avec les nouvelles ambitions des SDI de 2^{ième} génération. Ils sont résumés sur la Figure 4 et détaillés dans cette partie.

1) Un élément incontournable et qui a reçu toute l'attention des premiers travaux est l'ensemble des **modèles standard et des méthodes de catalogage et intégration** qui permettront techniquement de connecter les usages aux données. Les propositions du comité technique TC211 de l'organisme International Standardisation Organisation ainsi que celles du consortium OpenGeospatial (que nous appellerons modèles ISO/OGC) sont un atout de taille pour la mise en place de SDI. Il ne s'agit pas de modèles standard pour décrire l'espace géographique mais plutôt de normes de méta-modèles portant sur la description de la nature des objets, la description du positionnement. Les SDI de première génération s'appuient sur ces modèles pour définir des schémas fédérés (par exemple les spécifications INSPIRE). Les SDI de deuxième génération s'appuient sur ces modèles pour conduire plus facilement des fonctions de catalogage, de visualisation ou d'intégration. Actuellement, les besoins d'interopérabilité se situent moins au niveau du schéma logique ou de l'encodage que du niveau conceptuel. La plupart des logiciels peuvent afficher des géométries de n'importe quel objet modélisé comme un Feature par contre aucun logiciel ne peut comprendre que l'objet de type `Tron_Hydro_FeatureType` est un élément de réseau hydrographique et ne doit pas être utilisé dans un calcul d'itinéraire routier. Au laboratoire COGIT, les problématiques d'intégration et de généralisation sont traitées dans des actions de recherche spécifique. Celles-ci se sont dédiées en priorité à des besoins de production IGN mais s'ouvrent actuellement à des besoins typiques des infrastructures géographiques. Par exemple, une thèse a commencé en 2007 sur la mise en ligne de services d'intégration pour aider à intégrer

des données anciennes sur le RGE (thèse d'Eric Grosso sous la direction d'Alain Bouju). En généralisation également, des chercheurs ont participé à une réflexion sur la mise en place de services web de généralisation. Enfin, une étude ponctuelle a été faite sur la généralisation de bâtiments en 3D pour faciliter la transmission de données en les comprimant. Nous avons aussi étudié **la transformation de schéma**. Ces travaux sont présentés dans la section 1.1.4.

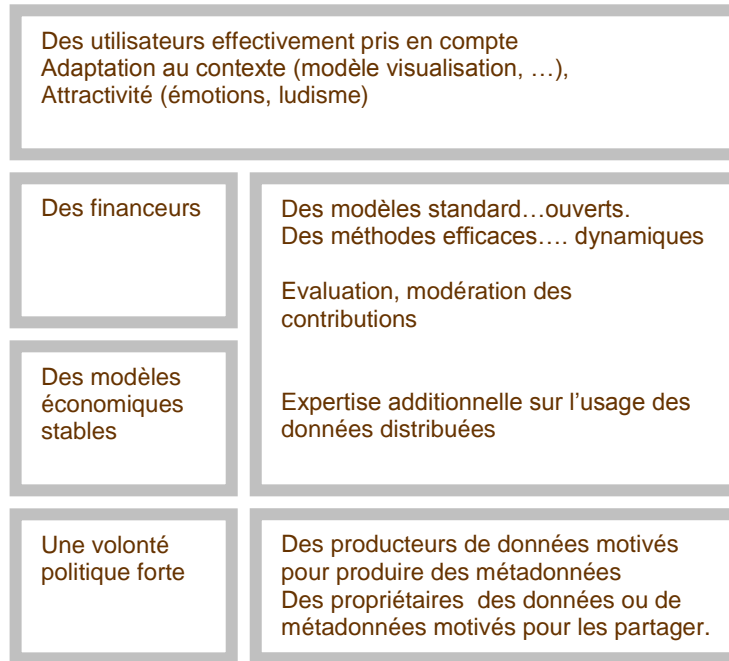


Figure 4. Défis des SDI.

2) Un autre défi consiste à **intégrer davantage les utilisateurs** dans le projet d'une SDI. Quand bien même le concept de 'communauté d'utilisateurs' est présent dans le modèle théorique de ces architectures, les cas d'utilisation avancés ne proviennent pas toujours d'"utilisateurs réels". Les utilisateurs 'finaux réels' sont aptes à évaluer et à réagir face à des propositions relativement abouties et complètes où tous leurs besoins sont traités. Lorsque la recherche explore des nouveaux domaines, il faudrait pouvoir demander à l'utilisateur testeur de faire abstraction de limites des prototypes qui correspondent à des aspects non traités (fluidité, rapidité, ergonomie, jeux tests restreints, cas d'utilisation restreints...). Dans le contexte du projet européen **SPIRIT** (plus spécifiquement sections 1.2.1.3 et 1.2.1.4) nous avons dû proposer des solutions à ces problèmes. Dans notre proposition sur l'aide à la construction de légende, nous avons voulu supporter des **processus de conception de légende fondés sur l'analyse du besoin** (section 2.2.1.2) et d'autre part nous avons également voulu adapter le processus au rédacteur de la carte grâce à une **méthode coopérative de conception et à une application de dialogue** (section 2.2.4).

3) Les ambitions des SDI de deuxième génération apporte un nouveau défi : proposer des **méthodes de catalogage et d'intégration génériques et dynamiques**. Tout d'abord les producteurs de données ne sont pas tous connus à l'avance. De plus les produits pertinents ne sont pas non plus tous connus à l'avance et l'infrastructure ne peut se contenter de donner accès à des produits fédérés figés, elle doit permettre à l'utilisateur de spécifier dynamiquement le produit dont il a besoin et déclencher des processus correspondant de transformation de schéma ou de généralisation par exemple. La proposition d'aide à la **transformation de schéma** de Sandrine Balley, mentionnée précédemment et présentée en section 1.1.4 vise la dynamique. Par ailleurs, nos travaux sur le **catalogage de traitements** veulent faciliter la spécification de nouveaux processus de traitements adaptés à un contexte et déclenchables en lignes. Ils sont présentés en section 2.1.4.

4) Ces mêmes ambitions demandent également des **modèles plus ouverts**. Par exemple, les modèles complexes d'INSPIRE ne sont pas accessibles à la plupart des producteurs de contenu et des utilisateurs. Construire une SDI de 2^{ème} génération couplée à INSPIRE nécessitera la mise en place de vues sur ces modèles complexes pour faciliter le développement d'applications alimentées par les données INSPIRE mais conçues par des personnes non nécessairement familières avec les techniques ISO/OGC préconisées dans INSPIRE. Comme le souligne (Singh 2004), les modèles OGC reflètent une vision experte des données géographiques difficile d'accès et de plus fortement marquée par les modèles des logiciels cartographiques. Le projet OpenStreetMap, pure produit du Web 2.0, a intégré dans sa Foire Aux Questions la question *'pourquoi ne pas utiliser les outils et formats OGC ?'* avec la réponse suivante : *'There are many existing tools which people often recommend for OpenStreetMap to use. The problem is, most of them are hard to use and maintain for a variety of reasons, and people are very reluctant to volunteer to help set them up and run them. We used to use MapServer for serving static versions of our maps, but unfortunately we found it to be unthreaded, slow and hard to extend - we replaced it with mapnik. It's not that OpenStreetMap as a community is against OGC standards, but OpenStreetMap has been built iteratively using the simplest approach that could possibly generate useful maps. The focus is on street data and maintaining a 'wiki' approach to editing where all changes are logged and can be rolled back. Off the shelf tools don't support that in the way we'd like.'* Le fondateur du logiciel mapnik utilisé par le projet OpenStreetMap précise sur son site qu'un atout de son projet est : *'[It] doesn't suffer from design decisions made a decade ago.'* Contrairement aux acteurs légaux, ces utilisateurs ne sont pas tenus d'adopter les modèles ISO/OGC. Dans le cadre du **projet TerraNumerica**, nous avons proposé une architecture intégrant des composants du Web sémantique avec des composants plus spécifiques à la géomatique (serveurs SparQL et serveurs WFS par exemple). Cette proposition est décrite en section 1.2.3.

5) Une tendance importante doit également être prise en compte par les SDI. Il s'agit du **besoin de l'utilisateur de contribuer** à la constitution d'un contenu géographique. Cette dernière tendance est très importante. Elle a été soulignée par (Goodchild 2007) dans un article où il introduit le concept de Volunteer Geography (information géographique produite par volontariat). Elle est illustrée par exemple dans de nombreux projets dédiés à la création collaborative de contenu sur le Web, présentés en section 1.2.4. La prise en compte de contributions d'utilisateurs, ou de producteurs non connus à l'avance, pose également le problème de l'évaluation de la qualité des données produites, voire également le problème de la modération des contributions. J'ai proposé un sujet de **thèse sur l'édition collaborative de données géographiques en ligne**, ce sujet est décrit brièvement en fin de section 1.2.4.

6) Un autre défi est celui du **financement d'une infrastructure**. Il est parfois difficile de trouver des financements car le retour sur investissement ne semble pas immédiat (Harvey and Tulloch 2006). Alors que les SDI de première génération étaient caractérisées par des financements ad hoc ou vouées à l'échec, la question du financement d'une SDI est traitée dans les recherches actuelles. Ces travaux préconisent l'introduction d'outils d'évaluation pour démontrer la valeur des SDI et convaincre les financeurs. Ainsi, (Giff and Cromptoets 2008) définissent des indicateurs de performance en s'appuyant sur un modèle logique de fonctionnement d'une SDI qui est repris sur la Figure 5. Un premier ensemble d'indicateurs de performance concernent le rendement de la SDI, c'est-à-dire des rapports sorties/entrées (output/input). La difficulté pour définir précisément des indicateurs de ce type est d'identifier précisément les sorties (par exemple un entrepôt de données) et les entrées (par exemple un financement) mais également de définir la contribution d'une entrée à une sortie. Deux autres ensembles d'indicateurs sont nécessaires pour mesurer le rapport résultats/sorties (outcomes/output) et le rapport impacts/résultats (impacts/outcomes). Le premier correspond par exemple au nombre d'utilisateurs qui ont trouvé le jeu qu'ils recherchaient grâce au catalogue de la SDI. Le deuxième correspond au nombre d'applications qui se sont développées au-dessus de la SDI. Les auteurs insistent sur l'importance d'indicateurs de performance d'une SDI qui permettent non seulement de convaincre les financeurs mais, avant même cela, d'évaluer si l'infrastructure remplit bien les objectifs qui avaient conduit à la mettre en place. La question du modèle économique d'une SDI demeure ouverte. Il semble pertinent d'étudier les modèles économiques du Web. En particulier, les liens entre des capacités d'acquisition comme OpenStreetMap ou Wikimapia et des responsabilités de production et de diffusion de données géographiques sur un territoire doivent être étudiés.

modèles animé par Paul-Annick Davoine et Thérèse Libourel qui vise à étudier l'apport de la modélisation dans les différentes disciplines et propositions conduites au sein du GDR, ou encore du projet sur la mutualisation animé par Florence Le Ber et moi-même qui vise à faciliter la mise en commun de ressources entre équipes MAGIS.

1.1.4 La transformation de schéma

Il n'existe pas de modèle universel pour représenter l'espace géographique. En particulier, si on considère la représentation sous forme d'objets dans des bases de données dites vecteurs, celle-ci implique plusieurs choix illustrés sur la Figure 6 (Balley 2007). Le producteur de données doit décider des objets du monde réel qui seront représentés dans la base de données et de la façon dont ces objets seront décrits. Or il n'existe pas de choix standard en la matière. Les personnes ont des ontologies différentes du monde réel, des modèles conceptuels et des schémas logiques différents (Raper 1996) (Mennis et al. 2000).

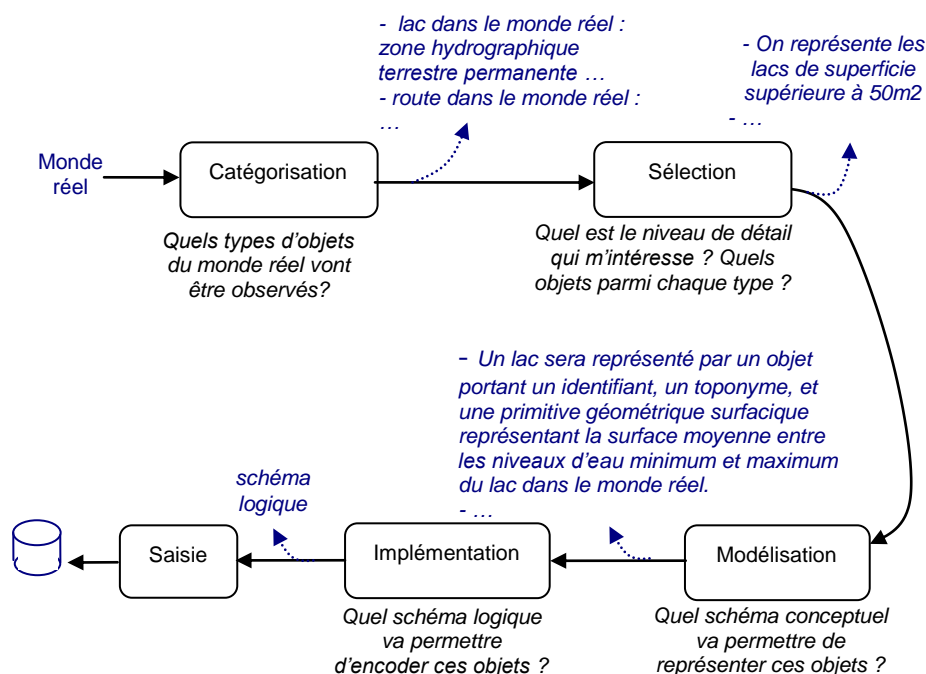


Figure 6. Processus de représentation de l'espace géographique dans une base de données (d'après (Balley 2007)).

La transformation de schéma est un processus important pour les SDIs. Dans les SDI de première génération, le producteur doit généralement transformer le schéma de ses données pour alimenter le produit fédéré. Dans les SDI de deuxième génération, il n'existe pas de schéma fédéré et il faut donc avoir des mécanismes de transformation flexibles qui s'adaptent à divers modèles de données en entrée et en sortie. La transformation de schéma est aussi nécessaire pour l'utilisateur qui acquiert des données et doit les adapter à son application. Nous nous sommes attachés à étudier ce dernier besoin. **Laisser l'utilisateur conduire lui-même une transformation de schéma pose plusieurs problèmes :**

- L'utilisateur doit posséder un outil de transformation.
- Il est possible que lors de la transformation des contraintes d'intégrité soient rompues sans que l'utilisateur en soit conscient.
- La transformation engendre généralement une rupture du lien avec le référentiel initial. Cela complique les opérations ultérieures de mise à jour. De plus, le fait d'utiliser un référentiel géographique commun doit faciliter la collaboration avec d'autres utilisateurs du même référentiel. Il faut pour cela garder les références vers les objets du produit diffusé.
- L'utilisateur doit faire évoluer les métadonnées pour que le jeu final soit pourvu de métadonnées cohérentes.
- A chaque nouvelle acquisition de données, le processus de transformation doit être réitéré.

Le travail de thèse de Sandrine Balley dirigé par Thérèse Libourel professeur d'informatique de l'Université de Montpellier et que j'ai encadré au laboratoire COGIT avait pour objectif de permettre à un utilisateur de spécifier une structure sur mesure de données afin que le producteur de données prenne en charge le processus de transformation, et lui délivre des données adaptées à son application et pourvues de métadonnées. Le processus proposé permet par ailleurs de vérifier des contraintes d'intégrité du jeu au fur et à mesure de la transformation –pour produire un jeu en conformité avec ses spécifications- et de conserver le lien avec le jeu de départ.

Techniquement, la complexité de la transformation de schémas géographiques provient d'une part du fait que le schéma logique de données ne suffit pas à lui seul à rendre compte de la structure des données et d'autre part de la complexité et de la diversité des schémas de données et des règles de grammaire des plate-formes d'encodage (comme la lecture des structures cycliques dans les graphes). **Son approche se fonde sur la représentation explicite de toutes les informations qui rendent compte de la structuration d'un jeu**, que ce soit le jeu initial, le jeu en cours de transformation ou le jeu final (Balley 2007). Une première proposition de ce travail de thèse est un modèle rendant compte de la structure complète d'un jeu en cours de manipulation repris sur la Figure 7.

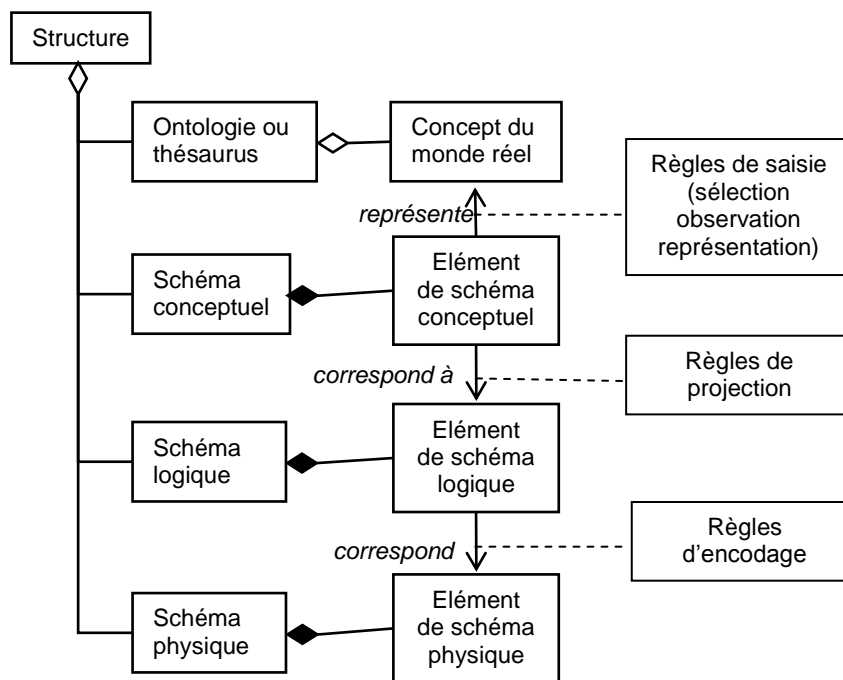


Figure 7. Modèle de représentation de la structure complète d'un jeu de données (d'après (Balley 2007)).

La structure est décrite en quatre niveaux : **le niveau 'physique', le niveau logique, le niveau conceptuel et le niveau du monde réel représenté**. Les schémas logiques et physiques sont associés à des 'plate-forme' qui peuvent comporter des règles de grammaire. Dans le prototype, ces règles sont implantées de façon procédurale dans certaines étapes de la transformation. De plus, l'application s'est concentrée sur un type de plate-forme : la plate-forme Geoxygene qui possède un schéma logique java couplé à un schéma relationnel de stockage (donc un double schéma logique).

Le modèle utilise les standards ISO existants pour décrire ces niveaux (voire Figure 8).

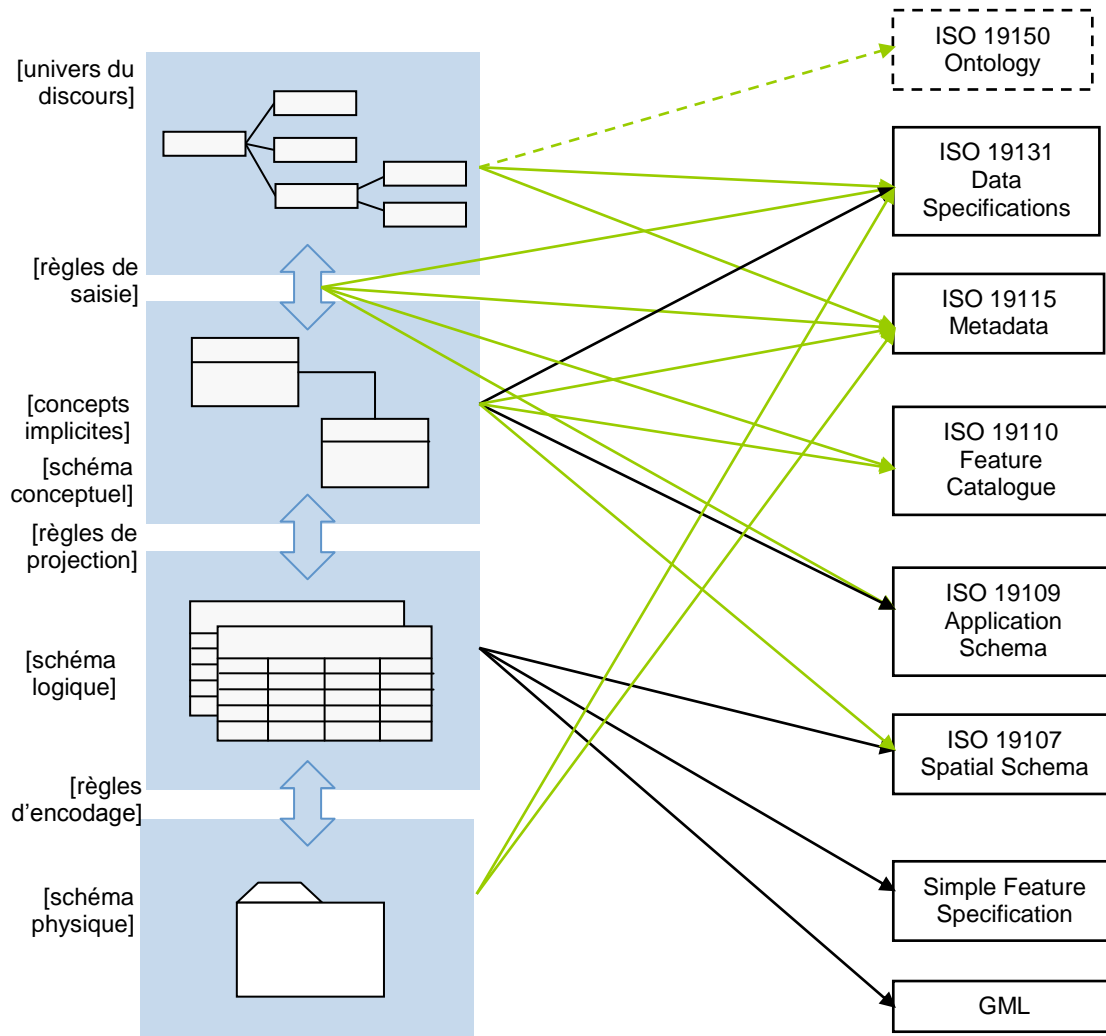


Figure 8. Connaissances impliquées dans la structure d'un jeu de données et modèles standard existants pour les représenter (d'après (Balley 2007)).

Plus précisément, Sandrine Balley a étendu les normes ISO en ajoutant la possibilité de décrire un domaine de valeur d'un attribut (GF_AttributType) à l'aide d'une liste de valeurs provenant du modèle ISO19110 (FC_AttributeValue). Elle a aussi ajouté au General Feature Model ISO19109 l'attribut 'isexplicit' pour une classe ou une propriété. Cela permet d'intégrer dans le schéma des éléments qui peuvent être dérivées à partir d'éléments existants; ces éléments ont une valeur 'false' pour cet attribut alors que les éléments existants ont une valeur 'true'. La dérivation d'un élément implicite est décrite comme une opération de transformation de schéma. Cela est illustré sur la Figure 9. Les opérations de dérivation d'éléments implicites sont décrites plus loin.

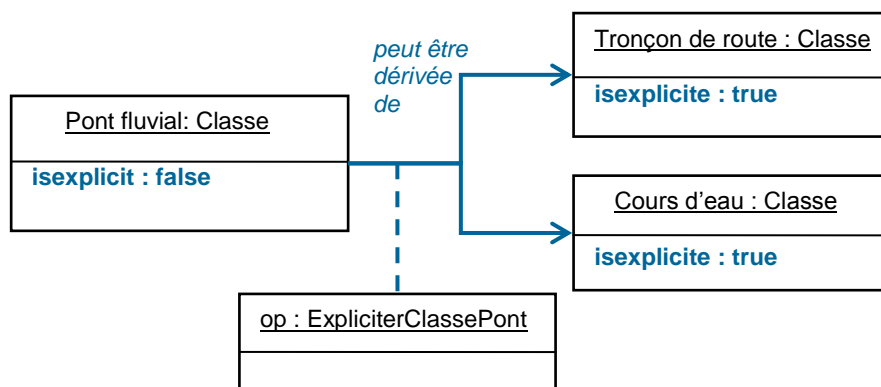


Figure 9. Notion de classe implicite dérivée de classes explicite par une opération de transformation de schéma spécifique (parfois complexe).

Le modèle utilisé pour décrire et conduire les transformations est un modèle à base d'activités repris sur la Figure 10 et qui s'inspire du modèle proposé par (Bucher 2005). L'élément central est la transformation élémentaire qui se compose d'une transformation de schéma conceptuel, d'une transformation de schéma logique et d'une transformation de données. Le processus de transformation de schéma est supporté comme un processus pas à pas. L'utilisateur spécifie une séquence de transformations élémentaires qui conduit au schéma adapté à son application. Dans ce processus, toute transformation élémentaire est bien connue (c'est-à-dire que les effets de ces transformations sur les spécifications IGN sont connus) et bien conduite (c'est-à-dire que les données et leurs métadonnées complètes sont transformées). Le modèle de Sandrine Balley étend par ailleurs le modèle de (Bucher 2005) avec un marqueur d'état d'exécution qui permet d'exécuter une activité en plusieurs fois. Ce marqueur peut prendre trois valeurs : 'exécution non commencée', 'exécution sur les métadonnée de structure', 'exécution sur les données et leur structure'. Grâce à ce modèle, l'utilisateur peut disposer d'une vue actualisée du schéma du jeu sur le client alors que l'exécution sur les données est différée.

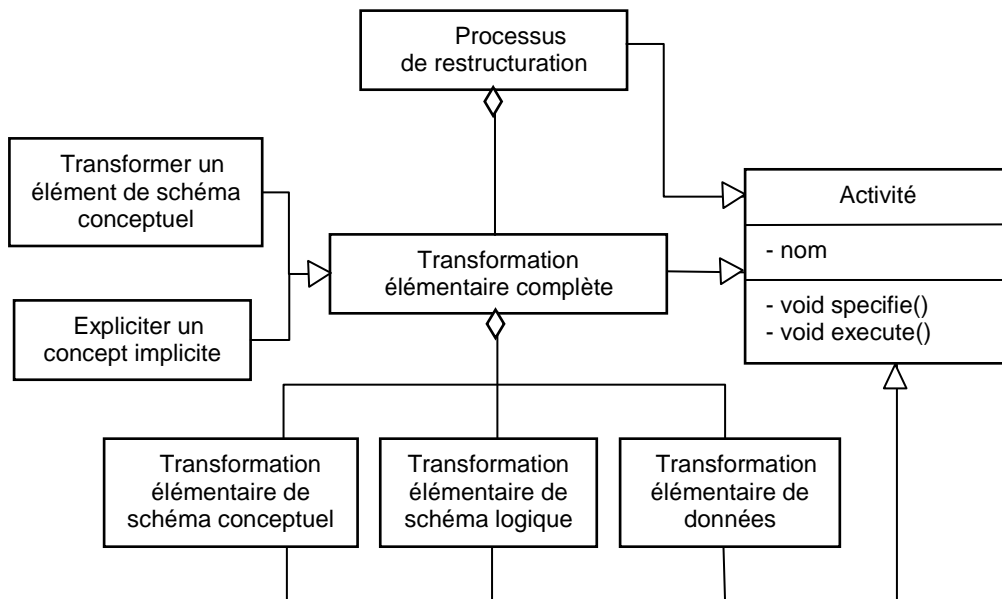


Figure 10. Modèle à base d'activités pour la représentation des connaissances de transformation de schéma (d'après (Balley 2007)).

Les transformations élémentaires correspondent à des opérations classiques d'évolution de schéma et à des opérations moins classiques: l'explicitation de concepts implicites. L'étude des spécifications de contenu permet à Sandrine Balley de déduire comment dériver une représentation sémantique et géométrique pour certains concepts qui apparaissent dans les spécifications mais pas dans le schéma. Dans une même optique, (Lüscher et al. 2008) étudient comment spécialiser des concepts d'intérêt de façon à pouvoir les calculer simplement dans une base de données topographiques. Les auteurs proposent un langage pour la spécification de concepts d'intérêt qui comporte deux niveaux : des concepts de haut niveau qui renvoient à des objets complexes de la réalité et des concepts de plus bas niveau qui renvoient à des propriétés calculables dans la base de données (comme la forme d'un objet ou une structure spatiale). Cette proposition vient donc enrichir considérablement celle de Sandrine Balley.

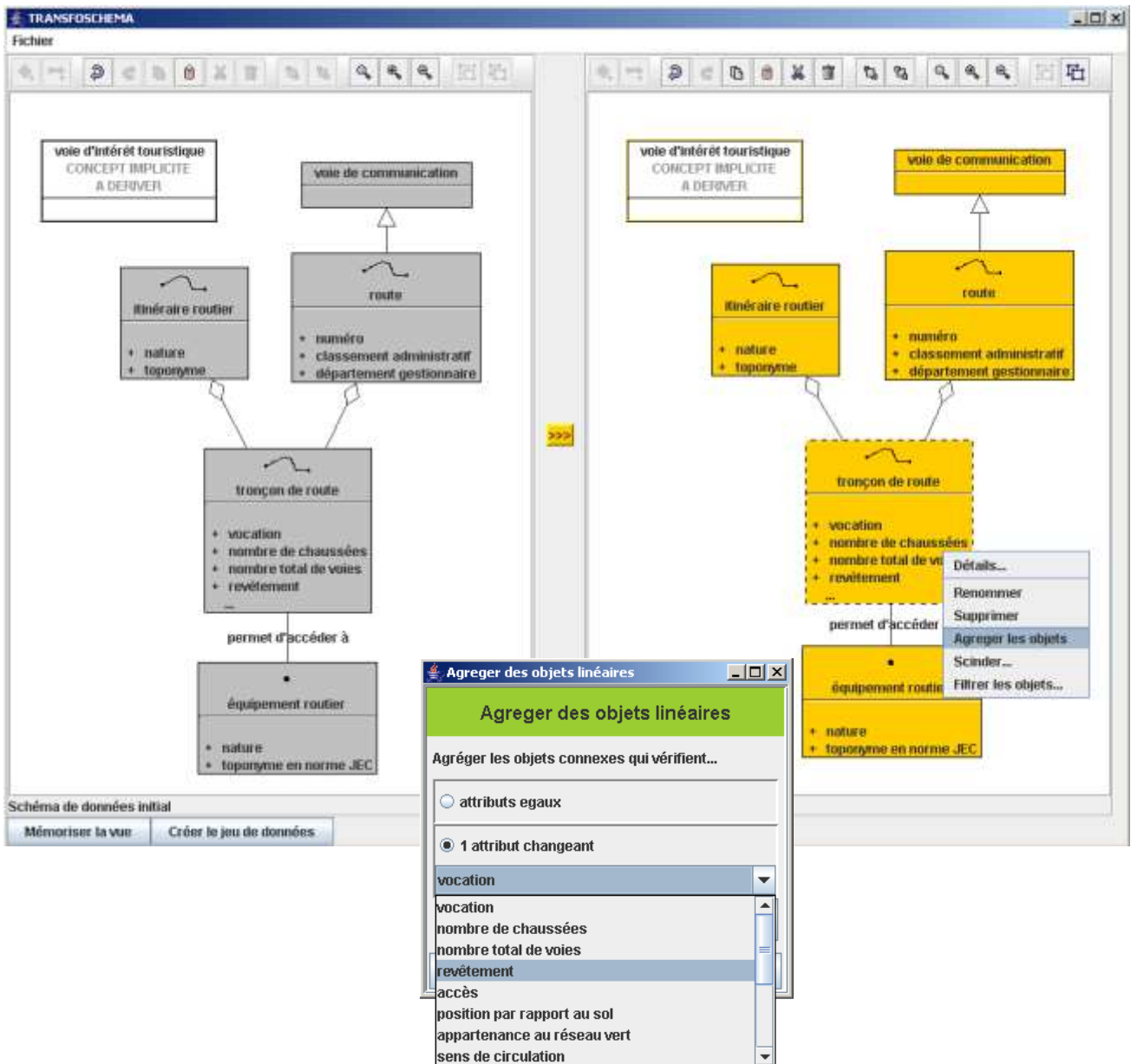


Figure 11. Prototype pour la transformation de schéma développé par (Balley 2007).

La spécification est conduite sur une interface sur laquelle l'utilisateur ne voit que le niveau du schéma conceptuel. L'utilisateur spécifie des transformations de schéma conceptuel que le système complète pour obtenir des transformations élémentaires complètes (Figure 11). Il est possible d'ajouter au système de nouvelles opérations de

transformation de schéma, du moment qu'elles suivent le modèle proposé pour une transformation élémentaire complète.

Une proposition similaire a été faite récemment par (Staub et al. 2008). La différence réside dans le mode de spécification de la transformation de schémas. Les auteurs proposent une spécification assertionnelle. Au lieu de faire spécifier à l'utilisateur une transformation de schéma, ils lui permettent de spécifier un schéma cible et un mapping du schéma source au schéma cible. L'expression du mapping comporte des requêtes de sélection et des transformations. Ainsi, l'utilisateur spécifie bien la transformation de schéma mais de façon portée par le schéma et non de façon organisée en tâches de restructuration. La proposition de (Staub et al. 2008) est plus adaptée au cas où le schéma cible est connu à l'avance et où la spécification de la transformation consiste davantage à établir comment aller de la source à la cible, c'est-à-dire la spécification du 'mapping'. La confrontation de cette proposition à celle de Sandrine Balley ouvre des perspectives prometteuses pour l'intégration de données.

1.1.5 Métadonnées et recherche d'information

Dans cette section, nous rappelons l'importance des métadonnées dans un processus de recherche d'information, que ce soit sur une SDI ou sur le Web à l'aide de moteurs de recherche.

Des métadonnées sont des données à propos de ressources qui permettent d'effectuer des opérations à propos de ces ressources soit parce que ces opérations ne peuvent être conduites directement sur ces ressources (par exemple parce qu'on ne possède pas les ressources) soit parce que ces opérations peuvent être conduites mais difficilement. Les métadonnées peuvent comporter des informations à propos de ressources qui ne sont pas dans la ressource, par exemple le fournisseur de données et le contact. Elles peuvent aussi reprendre des informations qui sont dans la ressource mais les structurer différemment pour en faciliter la manipulation ou encore les regrouper dans un gestionnaire centralisé pour plusieurs ressources, par exemple la couverture géographique d'un jeu de données.

De façon générale, **la diffusion de ressources, qu'elles soient disponibles ou dérivables à la volée, s'appuie sur des métadonnées à propos de ces ressources.** Pour illustrer le cas de ressources existantes, prenons l'exemple d'un tableau d'assemblage de cartes. Les 'coins de cartes' figurés sur le tableau d'assemblage sont des métadonnées des cartes qui permettent à l'utilisateur d'identifier les cartes qui couvrent sa zone d'intérêt. Pour illustrer le cas des ressources dérivables, prenons maintenant l'exemple d'un service de construction de carte. Ses métadonnées précisent quelles sont les étendues spatiales des données et couches utilisées pour le dessin, et indiquent également que l'utilisateur doit donner comme paramètre de la carte à

construire la couverture voulue, sous forme d'un rectangle. L'utilisateur détermine alors le rectangle englobant de sa zone d'intérêt, vérifie que les couches et données couvrent ce rectangle et demande au service de dessiner la carte correspondante à ce rectangle. Les métadonnées sont regroupées dans des catalogues dédiés (dans les SDI) ou dans des bases de métadonnées de moteurs de recherche généralistes. Dans le cas du Web sémantique, les métadonnées sont le contenu des bases de données organisées dans le réseau de contenu distribué appelé Linked Data décrit plus loin. Ainsi le terme métadonnées géographiques a deux acceptions. Il peut renvoyer à des métadonnées sur des données géographiques (c'est le cas dans les spécifications INSPIRE par exemple) et il peut renvoyer à des métadonnées géographiques sur des ressources localisées, c'est-à-dire à des données géographiques (c'est le cas dans les mashups fondés sur Google Maps par exemple).

Les métadonnées servent à confronter une expression du besoin de l'utilisateur aux possibilités de réponses apportées par les ressources. Par rapport aux systèmes documentaires classiques, un principe important dans les SDI ou sur le Web est que l'utilisateur peut ne pas savoir demander 'de façon directe' la ressource dont il a besoin. Bien plus, l'utilisateur peut ne pas vouloir 'trouver' une ressource mais enrichir sa connaissance sur les ressources.

Pour proposer à un utilisateur une exploration de ressources guidée par son besoin, il est courant d'organiser la consultation en deux phases :

1) La première phase est la **découverte**. Elle vise à identifier des ressources potentiellement intéressantes sur les critères principaux de l'utilisateur. La découverte se fait généralement par la requête. L'utilisateur entre un mot-clé dans une zone texte ou renseigne un formulaire de requête. La découverte peut également s'appuyer sur une navigation dans des index. Cette séparation entre requête et navigation est actuellement de moins en moins franche. Par exemple, lorsque l'utilisateur remplit un formulaire actif, ou lorsque un service de complétion automatique de requête lui est proposé (sur le mode appelé 'suggest'), selon le niveau atteint par la complétion la modalité d'interaction sera en partie une navigation. Enfin, des travaux proposent des modèles ad hoc qui traduisent une expression de besoin en une requête de ressource en situant l'expression du besoin dans le domaine de l'utilisateur et non dans celui des ressources. Un modèle d'interprétation est alors nécessaire pour traduire l'expression du besoin en une requête de ressource. Par exemple, le projet K-DIME (Kansei Distributed Information Management Environment) est un système d'information distribué qui fournit une interface de requête par mots-clé. Sa spécificité est de pouvoir prendre en compte des mots-clé éloignés du domaine des ressources et proches de la sensibilité de l'utilisateur grâce à un modèle interne de connaissance qui fait le lien entre des termes (et non des concepts) correspondant à des ressentis, par exemple 'romantique', et des critères de requête des ressources, par exemple 'une image claire'. Ce système apprend au fur et à mesure de son utilisation –il corrige et enrichit son modèle de connaissances- (Bianchi-

Berthouze and Hayashi 2002). Lors de la découverte, le moteur peut aider l'utilisateur en étendant sa requête lorsqu'il n'y a pas de réponse à cette requête, par exemple en suggérant des orthographe voisines pour ses mots-clés. Une aide intéressante est celle des réponses coopératives décrites plus loin (section 1.2.2.1).

2) La deuxième phase est l'**exploration**. Elle vise à étudier les ressources découvertes, leur pertinence par rapport au besoin et éventuellement à prendre en compte d'autres propriétés spécifiques à ces ressources. En effet, l'utilisateur peut ne pas trouver une ressource lui convenant parfaitement mais plutôt devoir choisir entre plusieurs possibilités ayant chacune leurs atouts et leurs inconvénients. L'exploration s'appuie généralement sur la lecture et la visualisation. L'utilisateur doit pouvoir comprendre la pertinence des ressources trouvées vis-à-vis de sa requête. Par exemple le site des pages jaunes affiche une carte de localisation en réponse à une requête utilisant un critère de proximité. Par exemple encore, un moteur de recherche ajoute à la liste des sites trouvés des extraits comportant les mots-clés utilisés dans la requête.

Dans le domaine des métadonnées sur des données géographiques, un problème récurrent est l'**utilisabilité de ces métadonnées**. Il est nécessaire de proposer à un utilisateur humain une aide pour comprendre la pertinence d'un jeu de données géographiques en s'appuyant sur ses métadonnées et également de comparer les pertinences de deux jeux différents à l'aide de leurs métadonnées. Certaines approches dans ce sens sont décrites dans d'autres sections: permettre à l'utilisateur de construire des requêtes de métadonnées en spécifiant des objectifs applicatifs (section 2.1.2.2) ou aménager des interfaces de visualisation des métadonnées (section 1.1.6). Une meilleure utilisabilité des métadonnées passe également par l'**intégration de davantage de sémantique dans les métadonnées**. (Kavouras and Kokla 2008) définissent le sens de données géographiques comme le lien entre ces données et le monde réel. A l'IGN, les spécifications d'acquisition de bases de données tracent le lien inverse puisqu'elles précisent comme le monde réel est observé et est modélisé, conformément au processus illustré en Figure 6. De façon générale, représenter la sémantique dans les métadonnées nécessite de proposer une ontologie formelle du monde réel d'intérêt (par exemple taxonomie d'un domaine) et d'annoter les éléments du modèle conceptuel de données (voire du schéma logique si on ne l'a pas) avec les concepts de l'ontologie. Ce travail est conduit au sein de l'action de recherche en intégration dirigée par Sébastien Mustière ; l'approche consiste à formaliser les spécifications d'acquisition des données. La thèse de Nils Gesbert dirigée par Thérèse Libourel a conduit à un modèle formel décrivant les spécifications des bases de données géographiques de l'IGN. Cette thèse permet donc de lier des éléments du schéma des données géographiques de l'IGN à des catégories d'éléments du monde réel. D'autres travaux de cette action de recherche se sont concentrés sur la construction d'une ontologie topographique à partir des spécifications. Un projet ANR est en cours sur la construction et l'alignement d'ontologies géographiques, le projet GeOnto. Notons que ce travail peut être conduit également sur des métadonnées de services transformant ou

fournissant des données géographiques. Dans sa proposition pour mettre en place l'interopérabilité sémantique de services web géographiques (ce qui revient à un Web sémantique des services géographiques), (Lemmens 2006) utilise également des techniques d'annotation de métadonnées classiques dans des ontologies. Sa proposition est décrite en section 2.1.3.3. (Lutz and Klien 2006) proposent eux un modèle de requête de plusieurs services WFS en utilisant une ontologie fédérée, alignée avec plusieurs ontologies locales utilisées chacune pour annoter les services WFS.

1.1.6 La visualisation de métadonnées géographiques

Dans le domaine de la qualité, plusieurs chercheurs ont étudié la possibilité de mieux comprendre les informations de qualité à propos de données à l'aide de techniques de visualisation. En effet, les métadonnées de qualité sont extrêmement complexes, même pour des experts, à tel point que les utilisateurs estiment que le fait de les documenter revient plus à dégager la responsabilité du fournisseur qu'à aider le choix de l'utilisateur et la bonne utilisation des données (Devilleers 2004). Réciproquement, la qualité est un champ d'application intéressant pour les techniques de géovisualisation. En effet, la qualité d'un jeu de données géographiques peut être vue elle-même comme un ensemble de données localisées qui peuvent être présentées à un niveau de détail fin (le niveau de l'acquisition de l'information de qualité) ou de façon synthétique. (Mackaness et al. 1994) ont dressé il y a plus de dix ans une bibliographie sélective d'articles sur le domaine de la visualisation de la qualité de données géographiques. Plus de 40 articles dans cette bibliographie portaient spécifiquement sur la visualisation de la qualité.

En 2002 j'ai invité au laboratoire COGIT la chercheuse Paula Ahonen de l'université d'Helsinki qui travaillait sur la présentation visuelle de métadonnées à l'aide d'outils graphiques statistiques comme les camemberts ou les diagrammes bâtons (Ahonen 2001). Toutes deux, nous avons particulièrement étudié la spécialisation de la **métadonnée "MD_BrowseGraphic" du standard ISO19115**. Celle-ci est définie dans la norme comme étant un "aperçu visuel" de la donnée. Plus précisément, dans le modèle ISO19115, une ressource est décrite par 1 ou plusieurs entités MD_Identification. Chaque entité MD_Identification est elle-même associée à zéro ou plusieurs entités MD_BrowseGraphics par la relation 'graphicOverview'. Le contenu en information de l'entité MD_BrowseGraphics n'est pas précisé. La norme spécifie à quoi sert cette entité, elle spécifie seulement qu'elle doit avoir 3 champs : le nom du fichier contenant l'illustration, le format d'encodage de ce fichier, la description textuelle de l'illustration (la légende de la figure). Une conclusion de nos réflexions communes est l'existence de deux types de telles métadonnées : les métadonnées BrowseGraphic

définies et construites par les fournisseurs de données, les métadonnées BrowseGraphic définies et construites par les services de catalogage (par exemple en utilisant d'autres métadonnées ISO19115).

Par la suite, j'ai étudié la définition de telles **métadonnées de visualisation pour des spécifications**. En effet, comprendre la qualité d'un jeu de données nécessite au préalable de comprendre les spécifications en regard desquelles la qualité est évaluée. Ce travail a fait l'objet du stage de DESS cartographie de Gyorgyi Göder, co-encadré par moi-même et par Sébastien Mustière (Göder 2003). Gyorgyi Göder a étudié le recours à la cartographie pour présenter à un utilisateur des spécifications de contenu de bases de données géographiques en supportant en particulier la comparaison entre spécifications. Plus exactement, nous lui avons demandé de représenter les différences entre les spécifications en s'inspirant de la typologie établie par (Devogele et al. 1998) :

- Différences de sélection. Ces différences se voient parfois sur le schéma conceptuel, par exemple les zones inondables ne sont pas représentées du tout dans la BDCarto et cette classe est absente du schéma. Parfois la sélection se fait au sein d'une classe d'objet, la classe existe mais toutes les instances du monde réel ne seront pas représentées dans la base. Par exemple, les surfaces hydro-terrestres permanentes sont représentées dans la BDTopo si leur surface excède 20m² et dans la BDCarto si leur surface excède 4ha.
- Différence de classification. Ces différences se voient au niveau du schéma conceptuel.
- Différence de représentation des attributs dont la géométrie. Les différences de représentation des attributs se voient au niveau du schéma conceptuel sauf si on ôte du schéma conceptuel le positionnement (auquel cas les différences géométriques ne se voient pas). De plus, la géométrie se prête bien à représentation visuelle. Ce qui se voit mal au niveau du schéma sont les conditions de représentation, c'est-à-dire par exemple sous quelles conditions un fleuve sera représenté comme un objet linéaire ou comme un objet surfacique.

Deux grands types de BrowseGraphics sont proposés. Les premiers sont des tableaux de synthèse centrés sur des catégories d'objets du monde réel. Le tableau repris Figure 12 illustre les différences de spécifications de la BDTopo et la BDCarto pour les cours d'eau.

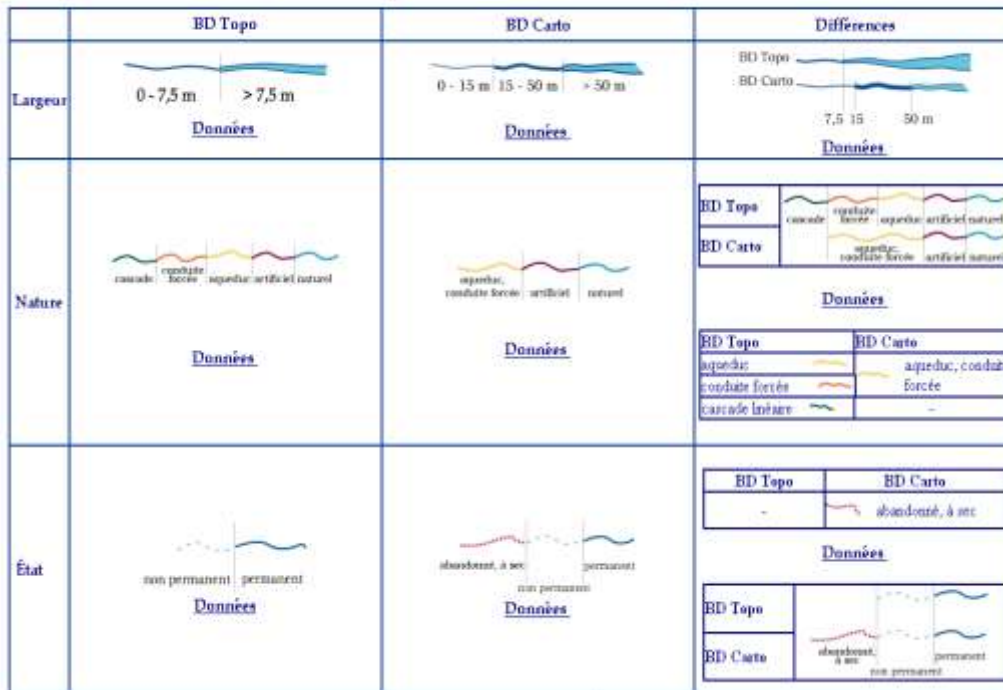


Figure 12. Visualisation des spécifications de produit de la BDTopo et de la BDCarto centrée sur la comparaison des attributs.

Les deuxièmes types de BrowseGraphics sont construits sur des extraits de données et visent à montrer à l'utilisateur, sur une zone donnée, le nombre d'objets concernés par la représentation.

Plus récemment, (Kent 2007) a proposé une méthode originale pour visualiser des spécifications de cartes topographiques de façon à pouvoir les comparer. Il utilise un diagramme radar ou cartogramme. L'auteur divise l'espace en autant de cadrans égaux que de thèmes étudiés. Ces thèmes ont été identifiés en étudiant des cartes topographiques européennes à grande échelle. Chaque axe est ensuite gradué de 0 (rayon nul) à 20 et l'auteur rapporte sur l'axe le nombre de symboles utilisés pour décrire ce thème dans la légende, ce qui correspond pour lui à la richesse du vocabulaire cartographique de la légende. Le diagramme final montre selon (Kent 2007) l'importance accordée à certains thèmes dans le vocabulaire graphique de la carte. La comparaison des formes des diagrammes permet de comparer les spécifications des différentes cartes. Ce travail appelle des critiques. La première est le fait que le diagramme radar n'est pas adapté à la comparaison car l'œil y voit avant tout les superficies et non les quantités mesurées sur les axes. La suivante est que l'importance accordée par le vocabulaire cartographique ne tient pas uniquement au nombre de symbole consacré à un thème mais aussi à l'importance des symboles : si les voies ferrées apparaissent en

traits extrêmement fins elles auront moins d'importance visuelle que les routes, un peu comme les paragraphes écrits en tout petit dans les contrats d'assurance.

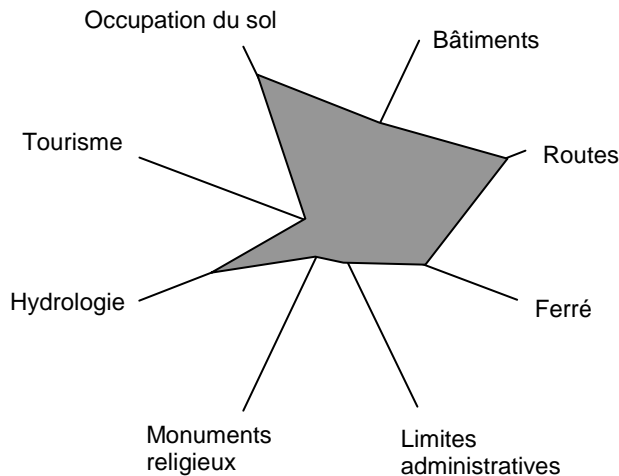


Figure 13. Principe de visualisation des spécifications de carte proposé par (Kent 2007). Les mesures portées sur les axes correspondent au nombre de symboles associé au thème dans la légende.

1.1.7 'Collages spatiaux' sur les infrastructures d'information géographique

1.1.7.1 L'importance de la navigation entre plusieurs niveaux de détails

L'appréhension de plusieurs niveaux de détails est une fonctionnalité importante sur les SDIs pour plusieurs raisons.

De façon générale, il n'y a pas une échelle unique pour étudier un phénomène géographique mais souvent plusieurs échelles et l'échelle initiale ne sera qu'un point de départ duquel pourra partir l'analyse (Mackness 2007). Un aspect important de ces cartes est donc de faciliter le passage d'un niveau de détail à un autre, en aidant l'utilisateur à ne pas se perdre entre les deux échelles.

Cette fluidité est plus spécialement importante dans un contexte ubiquitaire où la charge cognitive de l'utilisateur doit être allégée. La notion de collage spatial a été introduite par (Claramunt et Mainguenaud 1996) pour désigner une représentation permettant à un

utilisateur de faire la correspondance entre deux échelles dans un contexte d'aide à la navigation.

La deuxième raison est que, lors de l'utilisation d'une carte écran, l'utilisateur doit souvent s'appuyer sur une petite échelle pour avoir une vue d'ensemble car la taille de son écran est plus réduite que la taille d'une carte traditionnelle à l'échelle qui l'intéresse. Cela est particulièrement vrai dans le cas des petits écrans des terminaux nomades (Paolino et al. 2007). Dans un contexte mobile, l'utilisateur a souvent besoin de deux niveaux de détails de façon simultanée :

- un niveau appelé 'focus' qui comporte les informations de repérage immédiat dans la réalité (comme les noms de rues et les adresses pour un piéton),
- un niveau appelé 'contexte' qui comporte les informations utiles pour son activité en cours (comme les points d'intérêts accessibles en une demi-heure depuis sa localisation) (Paolino et al. 2007).

Des propositions prometteuses consistent à déformer la carte pour permettre de garder sur la même carte une ou plusieurs zones d'intérêt à un niveau de détail fin et le contexte à un niveau de détail plus grossier (Harrie et al. 2002). Une difficulté de ces approches est de s'assurer que le lecteur comprenne bien la distorsion et intègre le changement d'échelle dans sa lecture du territoire concerné. La proposition de (Takahashi 2008) propose de diminuer cet effet en prenant en compte, lors de la distorsion, de différences d'échelles dans la propre carte cognitive du lecteur.

Globalement, faciliter l'appréhension de différents niveaux de détails de façon fluide et continue va dans le sens de ce que (Cartwright 2003) désigne par 'Engineered Serendipity'. Engineered Serendipity se traduit littéralement comme 'mettre en scène des heureux hasards'. C'est selon l'auteur une qualité essentielle de la carte : faciliter l'exploration libre du lecteur tout en le guidant correctement par la sélection des informations pertinentes à afficher.

1.1.7.2 Propositions pour l'aide à l'exploration de plusieurs niveaux de détails

Un premier travail que j'ai encadré sur le sujet qui précède a consisté à reprendre une proposition de la littérature, **Framy de (Paolino et al. 2007)**. Celle-ci consiste à ajouter à une carte des bords colorés indiquant la densité de points d'intérêt (POI) dans les diverses directions autour de la zone affichée à l'écran comme illustré sur la Figure 14.

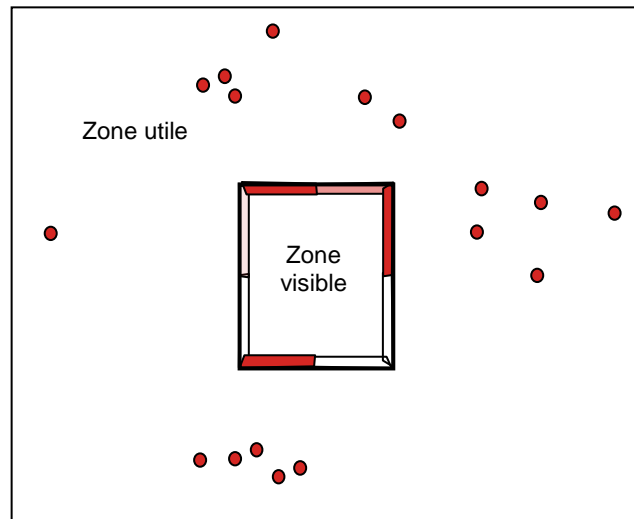


Figure 14. Principe de la méthode Framy de (Paolino et al. 2006).

Ce travail a fait l'objet du **stage de mastère CarthaGéo de Charlotte Hoarau** (Hoarau 2009). Charlotte Hoarau a implémenté un simulateur de petit écran sur OpenJUMP et a enrichi la proposition Framy en utilisant deux variables graphiques : l'intensité de la couleur du bord et l'épaisseur du bord. Cela permet de visualiser les valeurs de deux fonctions d'agrégation dans les différentes directions entourant la position de l'utilisateur. Une de ces variables rend par exemple compte de la densité de points d'intérêt dans cette direction et l'autre variable rend compte de la distance du plus proche point d'intérêt. Charlotte Hoarau a conduit des tests utilisateurs pour valider l'efficacité de ses propositions sémiologiques. Les tests ont montré que les utilisateurs apprennent rapidement à interpréter ces bords colorés d'épaisseurs différentes.



Figure 15. Illustration de la proposition de (Hoarau 2009).

Par la suite, j'ai proposé un autre sujet de stage destiné au monde 3D TerraNumerica. Les environnements de visualisation 3D actuels offrent un point de vue trop global (vue globale de paris) ou trop local (vue locale proche de l'expérience vécue dans la réalité) mais ne facilitent pas l'intégration de ces points de vue ensemble. Le visiteur de la représentation de l'espace a besoin d'information de type 'focus' correspondant à son

voisinage immédiat et d'information de type 'contexte' correspondant aux informations utiles pour son activité en cours. Le travail de stage de mastère SIG d'Ameur Boucheritte a étudié la conception et l'implantation de deux services pour l'aide à la navigation entre plusieurs niveaux de détails (Boucheritte 2009):

- Un service d'aide au 'zoom' qui calcule un nombre fini de propositions de zoom à partir d'une vue globale donnée et en fonction de plusieurs critères de pertinences. Les propositions de zooms sont ordonnées.
- Un service d'aide au 'pan' qui calcule des informations sur les directions autour d'une position locale en fonction également de plusieurs critères de pertinence. Ce dernier service se rapproche des fonctions d'agrégation proposées par Charlotte Hoarau (Hoarau 2009).

Nous avons prévu que ces services pouvaient être interrogés dans deux contextes : l'exploration d'un monde 3D virtuel et la consultation d'une carte sur un petit écran.

Le service de **proposition de zoom** est destiné à suggérer à un utilisateur des points de départ pour explorer une carte ou un monde virtuel à une échelle locale. Ce service prend en entrée, entre autres, les arguments suivants : une zone initiale sur laquelle est conduite l'exploration (par exemple Paris), une dimension de zone pour les zooms demandés, une couche de points d'intérêt de premier ordre (il s'agit des objets auxquels l'utilisateur s'intéresse en priorité), une couche de points d'intérêt du deuxième ordre (il s'agit d'objets dont la présence n'est pas nécessaire mais augmente la pertinence d'une zone). Deux méthodes de regroupement ont été testées : un automate cellulaire et une méthode de classification ascendante hiérarchique. Les points d'intérêts du deuxième ordre sont pris en compte pour favoriser la création de groupes des points du premier ordre. La Figure 16 montre un résultat de ces méthodes (en bleu la classification ascendante hiérarchique et en rouge l'automate cellulaire). Cette figure montre l'importance d'introduire d'autres critères, par exemple la sémantique des points d'intérêt, lorsque les POIs sont denses. Nous retenons comme intérêt de la méthode de l'Automate Cellulaire Géographique la possibilité d'écrire des règles d'agrégation fines et lisibles, qui peuvent prendre en compte tout l'espace. Cette méthode présente un inconvénient majeur qui est le biais introduit par la définition (arbitraire) des cellules de l'automate. La classification ne présente pas un tel biais mais il est plus complexe de synthétiser dans une mesure de distance toutes les connaissances utiles au regroupement.

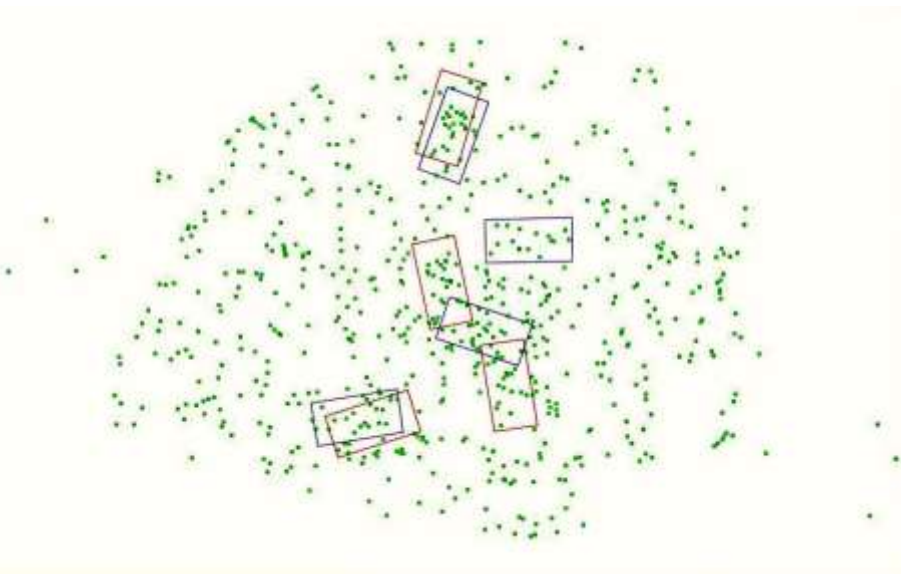


Figure 16. Résultats de la génération de propositions de zoom sur Paris en utilisant comme points d'intérêts les POIS culture et loisir de la BDTopo® et en utilisant comme points du deuxième ordre les bâtiments remarquables les plus connus (Tour Eiffel, Sacré Coeur, Notre Dame, etc). (Boucheritte 2009)

Le service **d'aide au pan** vise à aider un utilisateur qui est sur une vue locale à décider dans quelle direction déplacer sa fenêtre de visualisation grâce à une appréhension d'information de synthèse sur son voisinage proche. Cette information de synthèse est délivrée sous la forme d'un découpage du voisinage en secteurs angulaire et pour chaque secteur le nombre de POIs dedans et la distance du plus proche POI. Ameur Boucheritte a cherché à améliorer la proposition de (Hoarau 2009) en calculant les secteurs angulaires les plus pertinents en fonction de la distribution des points d'intérêt. Il a proposé et comparé plusieurs méthodes. Des résultats sont illustrés sur la Figure 17.

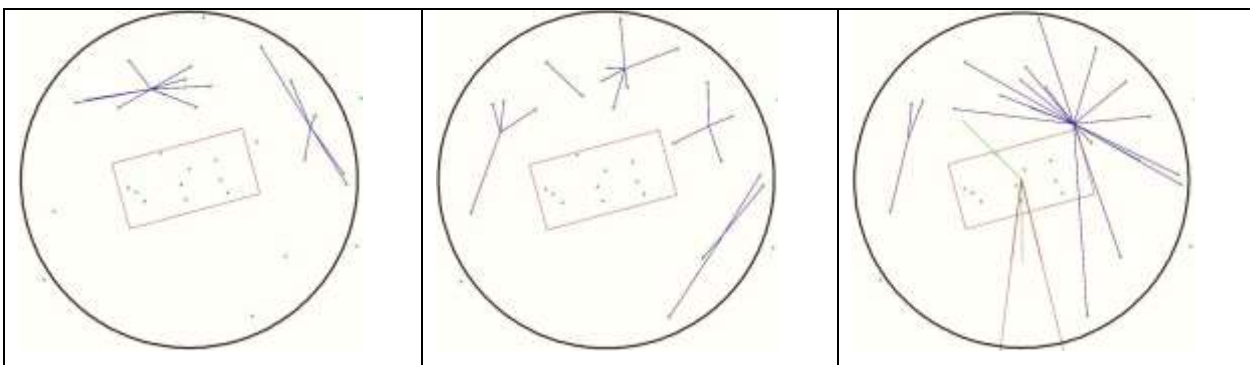


Figure 17. Résultats du calcul de groupes de POIs autour d'une zone de zoom avec la méthode de classification ascendante hiérarchique (à droite), l'histogramme polaire suivi de Kmine (au centre), la prise en compte des POIs de deuxième ordre pour créer les secteurs (à gauche). (Boucheritte 2009)

Une visualisation dans un environnement 3D des informations fournies par le service d'aide au pan a été expérimentée grâce à la contribution de Mickaël Brasebin de l'action de recherche sur l'analyse spatial, le temps et le 3D au COGIT (voir Figure 18). Un bord de faible largeur du disque indique que la proximité d'un POI dans ce secteur, une couleur de forte intensité indique une grande densité de POIs dans ce secteur.

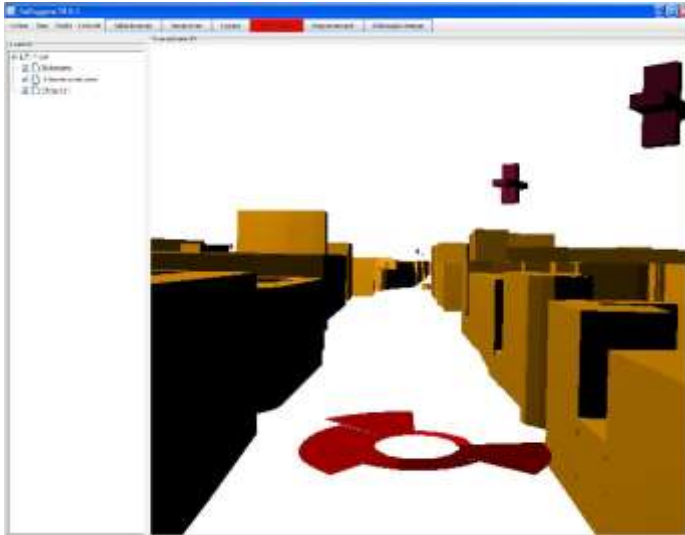


Figure 18. Visualisation dans un environnement 3D (Geoxylene 3D) des informations d'aide au pan. (travail réalisé par Mickaël Brasebin)

1.2 Le GéoWeb

Sur le Web, d'autres solutions contribuent à faciliter le partage d'information géographique par des parties prenantes. Elles se différencient des infrastructures précédentes : **autres modèles et autres acteurs**. Pour autant, ces solutions apportent aux SDI des **clés intéressantes**: elles utilisent des techniques autres, elles sont adoptées par de très nombreux utilisateurs, certaines possèdent un modèle économique apparemment réussi. Réciproquement, les experts des SDI classiques ont probablement un rôle à jouer dans les solutions provenant du Web car les démarches que nous allons présenter venant du Web sont confrontées aux difficultés bien connues des acteurs traditionnels comme la cartographie, la gestion de la qualité, la gestion des versions, la mise à jour des données ou l'intégration.

La section 1.2.1 décrit les moteurs de recherche géographiques et en particulier le projet européen Spatially Aware Information Retrieval on the Internet (SPIRIT) auquel j'ai participé. La section 1.2.2 introduit les notions clés du Web sémantique géographique, domaine sur lequel j'ai effectué une veille technique. La section 1.2.3 présente notre participation à un projet de plate-forme de diffusion et intégration de données géographiques dans lequel nous avons essayé de mettre en œuvre ensemble des principes des SDI et des principes du Web sémantique : TerraNumerica. Enfin, une dernière section 1.2.4 décrit quelques projets liés à la création collaborative de contenu.

1.2.1 Les moteurs de recherche géographique

1.2.1.1 Généralités

Des sites incontournables pour accéder à une information sur le Web sont les **moteurs de recherche**. Globalement, un moteur de recherche conduit deux types de processus :

- 1) Un processus de prétraitement.
- 2) Les processus de requête.

Au cours du **processus de prétraitement**, le moteur analyse les ressources auxquelles il veut donner accès et construit pour chaque ressource une description de cette ressource (**les métadonnées de la ressource**). Par exemple les métadonnées d'un

document html dans Google sont les mots contenus dans ce document (leur fréquence, leur situation dans le document) et également des informations comme les liens conduisant à ce document, le caractère officiel de l'url et ainsi de suite. Ensuite le moteur crée **des indexes**. Par exemple un index textuel qui associe à chaque mot les documents dans lesquels ce mot apparaît.

Au cours d'un **processus de requête**, le moteur propose une zone de saisie de la requête (mots clés généralement ou formulaire de requête avancée). Jusqu'à récemment les mots d'une requête étaient traités comme des chaînes de caractère vides de sens. Depuis peu, le moteur Google détecte les noms de lieux et propose des cartes, il détecte aussi les titres de films, les titres de livres et les noms de lieu. Le moteur confronte la requête aux indexes pour sélectionner des ressources pertinentes. A cette étape, le moteur peut ne pas tenir compte de la possibilité qu'un mot soit mal orthographié dans un document de sorte que ce document ne sera pas sélectionné en réponse à une requête. Ce n'est pas grave car il y a suffisamment d'autres ressources qui répondront à la requête. Un critère décisif de performance à cette étape est la rapidité du moteur. L'étape suivante consiste à **regrouper et ordonner les réponses**. Son importance ne doit pas être sous-estimée. Si la réponse à une requête est constituée de plusieurs centaines de réponses il est nécessaire que le moteur les ordonne pour que les réponses figurant sur les deux premières pages soient les plus pertinentes. Selon le cas, l'utilisateur regarde les cinq ou les vingt premières réponses mais pas les suivantes. Le regroupement est également important. Il consiste à regrouper les réponses qui ont des pertinences similaires pour pouvoir présenter à l'utilisateur dans les premières réponses au moins une réponse parmi tous les groupes différents.

L'information géographique est souvent présente dans la recherche d'information sur le Web. En 2001 une étude a parcouru 20 000 pages du domaine anglais (selon leur adresse IP) et a constaté que 85% d'entre elles comportaient une référence à un nom de lieu (McCurley 2001). (Zhang et al. 2006) ont analysé un échantillon de quatre millions de requête et calculé que 12,7% contenait un mot qui était un nom de lieu. En 1996, Ray Larson a parlé de '**Geographical Information Retrieval**' pour désigner la problématique de rechercher une information sur des critères spatiaux (Larson 1996). A la suite de cela, une communauté scientifique nouvelle s'est formée pour traiter cette problématique qui réunit des chercheurs de la recherche d'information et des chercheurs de la géomatique.

1.2.1.2 SPIRIT : l'architecture

De 2002 à 2005, le laboratoire COGIT a participé au **projet SPIRIT (Spatially Aware Information Retrieval on the Internet**, <http://www.geo-spirit.org/>), projet européen IST 2001 35047. L'objectif de ce projet était de concevoir un moteur de recherche qui

applique des techniques issues des recherches en géomatique pour améliorer la recherche d'information sur le Web, c'est-à-dire un moteur de recherche géographiquement intelligent (Jones et al. 2004) (Purves et al. 2007). Le consortium de ce projet était composé de partenaires géomaticiens, les universités de Cardiff, Zürich, Hanovre, Utrecht et le laboratoire COGIT ainsi que d'un partenaire en recherche d'information, l'université de Sheffield.

Un groupe technique réunissant un membre de chaque partenaire –et dans lequel je représentais le COGIT- a établi une proposition d'architecture distribuée pour le développement du prototype. Plusieurs composants ont été identifiés, ainsi que les échanges entre ces composants. Le protocole SOAP a été choisi pour ces échanges (Finch et al. 2003).

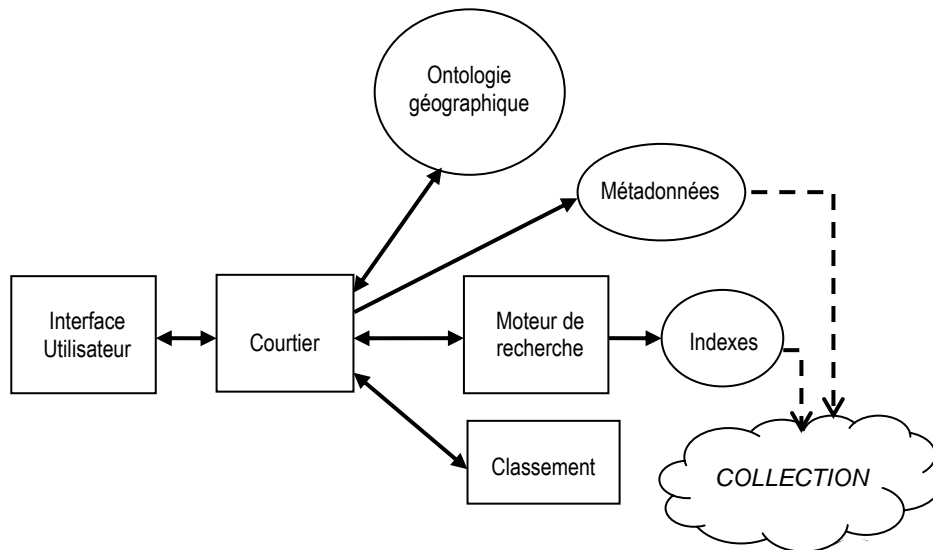


Figure 19. Composants impliqués dans le moteur de recherche SPIRIT.

Les composants présentés sur la Figure 19 interviennent de la façon suivante dans le traitement d'une requête.

- 1) L'utilisateur exprime son besoin sur l'interface graphique sous forme d'un triplet (concept, relation, lieu), par exemple (pêcher, dans, Ecosse). Les champs pour le concept et le lieu sont en texte libre alors que les relations sont choisies dans une liste.
- 2) L'interface transmet la requête au courtier qui détermine grâce à l'ontologie s'il y a une ambiguïté de lieu dans la requête. Si le lieu est ambigu, l'ontologie transmet au courtier les informations nécessaires pour lever cette ambiguïté. Dans ce cas, le courtier transmet ces informations à l'interface pour qu'elle demande à l'utilisateur de préciser le lieu auquel il fait référence.

- 3) Une fois l'ambiguïté levée, la partie (relation, lieu) de la requête est interprétée en une ou plusieurs zones polygonales d'intérêt grâce à l'ontologie géographique.
- 4) La nouvelle requête composée de mots clés et d'une empreinte géométrique est alors transmise au moteur de recherche.
- 5) Ce dernier s'appuie sur des indexes pour obtenir au plus 1000 documents répondant à la requête. Les indexes sont un index spatial et un index textuel.
- 6) Les identifiants des documents trouvés sont ensuite transmis au module de classement. Ce module obtient du courtier les métadonnées des documents qu'il utilise pour calculer leur note de pertinence et les regrouper. Le regroupement se fait sur des critères spatiaux et thématiques.
- 7) Les réponses classées sont transmises à l'interface utilisateur qui les met en forme et les présente à l'utilisateur (voir Figure 20).

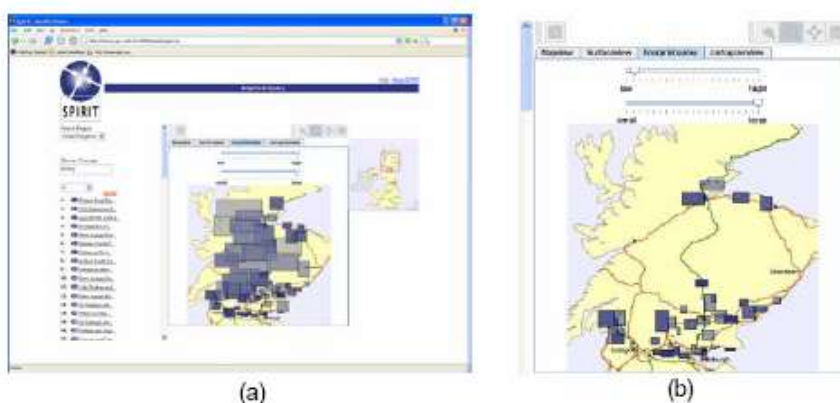


Figure 20. Deux exemples de visualisation des empreintes spatiales des réponses à une requête sur SPIRIT. A gauche, (a), tous les documents. A droite, (b), uniquement les documents très localisés dont l'empreinte est de taille réduite. (d'après (Purves and Yang 2005))

Un élément essentiel pour le prétraitement et le traitement de la requête est l'**ontologie géographique** de SPIRIT. C'est une base de données géographiques dans laquelle certaines relations sont précalculées et dans laquelle certaines objets ont été ajoutés pour les objectifs de SPIRIT. Ces objets correspondent à des lieux le plus souvent aux frontières floues, comme le Sud de la France, le Nord de Paris ou les Midlands. Les frontières géographiques de ces lieux ont été calculées de deux façons. La première consiste à appliquer des données géographiques, des outils d'analyse spatiale et des hypothèses pour extraire des zones comme : le centre historique (en s'appuyant sur la forme des rues), la zone industrielle ou la banlieue (Heinzle et al. 2003). La seconde s'appuie sur les autres noms de lieux (que l'on sait localiser) qui apparaissent dans les documents répondant à la requête composée du nom de lieu à localiser (Purves et al. 2005). De plus, certaines relations spatiales entre entités ont été explicitées pour ne pas devoir être recalculées lors de la consultation.

La phase de prétraitement (constitution des indexes et des métadonnées de documents) s'est fait globalement de la façon suivante. Au lieu de parcourir effectivement le Web, les partenaires du projet ont obtenu de Google une partie de sa copie du Web correspondant à 1 Teraoctet de données (Joho and Sanderson 2004). Ces documents ont ensuite été analysés un à un pour obtenir une base de métadonnées sur eux. Ces métadonnées étaient en partie des métadonnées classiques utilisées en recherche d'information (fichiers inversés) et également une métadonnée de localisation. Le processus de construction d'une telle métadonnée s'appelle geo-tagging. Le geo-tagging est conduit en plusieurs phases dans SPIRIT :

- détection des noms de lieux dans le document ('geoparsing') et des codes postaux, en s'appuyant sur les toponymes et codes postaux présents dans l'ontologie SPIRIT.
- validation du nom de lieu ('geocoding'). Une désambiguïsation est parfois nécessaire. Elle est conduite en s'appuyant sur le contexte de la page et sur les termes qui entourent le lieu désigné par ce nom dans l'ontologie.

A la suite de cette phase, les index de SPIRIT sont mis à jour : un index textuel classique et un index spatial indexant les noms de lieu.

1.2.1.3 SPIRIT : analyse du besoin

La participation du COGIT a également consisté en la proposition d'une approche centrée utilisateur dans la conception du système, c'est-à-dire la conduite d'une analyse du besoin qui sera utilisée dans la phase d'évaluation pour confronter la réponse aux besoins exprimés.

Première analyse des attentes des utilisateurs finaux

Une première étude des besoins a permis d'identifier deux groupes d'utilisateurs (Balley et al. 2002):

- Les personnes sans expertise particulière recherchant sur le Web des informations à propos de 'quelque chose associé à quelque part'. Ce groupe avait besoin d'aide pour l'expression du besoin et l'exploration des ressources d'intérêt.
- Les utilisateurs de SIG recherchant des données géographiques. Ce groupe avait besoin d'un SIG en ligne.

A l'époque, cela renvoyait à deux besoins si différents que la décision a été prise de ne s'adresser qu'au premier groupe. En effet, les besoins du deuxième groupe étaient théoriquement traités par les travaux sur le catalogage de données géographiques et la proposition de standard de métadonnées pour le catalogage (ISO19115) ainsi que sur les propositions d'extension Web des SIG.

Première analyse des propositions techniques envisagées par les partenaires

Une étude a également porté sur **les possibilités du système que les partenaires entendaient mettre au point**. Chaque membre du projet a été sollicité pour expliciter ce qu'il entendait par *spatially aware information retrieval on the internet* à l'aide d'un scénario détaillé. Des extraits d'un de ces scénarios sont repris ci-dessous pour illustrer cette notion.

I am planning to spend some holidays on the Mediterranean coast of France or Italy. I would like to reside in a big city (i.e. with restaurants and shops and many things to do) to occupy my evenings. And I would like to practice windsurf in a place nearby but not in the city itself where I expect to find too many tourists during the day.

As a result, I expect the URLs of : sites proposing hotels to stay in relevant cities, sites proposing renting of windsurf in places nearby, and I would like that the places concerned by these sites appear on one map.

[..]

4) Then I submit the query "windsurf renting near Nice" to SPIRIT

[..]

6) SPIRIT interprets the query :

what : windsurf renting
where : near Nice

SPIRIT asks me to specify what I mean by "near" by specifying among the following choices:
within a ..mn walk distance, within a ...mn driving distance, within ...km

I specify "within a 20mn driving distance".

SPIRIT displays a map of villages around Nice. I select the villages near the sea.

[..]

If there are answers, SPIRIT puts on a map symbols corresponding to the activity windsurf renting, located on the villages and linked to the corresponding Web sites.

If there are no answer, SPIRIT looks through its ontology of activities to find other activities attached to the geographic area mediterranean coast. It finds: diving, sailing. It proposes me to look for Web sites supporting these activities, I select diving... (auteur : Sandrine Balley)

Synthèse et validation des analyses du besoin

Suite à ces deux analyses, **les possibilités techniques ont été croisées avec les expressions de besoin pour aboutir à une liste de fonctionnalités attendues** par les utilisateurs visés et correspondant à ce que les partenaires pensaient que SPIRIT pouvait atteindre (Bucher et al. 2004). Ces fonctionnalités ont été illustrées sur une maquette graphique de SPIRIT et de nouveaux **tests utilisateurs ont été organisés pour les valider et les raffiner**. Ces nouveaux tests ont été conduits dans tous les sites des partenaires à l'aide d'un même diaporama et de questionnaires identiques (à la traduction près) qui ont été mis au point au COGIT. Lors des tests, le testeur faisait défiler le diaporama illustrant les fonctionnalités prévues dans SPIRIT et les utilisateurs

rédigeaient leurs réactions sur des documents papiers qui ont ensuite été numérisés et compilés.

Mike scenario

Oral comment:

To refine the query, Mike draws an area of interest on the map AND clicks on search.



Questions :

17- What do you think of using the map to express an area, is it very important, important, useful, non useful?

18- Would you prefer other ways like select dots, pick a place name, express a geographic constraint like near the lake, near the ski slopes?

Figure 21. Exemple de diapositive utilisée lors de la validation des fonctionnalités utilisateurs prévues dans SPIRIT. Les textes à côté correspondent aux explications orales que le modérateur du test doit fournir lors du test.

Ce dernier test a permis de valider nos propositions et de montrer que les utilisateurs attendaient de SPIRIT non seulement de l'aide concernant la recherche d'un site Web mais également une aide spatiale par ailleurs. Autrement dit, la carte proposée par SPIRIT en réponse à une requête ne doit pas servir uniquement à localiser les sites Web trouvés –en cartographiant leurs empreintes géographiques- mais doit également apporter une information d'intérêt dans le contexte de la requête.

1.2.1.4 SPIRIT : évaluation

Une méthode générique d'évaluation d'un moteur de recherche géographique a été proposée au COGIT lors du travail de stage de Guillaume Aumaitre (Aumaitre et al. 2004) qui comporte une phase centrée utilisateurs et une phase centrée système.

Les tests centrés utilisateur ont été conduits dans différents sites du projet sur le premier prototype, en 2004 (Aumaitre et al. 2004b). Dans chaque site de test, les mêmes tâches et le même questionnaire (traduits) ont été fournis aux utilisateurs conformément à la méthodologie proposée. Il est apparu globalement que le fait de donner des tâches trop précises aux utilisateurs les empêchait de juger le prototype. Le test du prototype final a donc nécessité d'adapter cette méthode générique. Nous avons donné une plus grande liberté aux utilisateurs. En contre-partie, cela impliquait de les prévenir des biais du prototype. Ce dernier ne parcourt pas le Web mais une partie du Web et son intelligence spatiale se concentre sur certaines zones géographiques. Nous avons également inclus un test comparatif avec des outils similaires¹. Globalement les utilisateurs ont apprécié la sobriété de l'interface, le fait de pouvoir utiliser des relations spatiales dans la requête, et la carte de présentation des résultats. Tous se sont plaints de la lenteur du moteur, en particuliers en comparaison avec les autres outils.

L'évaluation de la performance du moteur de recherche a été conduite parallèlement. Une méthodologie générique d'évaluation de la performance de SPIRIT a été proposée (Sanderson and Joho 2004) et raffinée dans (Aumaitre et al. 2004a). Ce processus a nécessité de construire une **collection test** composés de : 1) documents à indexer, 2) requêtes, 3) liste des documents pertinents pour chaque requête. Les collections tests sont un outil classique en recherche d'information dont la conception est une tâche si coûteuse qu'elle est mutualisée. Des campagnes de test sont organisées régulièrement dans ce domaine au cours desquels les développeurs de moteur de recherche peuvent tester leurs outils. La recherche d'information a progressé grâce à la mise en place de ces campagnes de tests permettant d'évaluer concurrentement plusieurs moteurs de recherche et de comparer les performances de techniques diverses. Bien entendu, il n'existait pas de collection test disponible pour un moteur géographique.

¹ <http://local.google.co.uk> : répertoire de type pages jaunes requêteable par un lieu et un thème.
<http://www.yell.co.uk> : les pages jaunes en ligne du Royaume Uni.
<http://www.map24.com> : application fournissant des cartes et des calculs d'itinéraires.

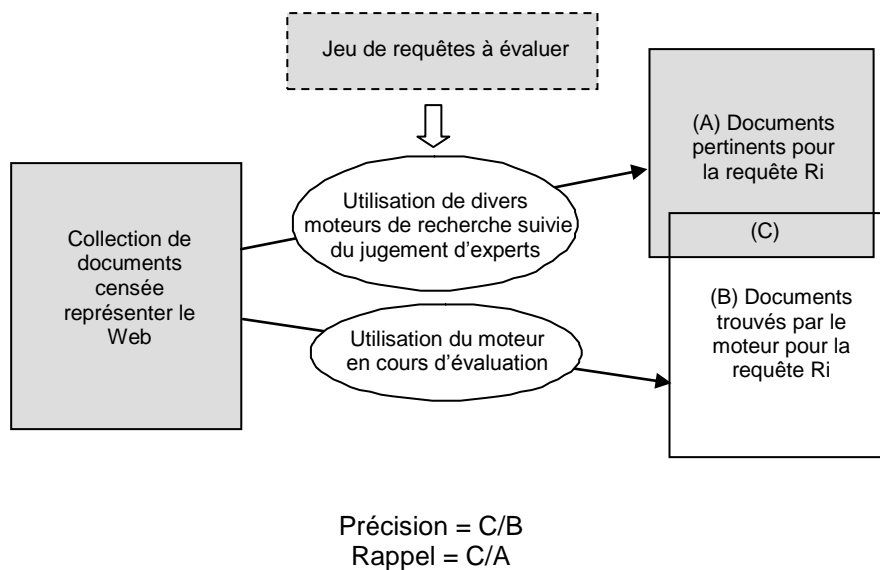


Figure 22. Eléments (grisés) d'une collection test utilisés pour évaluer les performances d'un moteur de recherche selon deux critères : la précision et le rappel.

La **constitution de la collection test** a soulevé de nombreux problèmes techniques (Bucher et al. 2005). Nous avons d'une part proposé de construire une collection test en utilisant comme requête des requêtes nécessitant une intelligence spatiale particulière (pour concentrer le test sur l'intelligence spatiale du moteur). Il s'agit des requêtes présentant les caractéristiques suivantes :

- Le nom de lieu est également un mot (ex : Battle en Angleterre)
- Le nom de lieu est ambigu (ex: Vienne)
- Le nom de lieu renvoie à une région imprécise (ex: le Sud de la France)
- Le nom de lieu est rare sur le Web (ex: blogs près de Vélioux)
- Le thème de la requête comporte un nom de lieu (ex: une forêt noire)
- La requête comporte une relation spatiale différente d'une inclusion (ex : des parcours de golf au nord de Saint Mandé)

Les requêtes de la Collection Test ont ensuite été construites en tenant compte des zones sur lesquelles nous possédions suffisamment d'information dans l'ontologie.

Nous avons ensuite défini une grille d'évaluation de la pertinence d'un document à destination des experts impliqués dans la constitution de la collection test. Cette grille comporte deux dimensions de pertinence.

- Pertinence thématique :

- Niveau 1 : le document est pertinent car il mentionne le concept.
- Niveau 2 : le document ne fournit aucune information sur le concept demandé.
- Pertinence spatiale
 - Niveau 1 : le document renvoie à une localisation qui est celle demandée ou une localisation très proche et vous estimez que le document comporte suffisamment de détails pour que vous trouviez ce lieu sur une carte de la zone en question.
 - Niveau 2 : le document renvoie à une localisation qui est celle demandée ou une localisation très proche (avec la même granularité spatiale) mais vous pensez qu'il n'y a pas suffisamment d'information dedans pour trouver ce lieu sur une carte.
 - Niveau 3 : le document ne correspond pas à la localisation recherchée.

Cette grille a été obtenue après une itération sur une première grille qui a été testée auprès de plusieurs évaluateurs. Le but de ce test était de vérifier que les évaluateurs étaient cohérents.

La collection test a ensuite été appliquée à plusieurs configurations de SPIRIT pour pouvoir évaluer plusieurs méthodes d'indexation (y compris l'indexation purement textuelle). Les résultats ont montré que dans les cas où l'index textuel seul donnait de bon résultat, l'ajout d'un index spatial améliorerait toutefois la 'précision aux 10 premiers documents'. Enfin, ils ont montré que dans certains cas, l'index purement textuel ne donnait aucun résultat alors que l'index spatial permettait de trouver des documents pertinents. Globalement, les résultats de cette évaluation étaient donc extrêmement encourageants.

1.2.2 Web sémantique

Cette section présente les principes clés du Web sémantique. Ceux-ci sont importants dans un contexte de mise en ligne d'information pour les raisons suivantes:

- Le Web sémantique est en passe de devenir une infrastructure de données distribuées (projet Linked Data décrit plus loin dans cette section).
- Le Web sémantique a proposé des langages issus des techniques d'intelligence artificielles dont les capacités sont très intéressantes dans un contexte de découverte, partage et intégration de ressources.

1.2.2.1 Langages du Web sémantique

Comme l'explique (Horrocks 2008), **le Web que nous connaissons est extrêmement simple**. Il se compose presque uniquement, d'un point de vue informatique, **d'hyperliens** (créé par les auteurs des documents) **et de chaînes de caractères**. Ce Web est riche en contenu mais actuellement, les moteurs de recherche s'appuient essentiellement sur les chaînes de caractère, les services s'appuient sur de la syntaxe. Ce constat est similaire de celui qui a conduit à concevoir le moteur de recherche SPIRIT pour prendre en compte l'aspect géographique du contenu du Web. Rappelons que l'approche de SPIRIT a consisté essentiellement en la création de métadonnées géographiques sur les documents et la construction d'une ontologie géographique (voir section 1.2.1).

Le Web sémantique vise à dépasser ces limites du Web actuel. C'est une vision exposée par (Berners-Lee et al. 2001) dans laquelle le Web est accessible aux machines, ou plus exactement la sémantique des contenus Web est accessible aux machines. Du point de vue des fournisseurs de contenu, le Web sémantique est le Web des machines –le Web dans lequel il faut qu'ils diffusent leur contenu pour des machines et non pour des humains-. Du point de vue des développeurs de logiciels, c'est le Web des données –où le mot données signale un contenu traitable par un programme-. Le Web sémantique ne correspond pas au Web 2.0, il est d'ailleurs parfois appelé Web3.0. Techniquement, (Berners-Lee et al. 2001) fondent le **Web sémantique sur l'intégration de techniques de modélisation venues de l'intelligence artificielle symbolique**. Celles-ci doivent permettre de développer des moteurs de recherche qui s'appuient non pas sur les mots mais sur le sens pour rechercher une ressource d'intérêt et intégrer plusieurs réponses.

RDF

Pratiquement, un premier niveau de Web sémantique consiste à '**annoter**' le contenu – en créant des données à propos de ce contenu qui seront, elles, manipulables par un programme. Cela correspond au principe essentiel de **RDF** (Resource Description Framework). Dans ce cadre, toute 'chose' qui peut être identifiée par un identifiant est une ressource et peut être décrite par des propriétés. RDF organise une description en triplets {uri, label, valeur} qui se lisent « la ressource identifiée par l'*uri* a pour propriété *label* la valeur *valeur* ». Les triplets se représentent facilement sous forme de graphes. Cela est illustré en Figure 23 et en Figure 25.

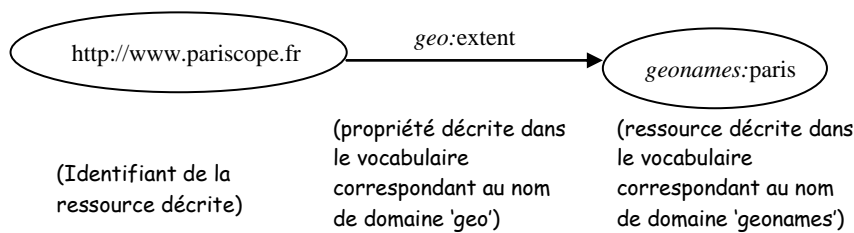


Figure 23. Schéma d'un énoncé RDF : le pariscope a pour couverture spatiale Paris.

Ainsi, RDF se distingue de XML par le fait qu'il correspond à une structure de graphe et non une structure d'arbre. Alors que XML est bien adapté pour décrire un document textuel structuré, RDF semble mieux adapté pour décrire des données ne serait-ce que pour écrire des relations entre les nœuds.

La valeur de la propriété peut être une ressource ou un littéral. La propriété peut être elle-même un mot ou une ressource. Généralement, des propriétés sont définies pour être utilisées pour décrire les ressources d'un domaine donné et le fait de les munir d'identifiants permet d'y faire référence dans des descriptions. De telles propriétés sont appelées dans leur ensemble un **vocabulaire de description**. Comme l'explique (Horrocks 2008) le recours à des vocabulaires spécifiques et partagés de description peut être suffisants dans des contextes spécifiques, par exemple dans le contexte du catalogage de documents textuels on trouve les propriétés définies par le Dublin Core. Un exemple de vocabulaire RDF est le Simple Knowledge Organisation System (SKOS). Depuis août 2009 il s'agit d'une recommandation du W3C pour partager des modèles statiques de connaissances sur le Web. La Figure 24 reprend un exemple de concept décrit dans SKOS, exemple proposé par Alistair Miles. D'autres exemples sont BasicGeo (pour décrire des coordonnées), Friend Of a Friend (FOAF) pour décrire des relations entre personnes ou le Dublin Core pour décrire des documents textuels. L'exemple suivant est extrait du tutoriel BasicGeo et utilise ces trois vocabulaires pour dire que la personne nommée Dan Brickley a pour homepage <http://danbri.org> et est localisée au point dont les coordonnées géographiques sont indiquées.

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:geo="http://www.w3.org/2003/01/geo/wgs84_pos#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns="http://xmlns.com/foaf/0.1/">
  <Person>
    <name>Dan Brickley</name>
    <homepage dc:title="Dan's home page"
      rdf:resource="http://danbri.org"/>
    <based_near geo:lat="51.47026" geo:long="-2.59466"/>
```

```
<rdfs:seeAlso rdf:resource="http://danbri.org/foaf.rdf"/>
<!-- more RDF here, using any RDF vocabularies -->
</Person>
</rdf:RDF>
```

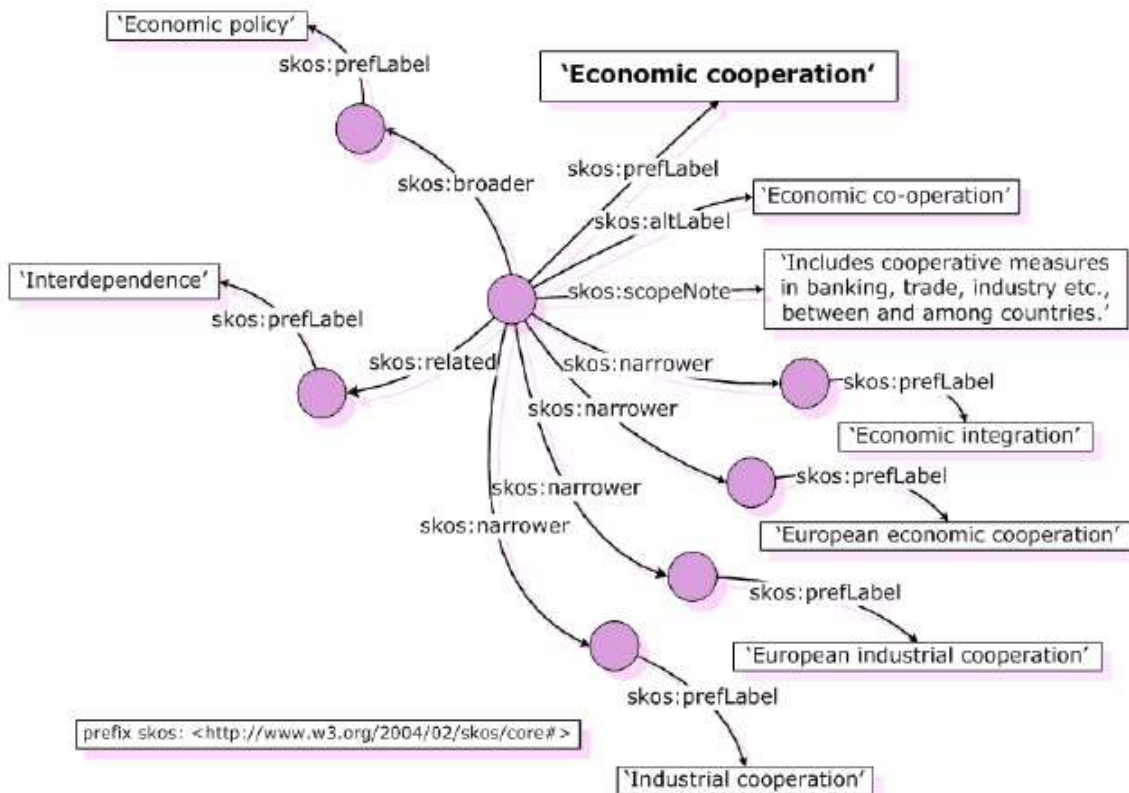


Figure 24. Description du concept étiqueté 'Economic cooperation' à l'aide de SKOS (repris d'un tutoriel d'Alistair Miles en 2008 sur SKOS).

RDF Schema

Il est parfois nécessaire d'étendre un vocabulaire ou encore de manipuler ensemble plusieurs vocabulaire. Pour cela, il faut définir pour le Web sémantique un **langage partagé de définition de vocabulaires** de description de ressources sur le Web. Le formalisme RDF-S vise justement à supporter l'écriture de vocabulaires pour décrire des ressources en RDF. RDF-S permet de définir des modèles conceptuels avec des hiérarchies entre classes et des hiérarchies entre relations. Un exemple d'ensemble d'énoncés RDF-S est illustré sur la Figure 25.

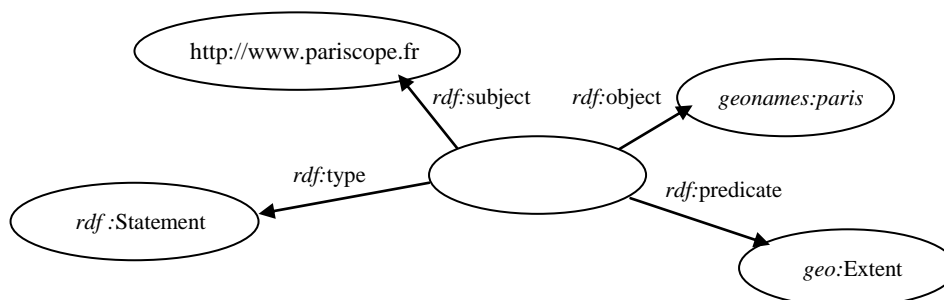


Figure 25. Description en RDFSchem de l'énoncé précédent. La ressource anonyme correspond à l'énoncé RDF en question.

OWL

L'intelligence artificielle offre davantage au Web sémantique pourvu que celui-ci puisse intégrer des langages plus avancés de description de contenu. Plus précisément ce sont actuellement les logiques de description qui doivent permettre au Web de remplir les promesses du Web sémantique (Horrocks 2008). **Les logiques de description** favorisent par exemple l'exploration de contenu dans un contexte où on ne sait pas construire une requête exacte car elles permettent les **réponses coopératives** (Benamara et al. 2003). Les auteurs décrivent plusieurs types de coopération. Une réponse coopérative peut par exemple décrire la violation par la question d'une présupposition existentielle. *Y a-t-il des villes de taille importante sur le bord de mer suisse ? Non, il n'y a pas de bord de mer suisse.* Elle peut aussi répondre par une approximation. *Les hôtels parisiens ont-ils le wifi ? Certains.* Elle peut aussi rappeler une contrainte générale que la question semble ignorer. *Y a-t-il une mairie à Montreuil ? Oui, toutes les communes ont des mairies.* Les logiques de description jouent également un rôle important dans l'intégration de contenus.

Dans le contexte du Web sémantique, un langage standard a été proposé pour écrire des ontologies : OWL. OWL vise à munir des contenus Web d'une sémantique formelle pour faciliter la découverte, la manipulation et l'intégration de ces contenus par des programmes. Le vocabulaire SKOS mentionné précédemment est par exemple défini à l'aide d'OWL.

La structure OWL est intéressante à bien des égards pour structurer une information géographique comparativement aux structures relationnelles largement utilisées. Un énoncé absent d'une base OWL est considéré comme 'non su' et non comme 'faux'. Une base OWL permet de mélanger des classes et des instances. Lors de l'analyse pour l'architecture TerraNumerica nous avons pu constater que cela pouvait s'avérer extrêmement pratique. L'expression de contraintes d'intégrité active est un problème ouvert des BDG. OWL permet d'exprimer les contraintes d'intégrité sous forme de

capacités d'inférences (Horrocks 2008). En OWL, un même objet peut avoir plusieurs identifiants. Cela semble utile pour se permettre de conserver dans une base des descriptions d'objets dont on ne sait pas encore nécessairement si ce sont les mêmes – mais qui le sont peut-être-.

1.2.2.2 Linked Data

Le Web sémantique compte d'ores et déjà de nombreuses réalisations. Il me semble que le plus important est le **projet LinkedData**.

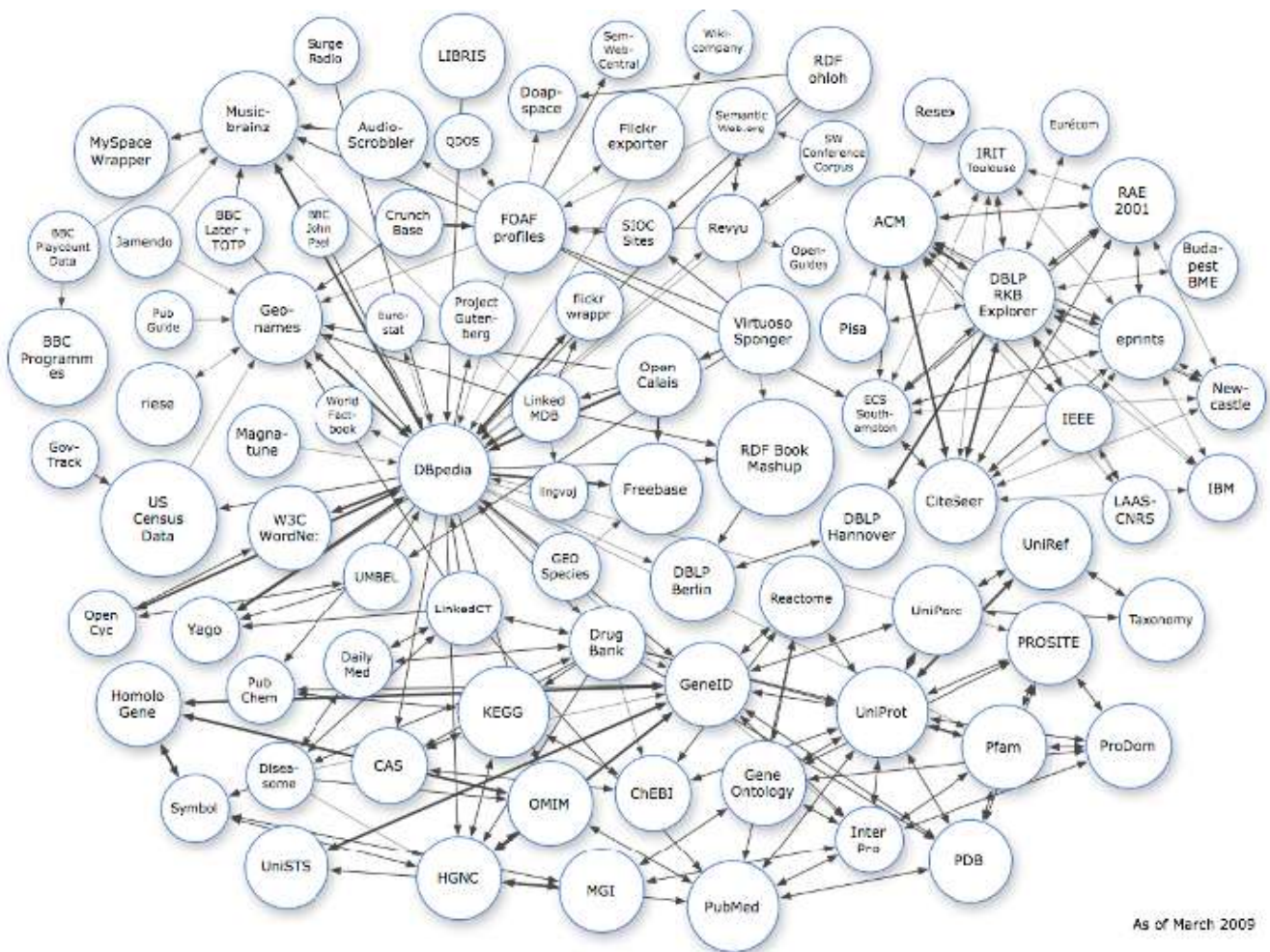


Figure 26. Ensemble de fournisseurs de contenu intégré dans l'architecture Linked Data (<http://linkeddata.org/>).

Celui-ci vise à mettre en place une interface de requête des contenus en ligne sur le Web conforme au langage SparQL. Le langage SparQL permet de faire des requêtes

pour extraire des sous-graphes de graphes d'origine. Le graphe d'origine est constitué de la base d'énoncés rdf et le sous-graphe à extraire correspond à la réponse à la requête. Le terme « nœud Sparql » renvoie à un fournisseur de contenu intégré dans l'architecture Linked Data. Un exemple de nœud SparQL est *DBPedia*. Cette base RDF est construite automatiquement à partir du contenu Wikipedia. Le Web devient ainsi un véritable système d'information distribué et permet de requêtes de la forme. La Figure 26 montre l'ensemble des fournisseurs de contenu qui ont intégré leur contenu dans l'architecture LinkedData. Voici un exemple de requête suggérée par Bernard Vatant de Mondeca : les noms en chinois de monuments remarquables de Paris, avec leur localisation geonames, et des images sur Flickr. Cette requête s'écrit comme suit :

```
PREFIX cat: <http://dbpedia.org/resource/Category:>
SELECT ?x ?nom ?geo ?images
WHERE {
    ?x skos:subject cat:Buildings_and_structures_in_Paris.
    ?x rdfs:label ?nom.
    ?x owl:sameAs ?geo.
    ?x dbpedia2:hasPhotoCollection ?images.
    FILTER ( lang(?nom) = "zh" ) }
```

Elle doit être écrite sur l'interface de requête de DBpedia : <http://dbpedia.org/snorql/>.

Et le résultat obtenu est affiché en Figure 27.

Soulignons que cette requête s'appuie sur l'existence d'une catégorie DBPedia 'Buildings_and_structures_in_Paris'. Autrement dit, l'aspect spatial n'est pas encore traité en tant que tel. Pour autant, RDF, SparQL et le projet Linked Data font peu à peu du Web une infrastructure de contenu qui devrait permettre de répondre à des requêtes complexes.

SPARQL results:

x	nom	geo
:Tour_Montparnasse	"蒙帕納斯大樓"@zh	< http://sws.geonames.org/6452893/ >
:Tour_Montparnasse	"蒙帕納斯大樓"@zh	< http://rdf.freebase.com/ns/guid.9202a8c04000641f800000000079b97 >
:Axe_historique	"历史轴"@zh	< http://rdf.freebase.com/ns/guid.9202a8c04000641f80000000006fb9 >
:Le_Bateau-Lavoir	"洗濯船"@zh	< http://rdf.freebase.com/ns/guid.9202a8c04000641f8000000000af140 >
:Val-de-Gr%C3%A2ce	"圣宠谷"@zh	< http://rdf.freebase.com/ns/guid.9202a8c04000641f80000000002b693a >
:Luxembourg_Palace	"卢森堡宫"@zh	< http://rdf.freebase.com/ns/guid.9202a8c04000641f8000000000de355 >
:Parc_des_Princes	"王子公园体育场"@zh	< http://sws.geonames.org/6457251/ >
:Parc_des_Princes	"王子公园体育场"@zh	< http://rdf.freebase.com/ns/guid.9202a8c04000641f8000000000349712 >
:Grand_Palais	"巴黎大皇宫"@zh	< http://sws.geonames.org/6452938/ >
:Grand_Palais	"巴黎大皇宫"@zh	< http://rdf.freebase.com/ns/guid.9202a8c04000641f80000000007fc5b0 >
:H%C3%B4tel_de_Ville%2C_Paris	"巴黎市政厅"@zh	< http://sws.geonames.org/3013131/ >
:H%C3%B4tel_de_Ville%2C_Paris	"巴黎市政厅"@zh	< http://rdf.freebase.com/ns/guid.9202a8c04000641f800000000056511d >
:Palais-Royal	"巴黎皇家宫"@zh	< http://rdf.freebase.com/ns/guid.9202a8c04000641f80000000003b3898 >

Figure 27. Réponse à la requête 'nom en chinois des monuments remarquables parisiens avec leurs photos flickr et leur localisation geonames'.

1.2.3 Le projet TerraNumerica

Le projet TerraNumerica est un projet labellisé du pôle de compétitivité Cap Digital visant à produire des données 3D sur l'île-de-France et à développer divers moteurs de rendu dédiés pour des applications de réalité augmentée, de réalité virtuelle ou de cartographie ubiquitaire. Les partenaires de ce projet étaient : une grande entreprise (THALES), des PME (Star Apic, Trimbe Mensi, Mondeca, Hyptique, Betomorrow, Tecdev) et des établissements et laboratoires publics (l'IGN, le GET, l'INRIA, l'ENSMP, l'ECP, l'Université Paris Est, le CITU, le LEDEN). Une tâche générale d'ingénierie système a rassemblé un membre de chaque partenaire. Deux groupes ont été constitués pour

étudier d'une part l'ingénierie de la plate-forme de production de données, et d'autre part, l'ingénierie des interfaces d'exploitation de ces données. J'ai participé à ce deuxième groupe. Par la suite, ma participation était liée au WP3.1.3.2 dirigé par Mondeca. Ce workpackage était consacré à la spécification de services pour permettre d'enrichir le monde 3D avec du contenu provenant d'autres sources.

1.2.3.1 Glossaire

Un premier apport du WP3.1.3.2 a été la **définition d'un glossaire** de termes utilisés dans TerraNumerica (Vatant et Bucher 2007a). Le besoin d'un glossaire a été ressenti lors des réunions d'ingénierie système, face à la **diversité des domaines et des vocabulaires techniques des partenaires** du projet. Ce glossaire a donc été ajouté comme livrable au WP3.1.3.2. Le glossaire contient 46 termes à utiliser de façon préférentielle dans les documents de communication internes et externes du projet. Des termes alternatifs ou synonymes sont également définis dans la mesure où ils correspondent à une pratique constatée.

Le terme *Modèle* est représentatif des ambiguïtés rencontrées lors de ce projet. Voici comment nous avons proposé de le définir dans ce glossaire :

Modèle : Un modèle est de façon générale une représentation construite pour décrire/imiter/numériser quelque chose, par exemple un bâtiment ou un processus. En informatique, on sort généralement les données constituant la représentation complète du sens du mot modèle. Le modèle est donc ce qui sert à structurer les données, une 'façon de représenter'. Un modèle est typiquement un ensemble de classes, de propriétés de ces classes et d'associations entre classes. Il s'appuie sur des domaines de valeurs (comme les entiers, une liste énumérée de valeurs, ...).

Discussion : ATTENTION TERME AMBIGU DANS TERRANUMERICA !!! Certains utilisent modèle pour désigner un ensemble de données (un modèle 3D), d'autres pour désigner une méthode de dérivation (un modèle de végétation) d'autres pour désigner un élément d'une méthode de synthèse (le modèle générique), d'autres pour désigner un schéma conceptuel.

Dans ce document, on suggère, sauf dans les cas où une expression consacrée existe, de recourir à la terminologie de substitution suivante:

- 'façon de modéliser' = modèle conceptuel, structure, schéma conceptuel ou schéma de données.
- 'données modélisant un ou plusieurs phénomènes' = représentation, modèle d'un objet précis.
- 'méthode' = modèle de dérivation.
- 'modèle générique' = modèle générique, modèle géométrique d'un type d'objet.

D'autres définitions intéressantes issues de ce glossaire sont celles relatives à la notion de référentiel. Ce mot était souvent utilisé dans les discussions d'ingénierie système mais s'est révélé ambigu.

Référentiel : Système servant à spécifier/construire/identifier des références. La notion de référentiel renvoie donc à la propriété d'une ressource : pouvoir servir à spécifier des références. Mais elle ne dit pas forcément quelque chose sur la nature de cet objet (ce qu'est un référentiel).

En particulier : référentiel ontologique, référentiel 3D, référentiel spatial.

Référentiel ontologique : Le référentiel ontologique est une ontologie qui est reconnue comme une (bonne) référence sémantique dans Terra Numerica : une ontologie pivot, ou alors une ontologie de domaine spécifique rattachée à l'ontologie pivot.

Référentiel spatial : Système permettant de définir des localisations indirectes (appelé rigoureusement 'système de référence spatial') par opposition aux systèmes de coordonnées géographiques qui fournissent des localisations directes.

Un tel système permet en pratique d'associer à un objet symbolique de type simple (chaîne de caractères), comme par exemple un nom de pays, une position. Des exemples de référentiels spatiaux sont les pays, les entités administratives, les adresses.

Référentiel 3D : Représentation en 3D d'objets du terrain permettant de construire des références spatiales 3D d'objets. Les objets en question et leur représentation ont des propriétés qui leur confèrent ce statut de référentiel.

- *Phénomène de référence* : Les objets représentés sont des objets géographiques servant classiquement à se repérer dans l'espace, comme les routes et l'emplacement au sol des bâtiments.

- *Propriétés de référence* : La représentation de ces objets se concentre sur des propriétés qui permettent de reconnaître ces objets (le numéro d'une route, l'adresse d'un bâtiment) et également de les localiser dans un système de référence spatial direct (un référentiel mathématique).

- *Représentation de référence* : De plus cette représentation présente des caractéristiques de qualité 'garanties' par le fournisseur. A cet égard, le référentiel doit être mis à jour régulièrement au fur et à mesure de l'acquisition de nouvelles données sources. Autrement dit, il peut être vu davantage comme un service de fourniture de données que comme des données figées. Outre le référentiel proprement dit, des données particulièrement importantes à gérer sont les métadonnées sur les données sources servant à construire le référentiel, et les métadonnées d'accès au référentiel.

Plusieurs référentiels 3D peuvent cohabiter.

1.2.3.2 Un processus générique d'enrichissement

Les modèles et techniques pertinents pour permettre l'intégration de contenu dans une plateforme de données géographiques (3D en l'occurrence mais cela pourrait être aussi 2D) proviennent du domaine de la géomatique et du domaine du Web (Vatant et Bucher 2007b). Nous avons proposé un **processus extrêmement générique d'enrichissement**, le premier à notre connaissance dans le domaine de la géomatique, qui intègre ces deux domaines et s'appuie sur une architecture à base de services

(Bucher et Vatant 2008). Une variété de scénarios d'enrichissement ont été pris en compte:

- l'IGN souhaite enrichir TN avec une représentation vecteur du réseau routier,
- une DDE souhaite publier un projet d'aménagement de rond-point,
- la mairie de Paris souhaite faire apparaître dans une visualisation TN les bornes vélib' avec la consultation en temps réel des vélos disponibles,
- un architecte désire ajouter son cv et les projets sur lesquels il a travaillé.

Le point commun de tous ces scénarios est que l'information que souhaite apporter le fournisseur est une information localisée qui concerne la zone couverte par TN. Mais les modalités spécifiques d'interfaçage requises par chaque scénario diffèrent un peu. Premièrement, **le mode d'hébergement du contenu peut différer d'un processus à un autre**. Le fournisseur peut souhaiter conserver ses données chez lui, ou les transférer sur un serveur de contenu TerraNumerica ou enfin ce contenu peut correspondre à la sortie d'un service de dérivation. Deuxièmement, **le mode de liaison du contenu avec les données TerraNumerica peut différer**. Nous avons prévu que le fournisseur de contenu pouvait vouloir simplement recaler ses données ou créer des liens entre ses données et les données TerraNumerica.

Nous détaillons dans la suite de cette section notre processus générique d'enrichissement et les services sous-jacents. La Figure 28 reprend l'ensemble des services mentionnés dans (Bucher et Vatant 2008).

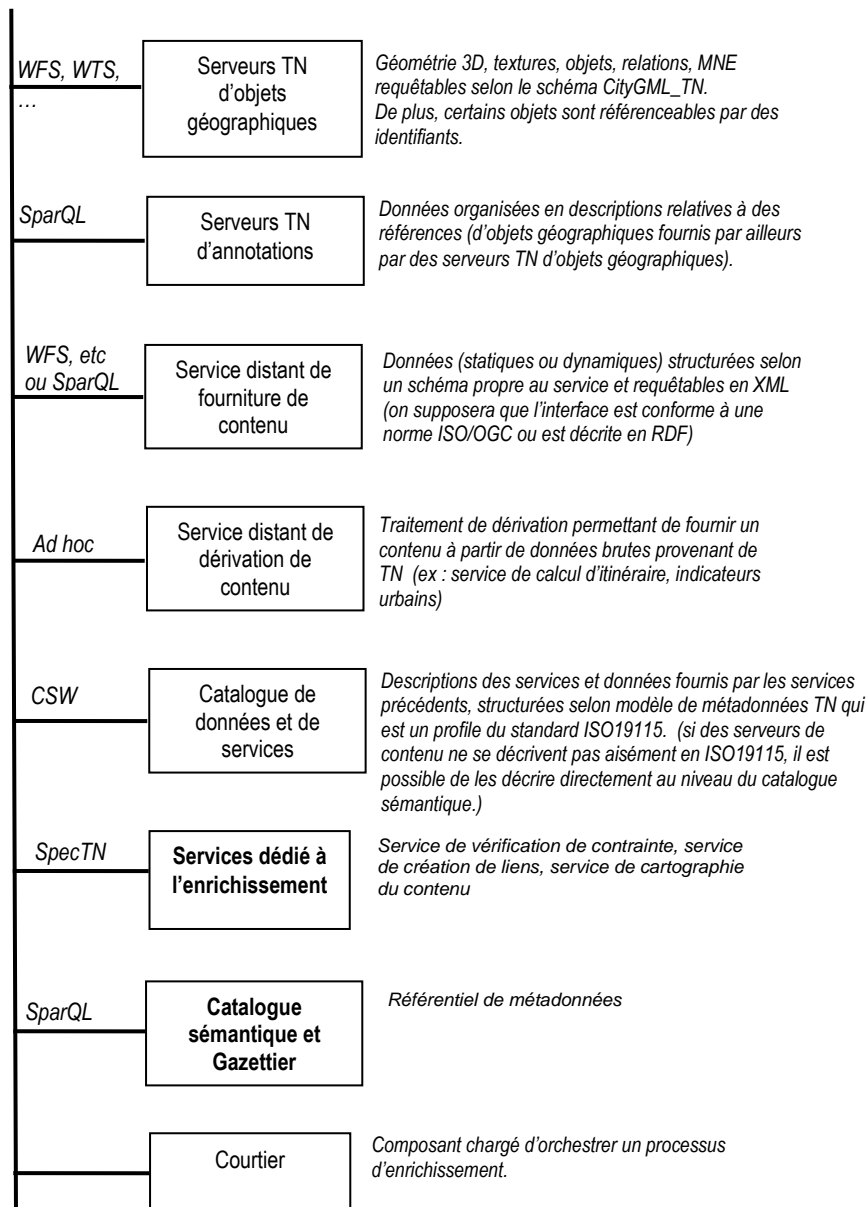


Figure 28. Architecture de services d'enrichissement de TerraNumerica. Les services en gras sont ceux qui ont été prototypés dans le contexte du WP3132.

1.2.3.3 Référencement dans les métadonnées

La première étape est le **référencement dans les bases de métadonnées TerraNumerica**. Autrement dit, le fournisseur doit fournir des métadonnées à propos de

ses données et ces métadonnées doivent être conformes au modèle de métadonnées TN. Ainsi, si deux fournisseurs enrichissent TN avec des informations à propos de la destination (i.e. la fonction) des bâtiments, même s'ils utilisent des modèles différents et spécifiques pour structurer leurs données, ils doivent tous les deux dire que ces informations portent sur la destination des bâtiments, concernent telle ou telle zone etc, et pour le dire ils utilisent un modèle commun : le modèle de métadonnées TN. Nous avons proposé comme modèle de métadonnées d'adopter les éléments 'cœur' de ISO19115 et d'ajouter des éléments utiles à la description de la structure d'un jeu (Bucher et Vatant 2008) :

- renseigner l'attribut ContentInformation_FeatureCatalogue sous la forme d'une liste d'uri vers les FeatureType définis dans le référentiel de métadonnées (ex: CityGML.ReliefFeature, CityGML_TN.Contexte_Existence).
- renseigner l'attribut ApplicationSchema en précisant une liste de contraintes d'intégrité (ou de critères de création de liens exprimées en OWL).

Cette première étape doit s'appuyer sur deux services:

- le catalogue sémantique qui comporte la spécification des types de contenu et
- le catalogue de données et services géographiques qui permet d'héberger des métadonnées ISO19115 et de pointer vers les sources de données.

La consultation du catalogue sémantique permet de connaître les types de contenus disponibles (et d'examiner s'il existe déjà dans ce catalogue des concepts utilisables pour référencer le nouveau contenu ajouté). Elle permet aussi d'enrichir ce catalogue sémantique si besoin. Nous préconisons l'usage de RDF-S comme structure du catalogue sémantique. Ce langage permet par exemple de pouvoir utiliser au niveau du catalogue des concepts (comme le concept de monument) et des instances (comme la Tour Eiffel). C'est intéressant car certaines instances sont aussi utiles que les concepts pour spécifier un contenu.

Concernant l'implantation, nous avons proposé l'utilisation de Sesame ou du logiciel ITM de Mondeca pour le catalogue sémantique et l'utilisation de MDWeb (<http://www.mdweb-project.org/>) pour le catalogue géographique structuré. MDWeb a été préféré à Geosource car il présente un modèle de métadonnées extensible qui permettait donc d'ajouter des champs spécifiques à TerraNumerica. Un prototype de tels services a également été développé au COGIT (Jolivet et Bucher 2008). Laurence Jolivet a implanté les fonctions suivantes du catalogue sémantique : l'utilisateur consulte une ontologie et annote éventuellement les éléments de son schéma avec les concepts de cette ontologie. Le client graphique affiche les éléments de schéma du jeu de l'utilisateur, lui demande de signaler quels éléments il veut partager et éventuellement d'annoter la classe et les attributs à l'aide d'une taxonomie. Cette taxonomie correspond à l'ontologie produite par l'action de recherche sur l'intégration.



Figure 29. Client graphique d'annotation des données de l'utilisateur.

1.2.3.4 Mise en cohérence des données

Une deuxième étape est la **mise en cohérence au niveau des données**. Le fournisseur précise par exemple que les géométries 3D des objets doivent être 'posées sur le sol' et non flotter en l'air ou disparaître en sous-sol. Cette opération peut être utile pour l'affichage ou le requêtage des données dans l'environnement TN. Cette étape peut être conduite de deux façons : soit par la simple évaluation de contraintes d'intégrité entre le contenu du fournisseur et le contenu TerraNumerica soit par le recalage.

1.2.3.5 Création des liens (références internes ou externes)

Une troisième étape consiste à **créer des liens au niveau des données entre le contenu et les données TerraNumerica existantes**. Cet enrichissement répond aux exemples de besoin suivant. Le premier exemple est celui de l'IGN qui souhaite ajouter des tronçons de rue vecteur dans une représentation de type Bati3D et créer des relations 'tronçon_rue donne_accès_à bâtiment' entre les rues ajoutées et les objets bâtiments existants dans Bati3D. Le deuxième exemple est celui d'une association qui souhaite indiquer les architectes ayant construit des bâtiments parisiens représentés dans TN. Nous proposons deux modes de création de liens :

- liens d'objet à objet (dans ce cas le contenu ajouté doit être stocké avec le contenu existant). C'est ce que nous appelons référence interne.
- liens d'objet à identifiant. C'est ce que nous appelons référence externe.

Nous proposons également que les liens soient calculés lors de l'enrichissement ou que le calcul du lien puisse se faire à la demande, lors de la consultation.

Dans le prototype développé par Laurence Jolivet, ce service est implémenté de la façon suivante (Jolivet et Bucher 2008). Le fournisseur spécifie un schéma de l'enrichissement

selon le modèle présenté sur la Figure 30. Ce schéma est constitué de la structure du contenu envoyé sur le serveur, indiquant les éléments qu'ils souhaitent effectivement publier dans TerraNumerica et indiquant les liens entre ces éléments et des éléments de schéma d'objets TerraNumerica existants. L'utilisateur précise aussi des critères de calcul de ces liens. Ce peut être une égalité d'attribut ou une proximité géométrique.

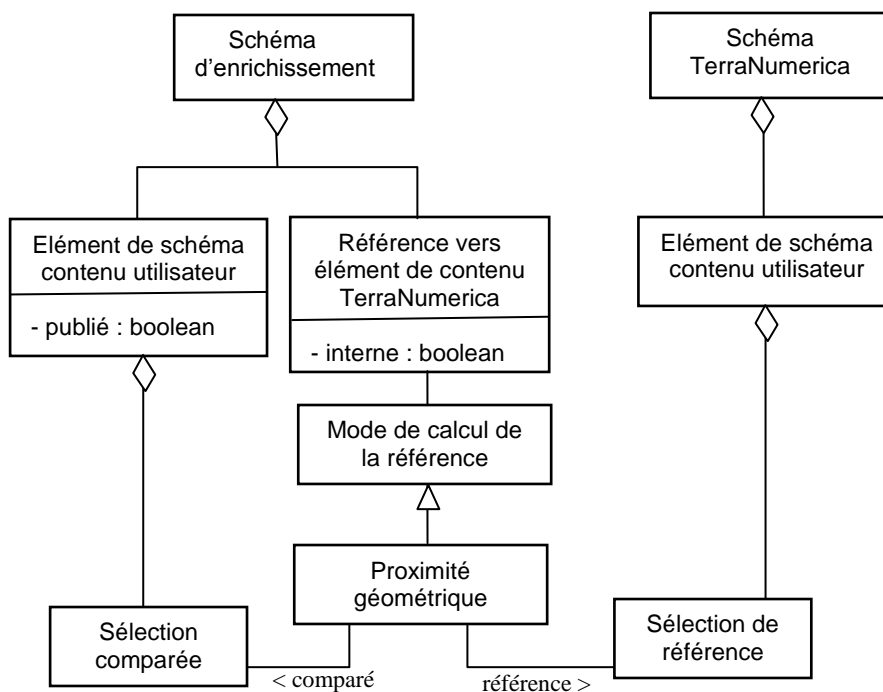


Figure 30. Modèle de spécification d'un enrichissement avec création de liens entre le contenu ajouté et le contenu TerraNumerica.

Laurence Jolivet a implémenté un service et un client graphique qui permet de construire une telle requête d'enrichissement. Son service utilise un appariement simple de données basée sur un critère de proximité : la distance euclidienne dans le cas de données ponctuelles, la distance de Hausdorff pour des données linéaires et la distance surfacique pour des surfaces.

1.2.4 Les projets dédiés à la création collaborative de contenu

Cette dernière section détaille quelques éléments d'une fonctionnalité qui doit être prise en compte dans les infrastructures de données géographiques. Il s'agit de la création de contenu par la personne qui consulte l'infrastructure. Elle comporte une courte analyse de l'existant et un pointeur vers une thèse qui commence tout juste sur ce sujet au COGIT.

La **création collaborative de contenu** sur le Web est un domaine qui prend une importance croissante. En effet, une notion-clé de Web – parmi celles ayant conduit à l'appellation Web2.0- est la suivante: celui qui consulte peut également contribuer au contenu. Bien plus, ce contenu peut être construit de façon collaborative avec d'autres contributeurs. Le projet de création collaborative de contenu sur le Web le plus connu est probablement *Wikipedia*, l'encyclopédie en ligne. Un nouveau terme est apparu, celui de néogéographes qui renvoie aux personnes utilisant les outils du Web2.0 pour éditer des données géographiques. Nous pouvons observer tout d'abord des projets croissants de localisation de contenus existants. Un projet actuel de *Wikipedia*, '*Wiki project on geographical coordinates*', vise à mettre en place une façon homogène de construire des métadonnées de localisation pour les articles. Ainsi, *Wikipedia* contient de plus en plus de documents localisés. *Flickr* propose un outil de localisation des photographies qui s'appuie sur l'API Google¹ et étudie aussi le recours à des techniques automatique pour localiser automatiquement une image d'après son contenu. Il y a également un nombre croissant de projets de création de contenu localisé comme par exemple UCrime qui vise à constituer une cartographie des 'crimes' (au sens large) dans le voisinage des Universités américaines. Dans ces projets, (Coote and Rackham 2008) distinguent les groupes d'intérêt qui se focalisent sur un domaine précis (la randonnée, les agressions..) et les communautés locales qui se focalisent sur un lieu relativement petit et, sur ce lieu, sont intéressées sur une thématique vaste ([http://www. OpenMapSihlwad.ch](http://www.OpenMapSihlwad.ch)). Cette tendance a conduit (Goodchild 2007) à parler de 'volunteered geography' où le citoyen se voit comme un capteur (c'est le titre de son article) et permet à d'autres utilisateurs d'exploiter ses capacités d'acquisition.

¹ Le contributeur voulant annoter sa photographie navigue dans une cartographie GoogleMaps jusqu'à désigner le point géographique auquel il a pris la photographie en question



Figure 31. Carte du crime sur l'American University à Washington.

Parmi ces projets, *GeoNames* est un exemple intéressant. Il a été conçu dès le départ comme un gazetier libre de droit et collaboratif et est devenu une référence en la matière. *Wikimapia* est un autre projet dont le but annoncé est de cartographier de façon collaborative des lieux, c'est-à-dire de construire un gazetier. *Wikimapia* permet de saisir une forme géométrique et non uniquement un point comme illustré sur la Figure 33. De plus, l'intérêt de ce projet est l'aspect wiki : l'interface de saisie est extrêmement simple. Enfin, les contributeurs peuvent saisir des commentaires à propos d'un lieu et surtout peuvent collaborer pour définir ensemble ce lieu comme illustré sur la Figure 34 où des contributeurs discutent pour définir le meilleur polygone décrivant Paris. Mais *wikimapia* est moins complet que *GeoNames*.

Un autre projet phare est *OpenStreetMap*. Ce projet s'est créé en réaction à l'absence de données libres de droits. Il invite toute personne à contribuer à l'acquisition de données géographiques libres de diverses façons. La plus répandue est de charger des informations géométriques sous forme de traces GPS, ou des informations sémantiques comme des noms de lieux. Certains membres scannent et assemblent des cartes, d'autres font de l'orthorectification, du dallage.

Ces exemples illustrent les différences de collaborativité qui existent entre de tels projets. (Noucher 2009) insiste sur l'importance de distinguer plusieurs types de création collaborative de contenu pour traiter différemment diverses pratiques et diverses attentes:

- Des utilisateurs produisant seuls pour les autres. Cela s'appelle le crowd sourcing. Les contributions sont agrégées de façon indépendante des utilisateurs (comme sur *OpenStreetMap*).
- Des utilisateurs coproduisant des données ensemble. Il s'agit de démarches participatives (comme sur *Wikimapia*).

LES INFRASTRUCTURES D'INFORMATION GEOGRAPHIQUE

Les infrastructures de données géographiques

GeoNames Home | Postal Codes | Download / Webservice | About

Search: Country:

[advanced search](#)

9 records found for "saint gildas"

Name	Country	Feature class	Latitude	Longitude
1 Saint-Gildas-de-Rhuys 36730, Abeland, Abeland, Lohentag, Saint-Gildas, Saint-Gildas de Ruy, Saint-Gildas-de-Ruis, Saint-Gildas-d...	France, Brittany	populated place population 1.333	N 47° 30' 0"	W 2° 50' 0"
2 Saint-Gildas-des-Bois 44530, Gweltaz-Lambourig, Saint-Gildas, Saint-Gildas-des-Bois	France, Pays de la Loire	populated place population 3.239	N 47° 30' 58"	W 2° 2' 11"
3 Saint-Gildas 22200, Gildas-du-Chateau, Saint-Gildas, Sant-Weltaz	France, Brittany	populated place population 251	N 48° 25' 0"	W 3° 0' 0"
4 Pointe de Saint-Gildas	France, Pays de la Loire	point	N 47° 8' 0"	W 2° 15' 0"
5 Ile Saint-Gildas	France, Brittany	island	N 48° 51' 0"	W 3° 18' 0"
6 Saint-Gildas-des-Bois	France, Pays de la Loire	fourth-order administrative division	N 47° 30' 58"	W 2° 2' 11"
7 Saint-Gildas	France, Brittany	fourth-order administrative division	N 48° 25' 0"	W 3° 0' 0"
8 Saint-Gildas-de-Rhuys	France, Brittany	fourth-order administrative division	N 47° 30' 0"	W 2° 50' 0"
9 Savins de Saint-Gildas	France, Pays de la Loire	forest(s)	N 47° 32' 0"	W 2° 1' 0"

GeoNames Home | Postal Codes | Download / Webservice | About

Map center: N 47° 30' 0" W 2° 49' 58"

ExploreOurPlace.net google earth tag

Plan Satellite M...

Scale: 2 mi / 2 km

Imagerie ©2009 TerraMetrics, Données cartographiques ©2009 Tele Atlas

GeoNames Wikipedia

Features

- city, village, ...
- mountain, hill, rock, ...
- stream, lake, ...
- country, state, region, ...
- parks, area, ...
- road, railroad
- spot, building, farm
- forest, heath, ...
- undersea

Name	country	feature	km to center
1 Saint-Gildas-de-Rhuys	France	fourth-order administrative division	0.02 km
2 Saint-Gildas-de-Rhuys	France	populated place	0.02 km
3 Pointe du Grand Mont	France	point	1.28 km
4 Anse de Cornault	France	coast(s)	2.25 km
5 Pointe de Saint-Jacques	France	point	3.1 km

Figure 32. Informations fournies par Geonames à propos du nom Saint Gildas (fenêtre du haut) puis du nom de lieu Saint Gildas de Rhuys plus précisément (fenêtre du bas).



Figure 33. Description du lieu Saint Gildas de Rhuys dans wikimapia. Il ne s'agit pas d'une empreinte ponctuelle, ni des véritables limites de la commune.

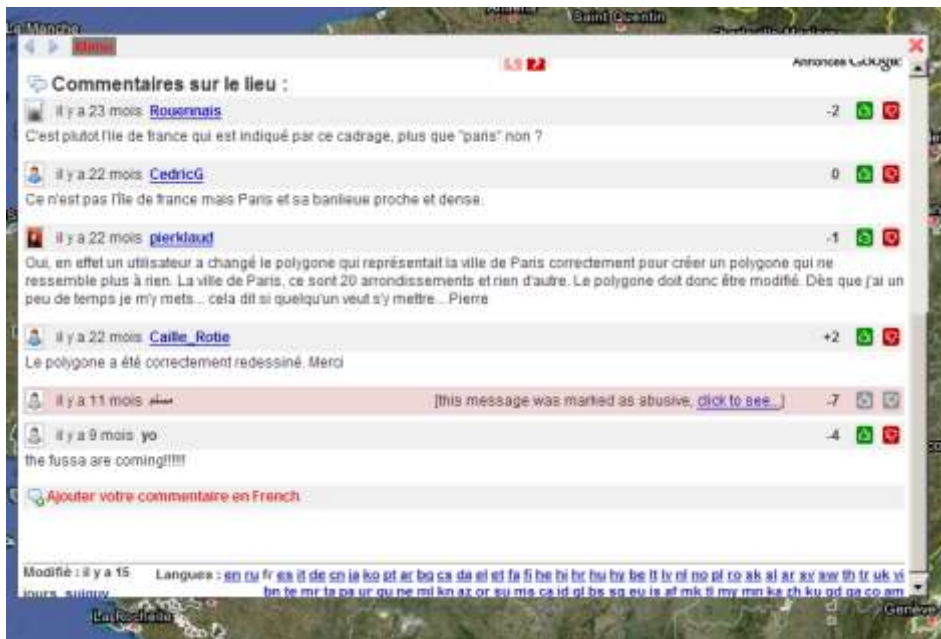


Figure 34. Commentaires associés au lieu 'Paris' sur Wikimapia.

La validité des informations construites est un thème récurrent. Le succès de *Wikipedia* a montré la possibilité de mettre en place des modèles de contrôle et validation suffisamment efficaces pour susciter une confiance certaine dans la valeur des articles *Wikipedia*. Le principal modèle de contrôle de la validité consiste à donner divers profils aux contributeurs et divers statuts aux contributions. Dans la plupart des systèmes de gestion de contenu, comme SPIP, une contribution peut être en attente de validation ou validée et en ligne. Dans *wikipedia*, parmi les contributeurs certains sont

modérateurs et ont pour rôle de valider les informations avant que celles-ci soient en ligne.

La **qualité** de données géographiques est l'adéquation de la représentation à l'usage voulu. Cette adéquation dépend des choix de représentation (voir le début de la section 1.1.4) et également du fait que le monde a été correctement observé suivant ce processus. **Classiquement, pour simplifier l'étude de la qualité de données géographiques, on distingue donc deux aspects de cette qualité :**

- La **qualité externe** qui mesure l'adéquation entre la représentation du monde indiquée dans les spécifications et le besoin.
- La **qualité interne** qui mesure l'écart entre la représentation du monde portée par les données et la représentation désignée par les spécifications.

Concernant la qualité externe, il semble que **les choix de représentation** faits par ces communautés pourraient être améliorés. Par exemple, sur wikimapia, la notion de lieu n'est pas claire car deux lieux peuvent apparaître avec des descripteurs différents alors qu'ils semblent être un lieu unique dans la réalité, par exemple Saint Gildas en tant que commune et Saint Gildas en tant que lieu habité. Il semble qu'un meilleur choix de modélisation consisterait à utiliser les classes 'entité administratives' ou 'lieu habité' comme des propriétés qui peuvent exister simultanément et non comme des classes de haut niveau. Dans OSM, le modèle est basé sur RDF. Un objet est décrit par des triplets de propriétés appelés tags. Un modèle de tag est proposé sur le site et peut être étendu par les contributeurs. Ce qui est frappant c'est que la ressource qui porte les triplets n'est pas l'objet géographique mais la géométrie. Parfois la géométrie d'un objet est modifiée alors que cet objet demeure, par exemple si on ajoute une extension à un bâtiment ou si on élargit une chaussée. Cette représentation d'OSM ne facilite pas par ailleurs des modélisations importantes en géomatique qui sont la représentation multiple, le partage de géométrie.

Par ailleurs, un élément nécessaire à l'évaluation de la qualité externe et interne est la **connaissance des spécifications du produit**. *Wikimapia* propose à sa communauté une charte de sélection des lieux qui ont du sens et peuvent figurer sur la carte (http://wikimapia.org/wiki/User_Guide:_Places_and_Place_Tags#Can_I_add_my_own_house.3F). Les principales règles sont les suivantes:

- 1) *A place may be added for existing, non-movable objects that are visible on the satellite view.*
- 2) *A place may be added for existing buildings that are not visible on the satellite view.*
- 3) *A place may be added for settlement boundaries such as cities, villages and neighborhoods if a category is properly assigned to the place.*
- 4) *A place must have at minimum, a properly constructed outline.*
- 5) *Places that do not physically exist are not permitted.*

6) A place may be added for a movable object if it is well-known, famous or interesting to a large audience and it does not clutter the map or visually conflict with other places.

7) A place may be added for an existing historical place if it is well-known, famous or interesting to a large audience and it does not clutter the map or visually conflict with other places.

Wikimapia appelle ses contributeurs à désigner et détruire les places ne répondant pas aux critères (<http://wikimapia.org/wiki/Deletion>).

UCrime propose aussi une classification des crimes reprise sur la Figure 35.

Crime Classifications
* Theft - The act in which property belonging to another is taken without that person's consent.
* Burglary - The criminal offense of breaking and entering a building illegally for the purpose of committing a crime in that building.
* Robbery - Using force or intimidation to take property away from another person in the presence of that person.
* Assault - A physical attempt or threat to use violence with the intent to do harm to another.
* Arson - The crime of intentionally setting fire to a building or property of another or the burning of one's own property to collect insurance.
* Shooting - The act of firing a weapon in order to hit, wound, or kill someone or something.
* Vandalism - The intentional destruction of or damage to the property of another.
* Arrest - The seizure of an alleged or suspected offender to answer for a crime.

Figure 35. Classification proposée par UCrime pour ses données sur le crime.

Globalement, nous pensons qu'il est possible d'améliorer l'expression de spécification de tels contenus pour améliorer la qualité et l'évaluation de la qualité de ces contenus. C'est pourquoi j'ai proposé un sujet de thèse sur cette problématique. Il s'agit de la thèse de Carmen Brando à l'Université de Marne la Vallée dirigée par Anne Ruas et que je co-encadre. L'objet de cette thèse, qui a début en octobre 2009, est de proposer une méthode permettant à une communauté de construire collaborativement un contenu géographique. Les verrous scientifiques abordés sont de permettre aux contributeurs de munir le contenu de spécifications d'acquisition et de contenu et de qualifier la donnée.

Enfin, concernant la qualité interne, il existe d'ores et déjà une étude sur la qualité des données OSM fondée sur la comparaison avec celles de l'Ordnance Survey (Haklay 2008). L'auteur constate des différences de qualité selon certains types de zones. Sur les zones où OSM a bénéficié de plusieurs contributeurs et d'un certain processus de

validation, les données sont de qualité comparable avec celles de l'Ordnance Survey. L'auteur relève les zones les mieux couvertes correspondent aux zones aisées. Sur UCrime, on peut constater que certaines universités sont vierges de tout crime alors que d'autres en affichent plus de 50. Cela peut s'expliquer dans la réalité (voisinage délinquant ou non, un même individu qui perpétue plusieurs crimes) mais aussi par un biais dans l'acquisition (dans certains lieux, UCrime n'est peut-être pas connu et n'est pas utilisé pour rendre compte des crimes).

1.3 Bilan

Cette première partie a donné un aperçu de la variété des domaines techniques relatifs à la diffusion d'information géographique en ligne. Les domaines mentionnés ont été abordés à l'occasion de collaborations sur des projets ou de travaux de stage. Pour des domaines nécessitant de nouvelles compétences techniques que celles possédées par un ingénieur de l'IGN, les collaborations présentent un intérêt indéniable. SPIRIT m'a permis de mieux connaître le domaine de la recherche d'information et TerraNumerica (plutôt le WP 3.1.3.2 de TerraNumerica) m'a permis de mieux connaître le Web sémantique. Ce réseau scientifique doit être utilisé pour essayer de mettre en relation les modèles des SDI et les modèles pour les infrastructures d'information provenant du Web comme Linked Data. Cette approche est déjà amorcée par des chercheurs de l'Ordnance Survey (Goodwin et al. 2009).

La méthode de transformation de schéma de Sandrine Balley peut être étendue pour répondre aux besoins des SDI en méthodes automatiques et dynamiques d'extraction, d'intégration, de mise à jour et d'interopérabilité entre données et traitements.

Concernant l'extraction, il me semble la proposition de Sandrine peut être enrichie, en intégrant effectivement un niveau ontologique et en intégrant le pilotage d'algorithme d'analyse spatiale à l'instar de la proposition de (Lüscher et al. 2008). Elle peut aussi être améliorée en l'associant à un langage assertionnel de spécification d'une extraction à l'instar de la proposition (Staub et al. 2008). Enfin, d'autres sources d'information peuvent être ajoutées à une interface de recherche et extraction. Il s'agit d'une part des contenus textuels du Web et des contenus structurés en RDF du Web sémantique. Il s'agit d'autre part de données image et de mesures issues de capteurs.

Concernant la fusion de données et la propagation de mises à jour d'un jeu du producteur au jeu de l'utilisateur, la proposition de transformation de schéma dans la thèse de Sandrine Balley pourra être combinée avec la proposition en cours de la thèse

de Nathalie Abadie qui vise à établir les correspondances entre des schémas (Abadie 2009). Si des correspondances entre schémas sont interprétables en un processus de transformation automatique de données, cela simplifie la fusion des données et le calcul de différentiels dans le modèle utilisateur.

Par ailleurs, sa proposition a également des applications pour le traitement en ligne de données géographiques via des Web Processing Services (WPS) (section 2.1.3). Pour utiliser, voire composer, de tels services il est souvent utile d'adapter les données à traiter au modèle de données du traitement. Nous avons élaboré un début de proposition dans ce sens qui vise à permettre au fournisseur d'un WPS de publier un prétraitement associé à son service qui adapte les données à partir d'une structure pivot (Bucher and Balley 2007).

L'analyse fonctionnelle conduite dans le contexte de TerraNumerica sur l'intégration de contenu dans un référentiel est également un apport précieux. Elle doit être reprise pour comprendre les **besoins d'intégration dans les infrastructures d'information et élaborer des réponses flexibles**. A cet égard, il me semble qu'un outil nécessaire est l'existence d'objets géographiques de référence. Que le modèle soit vecteur ou non, il est utile que les objets susceptibles d'être utilisés comme des références apparaissent en tant qu'objet et aient des identifiants. C'est important pour faciliter la mise à jour, pour attacher du contenu, pour indexer cette information, et, avec l'informatique ubiquitaire, pour relier la représentation au monde. Ce qui a évolué c'est que les utilisateurs ont besoin de pointeurs stables et utilisables par eux et non seulement de méthodes mises en œuvre par le producteur de données. Une proposition intéressante est celle de Takashi Morita et de carte à échelle réelle. Cette proposition vise à mettre en place un système de référencement indirect s'appuyant sur des objets de référence visibles dans le monde réel et sur des relations de positionnement manipulables par l'homme non expert (Morita 2007).

Un autre apport décrit précédemment consiste en notre proposition d'une méthode **d'aide au changement de niveau de détail**. Cette proposition vise à atteindre des procédés de 'collage spatiaux' pour reprendre l'expression de (Claramunt and Mainguenaud 1996) dans lesquels l'utilisateur passe en continu d'une appréhension locale à une appréhension globale. Notre travail s'appuie sur l'intégration dans une vue à une certaine échelle d'éléments graphiques faisant référence à l'autre échelle. Une innovation importante consiste à intégrer le point de vue de l'utilisateur dans la construction d'une représentation synthétique. Par exemple, un voisinage est décrit en fonction d'une répartition angulaire autour d'un point central qui est la position de l'utilisateur.

Le travail conduit dans SPIRIT sur la **spécification des besoins et l'évaluation des prototypes de recherche** a permis de mettre en évidence deux difficultés –et de proposer des solutions- :

- La difficulté d'associer des utilisateurs à une réflexion qui se situe encore dans le domaine de la recherche et est d'une part trop éloignée des outils auxquels ils sont familiers et d'autre part ne prend pas suffisamment en compte des critères ergonomiques. Notre proposition qui consiste à croiser les besoins imaginés des utilisateurs et les possibilités imaginées par les techniciens me semble très pertinente dans ce contexte.
- La difficulté et la nécessité d'évaluer les performances de systèmes de recherche d'information géographique pour pouvoir faire progresser les techniques de recherche. Notre proposition de construire des collections tests spécifiques me semble là encore la plus pertinente.

Il existe des domaines sur lesquels nous devons nous positionner comme la production de données localisées par d'autres acteurs que les acteurs traditionnels. A cet égard, il me semble que pour construire un contenu géographique utile –d'autant plus si la construction repose sur plusieurs contributeurs- il est important d'avoir des spécifications détaillées d'acquisition de ces données. Ce sont elles qui permettent de faire le lien entre l'information géographique et la réalité et qui permettent également d'assurer une relative homogénéité des données. Le principe d'explicitement toutes les informations rendant compte de la structure de données géographiques suivi dans la thèse de Sandrine Balley me semble intéressant également dans ce contexte. J'ai proposé un sujet de thèse sur la contribution à un contenu de façon distribuée en mettant l'accent sur l'importance de la prise en compte des spécifications et de la qualification du résultat. Il s'agit de la thèse de Carmen Brando qui commence au COGIT en octobre 2009.

De façon générale, **l'intégration dans les infrastructures des modèles de spécification d'acquisition et de contenu** pour faciliter des opérations automatiques ou pour faciliter la compréhension des utilisateurs –par le biais d'interfaces de visualisation telles celle proposée par Gyogyi Goder- me semble une piste de recherche essentielle.

2 La mise en ligne de processus adaptables à un contexte

Cette deuxième partie traite de la mise en ligne de processus adaptables aux contextes dans lesquels ils présentent une pertinence. C'est le domaine de la diffusion sur lequel nous avons conduit des travaux de recherche de façon constante et suivie.

La première section présente des travaux sur une proposition générique pour la mise en ligne de processus manipulant des données géographiques.

La section suivante présente des travaux focalisés sur un processus particulier. En 2007, j'ai choisi de dévouer la moitié du personnel de mon action de recherche à l'étude de la conception d'une carte sur mesure. C'est un domaine dans lequel il me semblait que l'IGN devait faire rapidement des propositions innovantes démontrant son expertise et dans lequel il était possible de finaliser rapidement des résultats exploitables par d'autres services de l'IGN. De plus, ce choix était cohérent avec les sujets des autres actions de recherches du laboratoire.

2.1 Le cas général

2.1.1 Généralités sur la description formelle de processus

Plusieurs domaines produisent des modèles formels dans lesquels les composantes de processus (objectif, entrées et sorties, opérateurs, méthode, etc.) sont explicitement représentées sous forme d'objets et non inclus dans des procédures.

L'Intelligence Artificielle symbolique propose des modèles des raisonnements humains pour les reproduire informatiquement ou les faciliter. (Clancey 1985) a montré que les connaissances de raisonnement sont plus efficacement représentées sous forme de tâches et de sous-tâches que sous la forme d'un ensemble non structuré de règles. L'efficacité renvoie ici à la facilité de maintenance du système. Depuis, de nombreux travaux ont porté sur la modélisation formelle de tâches. (Chandrasekaran et al. 1998) (Gomez et Benjamin 1999) préconisent de conserver une indépendance entre les connaissances de manipulation et les connaissances sur la nature des objets manipulés, c'est-à-dire les ontologies de domaines.

En ingénierie des projets et des organisations, de nombreux formalismes ont été proposés pour modéliser les processus internes importants d'une entreprise ou les activités commerciales entre plusieurs entreprises (BPML 2001) (Andrews et al 2003). Le composant est alors non pas un composant logiciel encapsulant un objet métier mais une vue sur une personne, un service ou une entreprise. La description de processus sert ici de deux façons :

- La contractualisation des activités permet aux parties prenantes de coopérer avec une certaine connaissance les unes des autres et également une confiance.
- La mise en place de nouveaux processus peut souvent s'appuyer sur des processus génériques déjà modélisés qui capitalisent une expertise sur la façon d'atteindre un objectif. Par exemple, dans le processus de retrait d'espèces dans un distributeur automatique de billets, il vaut mieux que la machine rende d'abord la carte à l'utilisateur avant ses billets. Ou encore, pour la conduite d'une réunion, il vaut mieux distribuer à l'avance un ordre du jour.

En ingénierie des projets logiciels, un point clé est d'identifier les composants les plus stables à travers l'ensemble des modèles produits (par exemple d'un projet à un autre). Actuellement, les ingénieurs logiciels s'accordent à penser que les composants les plus stables sont les objets manipulés (comme un compte client dans une banque) et non les méthodes de manipulation ou les objectifs de manipulation. Une démarche d'ingénierie d'un nouveau système à partir de composants existants s'attache à identifier d'abord les missions du système (usages) et les capacités du système sur lesquelles ces missions vont s'appuyer. En effet, ce sont les aspects susceptibles d'avoir évolué et de nécessiter un nouveau développement. Le reste de la conception logicielle vise ensuite à faire le lien entre les capacités spécifiées (qui sont nouvelles) et les objets métier existants (qui sont supposés être les éléments stables). Pour cela, le concepteur doit décrire comment les capacités sont réalisées par des séquences d'opérations portées par des composants logiques construits au-dessus des objets métiers. Le composant logique est donc d'abord spécifié par les opérations qu'il offre, soit encore sa spécification d'interface. Le processus décrit formellement est la réalisation d'une mission par des échanges de messages entre des composants, par exemple sous forme d'un diagramme de séquence.

Les plates-formes de composants distribués ont elles-mêmes évolué en reflétant la dualité données-méthode. Il y a quelques années, les implémentations proposées d'architectures de composants distribués étaient surtout des architectures d'objets distribués (comme CORBA). SOAP était initialement un langage pour accéder à l'interface d'un objet dans une telle architecture (Simple Object Access Protocol). Des implémentations de méthodes distribuées existaient aussi : XML-RPC et RMI. Actuellement, l'architecture reine est l'architecture orientée services ou SOA (Service Oriented Architecture). Des implémentations extrêmement en vogue d'une SOA sont celles des services Web, services grid et services peer-to-peer. Dans le service Web, le point de vue essentiel est le point de vue fonctionnel mais ces implémentations ont tout de même renforcé la place de l'objet par rapport au modèle de communication XML-RPC. En effet, dans les échanges de message entre services Web, les objets échangés sont davantage typés grâce à l'usage du protocole SOAP.

Finalement, il semble que **la réutilisation de données et la réutilisation de processus soient des problématiques dépendantes**. Un exemple de projet dans ce domaine est le projet PADOUE (2002-2006, <http://www-poleia.lip6.fr/padoue/>) qui vise le partage de données environnementales (Doucet 2006). Un aspect de ce projet était le partage d'indicateurs environnementaux, c'est-à-dire de méthodes. Le travail de Jean-Pierre Matsumoto a consisté à proposer un modèle de dataflow scientifiques permettant à des scientifiques de planifier la dérivation d'indicateurs à partir de données disponibles, partagées sur la plate-forme PADOUE. Un dataflow est un modèle de processus de manipulation de données dans lequel les données sont constamment maintenues en mémoire vive. Cela permet par exemple d'alimenter des entrepôts. Notre approche est

similaire à celle-ci, nous souhaitons coupler accès aux données et accs aux traitements sur les données. Pour l'utilisateur il n'est pas toujours important de savoir si l'information qu'il souhaite est disponible sous forme de données ou doit être calculée à partir de données existantes. Par contre, nous ne nous limiterons pas aux dataflow mais prendront en compte d'autres processus qui peuvent impliquer une interaction avec l'utilisateur. Nous nous rapprochons donc davantage de la proposition de (Lin et al 2009) qui vise à permettre la documentation et le partage de workflows scientifiques.

2.1.2 Modélisation de tâches

2.1.2.1 Un modèle dédié au développement collaboratif d'applications de traitement d'image

Dans cette section, je présente rapidement **une démarche de modélisation qui a inspiré ma proposition de thèse. Il s'agit de l'Atelier Logiciel** conduit par Marinette Revenu, professeur d'informatique à l'Université de Caen, Valérie Ficet et Régis Clouard. C'est une application d'aide à l'assemblage et au paramétrage de codes de traitement automatique d'images. J'ai voulu extrapoler cette proposition au domaine des données géographiques vecteur. La proposition de (Clouart et al. 02) s'adresse à des utilisateurs qui manipulent des images dans leur travail, par exemple dans un contexte médical, mais qui n'ont pas d'expertise en traitement automatique d'image. **Elle les aide à concevoir des programmes de traitement d'image qui automatisent les actions qu'ils souhaitent faire avec ces images.** Nous retenons de ces travaux l'importance de modéliser plusieurs niveaux de connaissances pour faire le lien entre l'expression du besoin faite par de tels utilisateurs et l'assemblage de codes existants pour y répondre. Ces niveaux sont décrits ci-après.

Les **tâches dites intentionnelles** se concentrent sur l'expression de l'objectif poursuivi par l'utilisateur. Elles sont décrites avec un vocabulaire générique du domaine des images comme "isoler les objets du fond" ou "former les objets par fusion des régions". Elles comportent les connaissances nécessaires à la spécification d'une méthode pour atteindre l'objectif de l'expert. Ces tâches intentionnelles portent donc des connaissances de type stratégique.

Les **tâches dites fonctionnelles** correspondent à des fonctionnalités bien connues de traitement d'image comme la "détection de contour" ou la "classification des pixels". Elles décrivent comment ces fonctionnalités sont atteintes en utilisant des outils de traitements d'image existant, les objets opérationnels. Elles portent donc des connaissances de type 'méthode'. La méthode d'une tâche intentionnelle est une décomposition en tâches fonctionnelles.

Les **objets opérationnels** correspondent à des algorithmes de traitement d'image comme "la classification de pixels basée sur une maximisation de la variance inter-classes".

A l'issue de premiers prototypes, (Clouart et al. 08) ont constaté que la spécification de l'objectif ne peut s'appuyer uniquement sur le domaine de l'image mais doit également prendre en compte des connaissances de contexte sur ce que représente l'image. Par exemple, pour savoir si un contour d'objet peut être modélisé comme une transition franche d'intensité il faut s'assurer que la scène ne comporte pas d'ombre, ou encore il est utile de savoir s'il s'agit d'une vue 3D, d'objets en mouvement ou non. Les auteurs ont mis en place une nouvelle solution dans laquelle un expert du domaine –qui connaît le contexte de l'image- et un ingénieur de la connaissance –qui est expert en traitement d'image- construisent ensemble les connaissances du contexte qui seront utiles à la conception du programme de traitement d'image. Ces connaissances portent sur le mode de production de l'image et également sur le domaine d'information sous-jacent (les objets présents sur l'image). La construction coopérative des connaissances de contexte se fait grâce à une interface dite de formulation de problème. Un problème en cours de formulation peut donner lieu à des générations de programmes de traitement d'image appliqués aux images de l'utilisateur. L'utilisateur s'appuie alors sur les résultats obtenus pour éventuellement reformuler le problème, par exemple il peut préciser qu'il préfère une surdétection d'objets ou encore il peut signaler que l'image a été altérée par une compression.

2.1.2.2 TAGE : un modèle de tâches géographiques pour la spécification coopérative d'utilisations de données géographiques

L'application proposée dans ma thèse s'inspire des travaux présentés dans la section précédente 2.1.2.1, en leur état à la fin des années 90. Le sujet de thèse était **l'organisation de métadonnées pour faciliter l'accès en ligne aux données géographiques**. Cette thèse a été dirigée par François Bouillé de l'Université Paris VI. Rappelons qu'une norme de métadonnées sur les données géographiques était en passe d'être adoptée, ISO19115, et que ce modèle était issu des producteurs de données et était destiné à des utilisateurs experts en données géographiques. J'ai donc élaboré une proposition complémentaire à destination d'utilisateurs moins experts en données géographiques dans laquelle l'utilisateur spécifie un besoin d'application et non un besoin de données. En quelque sorte, l'utilisateur spécifie *pourquoi* il a besoin de données et le système calcule *comment* utiliser *quelles* données pour atteindre son objectif. L'application utilise des plans génériques d'utilisation de données géographiques qui sont spécialisés de façon coopérative avec l'utilisateur. Le résultat est un plan d'utilisation qui comprend les requêtes de données à acquérir. Le modèle sous-

jaçant est un modèle à base de tâches et de rôles, appelé TAGE (pour tâches géographiques).

Cette proposition se fondait grandement sur l'exemple de l'Atelier Logiciel mais également sur le fait que les notions de tâches et de rôles géographiques étaient déjà présentes dans la littérature. La notion d'« affordance » a été introduite par (Gibson 1979) pour désigner des objets lors d'interactions économiquement intéressantes pour l'homme qui crée ces objets (la route qui me permet de me déplacer est une affordance de mon déplacement). (Jordan et al. 1998) reprennent les propositions de Gibson et insistent sur le fait qu'une localisation prend son sens dans le contexte d'une action de l'homme. La modélisation d'une localisation doit donc comprendre les activités que l'homme a dans cette localisation. Le domaine de la navigation routière a donné lieu à de nombreux travaux visant à expliciter des tâches et des rôles. (Timpf et al. 1992) analysent les tâches participant à l'activité de prévoir un voyage et de suivre un itinéraire : les tâches de planification, les tâches de guidage, les tâches de conduite. Les auteurs associent, à chaque groupe de tâches, un modèle d'objets dédiés à ces tâches. L'intérêt de ce modèle est de faire le lien entre une tâche à accomplir et les objets dont on doit avoir une représentation pour accomplir cette tâche. Il s'agit des rôles de ces tâches. (Claramunt et Mainguenaud 1996) distinguent également plusieurs catégories de connaissances de navigation : les connaissances procédurales, les connaissances topographiques, et l'espace perçu. Les connaissances procédurales renvoient aux tâches de guidage de (Timpf et al. 1992), les connaissances topographiques renvoient aux tâches de planification, et l'espace perçu aux connaissances de conduite.

Le modèle proposé dans ma thèse s'appuie sur les résultats de CommonKADS qui est une suite de techniques de modélisation utiles pour le cycle de vie d'un système à base de connaissances, en particulier en phase de conception (Schreiber et al. 99). Les aspects étudiés par CommonKADS sont par exemple l'analyse de la connaissance dont le système a besoin, analyse des experts impliqués et analyse des techniques de représentation de cette connaissance dans le système. Ce qui nous intéresse ici c'est la proposition des auteurs pour organiser des connaissances dans un système à base de connaissances. Cela nous donne des indications utiles sur la façon d'organiser des connaissances dans un système expert dédié à la manipulation à distance de données géographiques. CommonKADS propose de distinguer trois catégories de connaissances : les tâches, les inférences, le domaine.

Les **Tâches** sont des problèmes génériques pour lesquels on possède des méthodes de résolution. Un problème particulier se caractérise par un état final du monde à atteindre à partir d'un état initial. L'état final peut être décrit de façon plus ou moins précise. Une tâche se caractérise donc elle-même par un état générique du monde et un état générique final du monde (correspondant respectivement aux cas généraux des états du monde pour les problèmes particuliers). La résolution d'un problème s'appuie sur des opérations de changement d'état du monde et sur une méthode de résolution qui indique

comment enchaîner les opérations. Il est préconisé d'identifier les structures d'opérations récurrentes et d'en faire des sous-tâches pour rendre le modèle plus modulaire. Les tâches dont la décomposition ne comporte qu'une opération sont appelées tâches primitives (par opposition aux tâches complexes). Les opérations sont appelées inférences dans commonKADS. Enfin, les variables utilisées ou produites par les inférences et les tâches sont appelées des rôles. L'ensemble des rôles (les entrées, sorties ou paramètres) d'une tâche est souvent désigné comme son contexte. Les rôles prennent leur valeur dans le domaine (notion décrite plus loin). Lorsque plusieurs méthodes de résolution sont disponibles pour une même tâche, il est bon de modéliser également une stratégie pour cette tâche qui aide à sélectionner la méthode de résolution la mieux adaptée au contexte ; en pratique, la stratégie fait un lien entre des plages de valeur des rôles de la tâche et les différentes méthodes.

Les **Inférences** sont les opérateurs permettant de manipuler les objets du problème. Les entrées et sorties des inférences sont spécifiées fonctionnellement, sous la forme de rôles. Les rôles sont des variables qui prennent leurs valeurs dans le domaine.

Le **Domaine** décrit les objets manipulés : la structure et les instances. CommonKADS ne propose pas de structure pour le domaine. Comme je le rappelle dans (Bucher 2002) ceci est logique car le but d'une structuration est de supporter un type de raisonnement, voire d'optimiser des manipulations : les connaissances du domaine doivent pouvoir être structurées diversement selon le raisonnement que l'on souhaite effectuer dessus. Cette structuration se fait par l'intermédiaire des rôles. A une tâche correspond un ensemble de rôles qui forment le vocabulaire de base de cette tâche. Lorsque la tâche est mise en œuvre, le fait de valuer les divers rôles avec les éléments du domaine correspond à structurer dynamiquement le domaine en vue du raisonnement mené dans la tâche. Cela est illustré sur l'exemple extrait de (Bucher 2002) et repris sur la Figure 36.

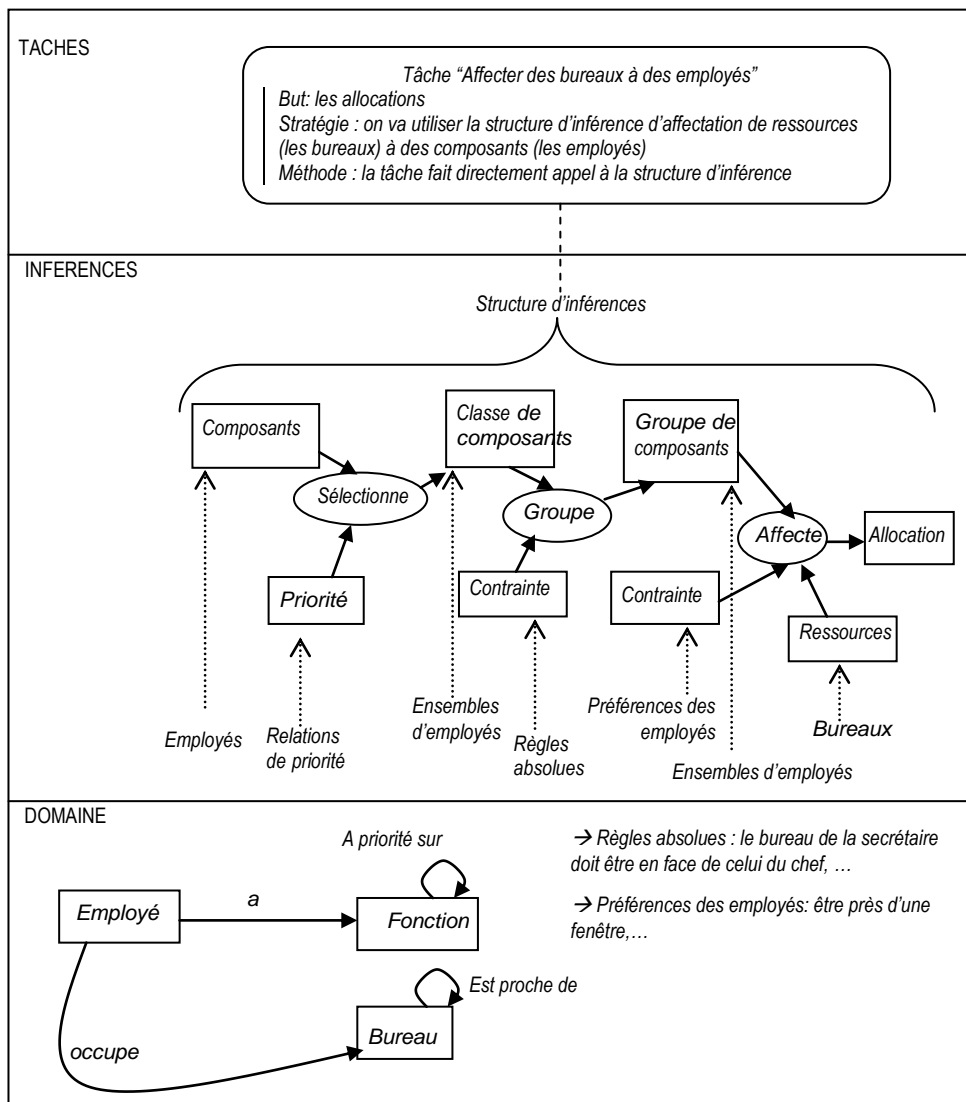


Figure 36 : Exemple d'utilisation du modèle KADS : la tâche d'affectation de bureaux.

Enfin, une notion importante est un peu négligée dans commonKADS : le but à atteindre dans la tâche. Il apparaît dans commonKADS sous forme d'un attribut textuel de la tâche. (Chandrasekaran et al. 98) qui ont proposé un modèle un peu similaire appelé 'Tâches génériques' expliquent qu'il n'y a pas de façon générique universelle de dire comment s'énonce un but. Certains buts s'énoncent uniquement à l'aide du vocabulaire proposé dans le domaine. D'autres buts intègrent dans leur énoncé des « termes d'attitude » comme 'souhaité' ou 'à éviter'. La proposition suivante apporte une solution sur la modélisation d'un but à atteindre : nous verrons que les auteurs proposent pour cela d'introduire un niveau supérieur de tâches : les tâches intentionnelles, dont la

description est dédiée à exprimer l'évolution du monde recherchée dans une manipulation (en amont de décrire comment s'effectue cette manipulation).

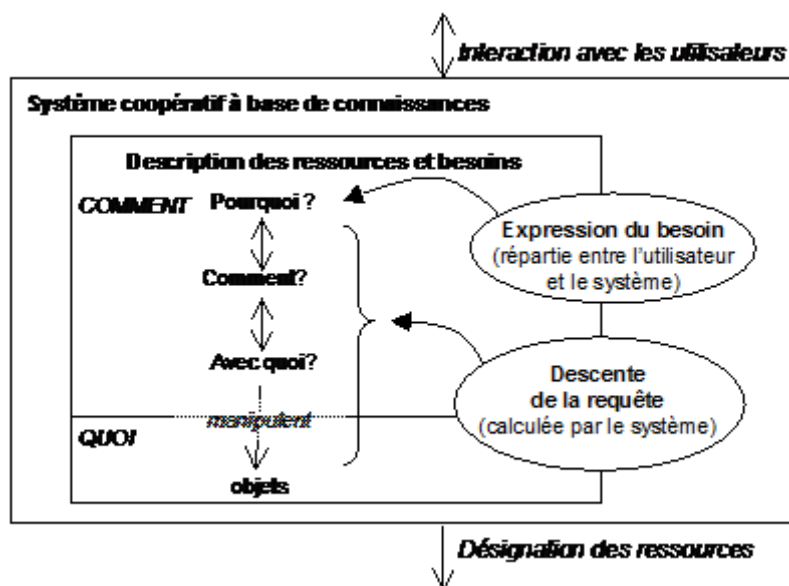


Figure 37. Exploitation d'un modèle intégrant des connaissances de type QUOI et COMMENT pour l'aide à l'accès. (d'après (Bucher 2003)).

Dans TAGE, une tâche représente une famille de problèmes selon deux aspects : son contexte générique et sa résolution générique. La représentation d'un contexte générique est composée de rôles entrées et d'un rôle sortie. La spécification du contexte revient à restreindre les valeurs possibles de ces rôles. Par exemple le contexte de la tâche Localiser_une_entité comporte un rôle localisation_recherchée qui peut prendre pour valeur une géométrie, une carte, des coordonnées GPS, des coordonnées linéaires, une entité administrative, ou des indications de navigation.

La représentation d'une résolution générique peut s'avérer difficile. Si une famille de problème peut se construire autour d'un objectif générique, il n'existe pas nécessairement de façon simple pour décrire génériquement les diverses méthodes résolutions correspondant à ces problèmes. Nous avons donc distingué d'une part les tâches primitives qui, à l'échelle du modèle, ont une résolution indépendante du contexte, comme "calculer une intersection" ou "superposer des couches de données". Cette résolution est appelée mécanisme. Les tâches primitives sont à rapprocher des tâches fonctionnelles du modèle de (Clourad et al. 2002). Et nous avons distingué d'autre part les tâches complexes dont la résolution varie en fonction du contexte, c'est-à-dire en fonction des valeurs prises par les rôles décrivant le contexte. Ces tâches sont à rapprocher des tâches intentionnelles de (Clouard et al 2002). La résolution générique des tâches complexes est constituée d'un plan générique et de connaissances stratégiques. Le plan générique se différencie d'un plan d'utilisation dans la mesure où il

peut comporter des tâches génériques et également dans la mesure où il peut comporter des structures de contrôle de type 'OU'. Les connaissances stratégiques permettant de résoudre les contrôle 'OU' du plan générique ou de restreindre les valeurs possibles des rôles (de la tâche principale ou des sous-tâches) en réaction à des spécifications des rôles de la tâche principale.

Comme illustré sur la Figure 38, le résultat d'une planification coopérative se présente comme un plan d'utilisation, c'est-à-dire une structure d'actions avec des conjonctions classiques "et", "si...alors..", "répète..jusqu'à..". Dans TAGE, une action est une tâche totalement spécifiée de sorte que sa résolution est déterminée ainsi que ses données d'entrée.

La difficulté sur laquelle je me suis concentrée était de permettre à l'utilisateur de spécifier de façon la plus flexible possible un plan générique. Le principe était de ne pas contraindre l'ordre dans lequel l'utilisateur spécifiait les différents rôles de sa tâche et d'obtenir un même plan quel que soit cet ordre. Pour ce faire j'ai introduit des rôles intermédiaires au niveau des plans génériques, appelés vocabulaire du plan.

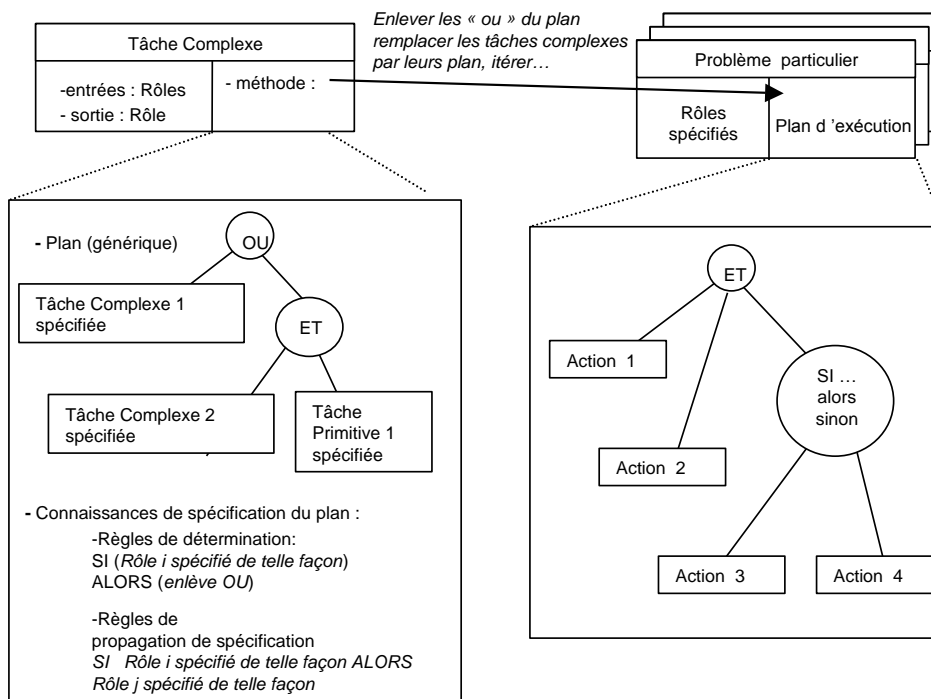


Figure 38. Représentation d'une méthode dans TAGE. La méthode est composée d'un plan générique et de connaissances de spécification de ce plan en fonction de la spécification du contexte de la tâche. Elle permet d'obtenir les plans d'utilisation correspondants. (d'après (Bucher 2003))

Parallèlement à la conception de cet environnement, la modélisation de plusieurs tâches géographiques a été conduite.

La tâche de localisation reprise sur la Figure 39 est très intensionnelle. Ce travail a consisté à analyser quelle était la forme de la réponse attendue à la question 'où est-ce' et comment la construire (les grands cas de figure). Une autre tâche intensionnelle a été modélisée : celle de la détection de proximité et de calcul d'itinéraire.

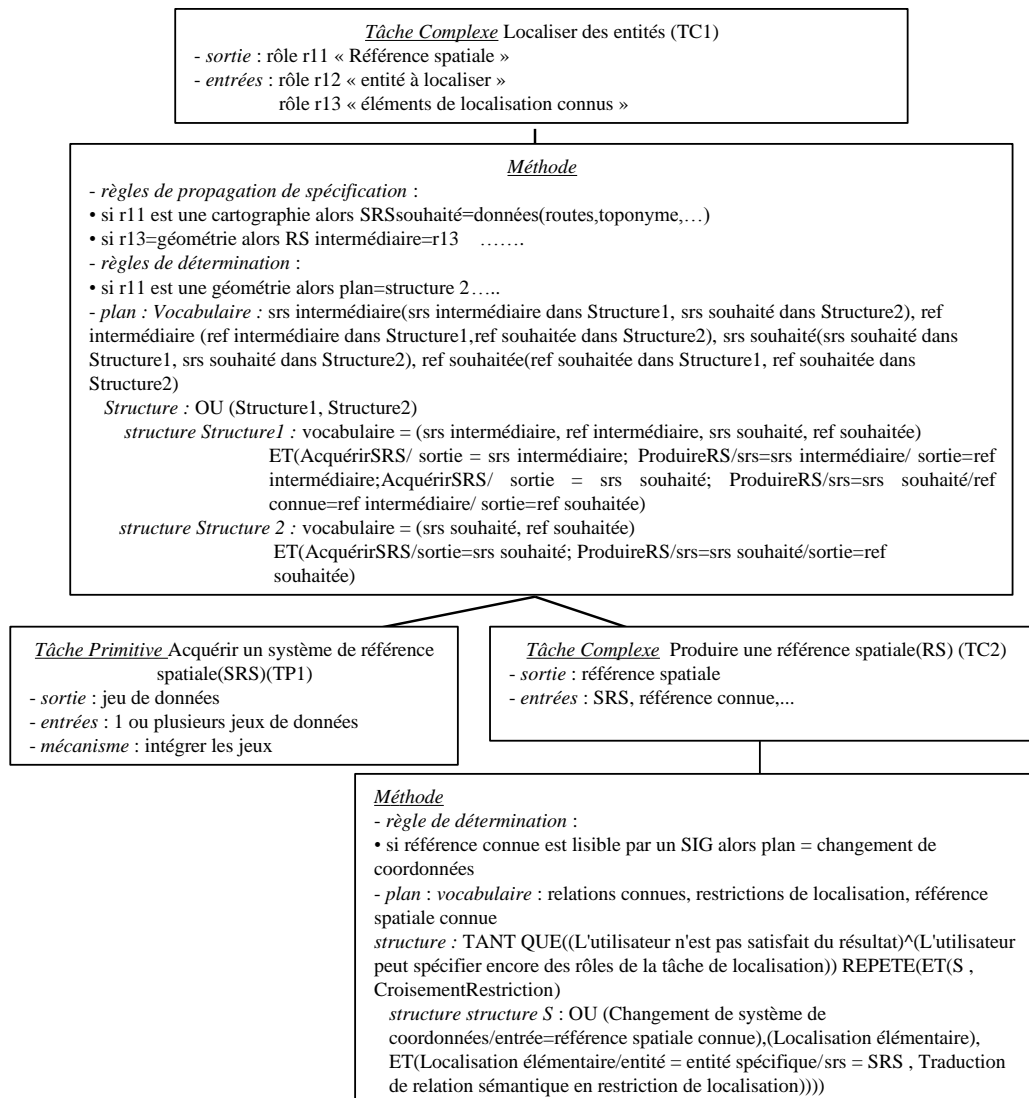


Figure 39. Modélisation de la tâche de localisation dans TAGE.

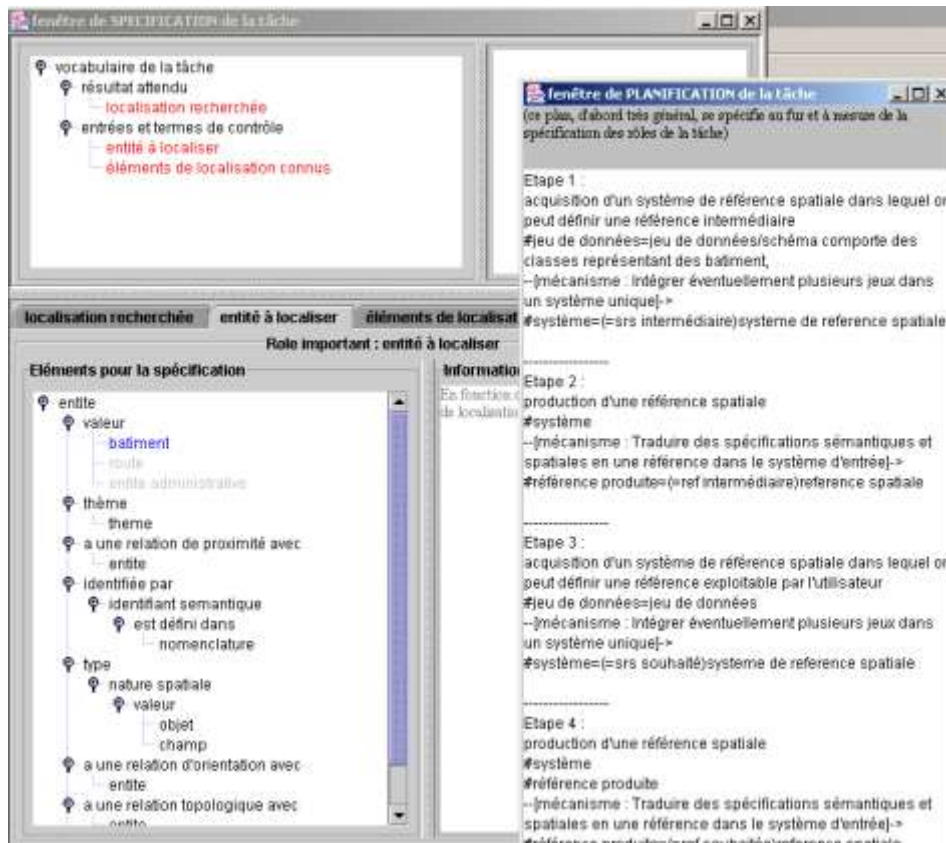


Figure 40. Prototypage d'aide à la planification d'applications de données géographiques.

Ce travail de thèse a conduit à plusieurs constats et enrichissements.

La modélisation à base de tâches et de rôles est efficace et prometteuse pour développer des processus spécialisés.

D'un point de vue modélisation, il semble a posteriori que viser d'emblée une représentation déclarative des connaissances était une faiblesse. Dans une optique de prototypage, il ne faut peut-être pas s'interdire le procédural qui rend le code plus lisible et plus facile à tester. TAGE a aussi montré la difficulté non seulement de décrire une variable et son domaine de valeur mais également de permettre automatiquement sa spécification et de décrire son état de spécification.

Toujours d'un point de vue modélisation, une difficulté de ce modèle est de déterminer la meilleure limite entre les tâches complexes et les tâches primitives. Selon la granularité du modèle de manipulation considéré –et selon la généricité recherchée– certaines tâches peuvent être vues comme des tâches complexes ou primitives. De même, dans les inférences il est parfois difficile de savoir si une notion doit être modélisée comme un rôle ou comme une structure du domaine.

Par ailleurs, j'ai expérimenté la complexité d'interagir avec des utilisateurs 'finaux' sur des prototypes de recherche extrêmement exploratoires. D'autres acteurs importants du système sont les experts qui connaissent bien des utilisations. J'ai essayé de construire des plans génériques avec certains et j'ai constaté que les concepts de tâches et de rôles ne sont pas toujours intuitifs.

Une difficulté intrinsèque du domaine de l'application est la complexité et l'hétérogénéité des modèles de données. Ceci concerne non seulement les données géographiques mais aussi les modèles de données des traitements considérés.

Une difficulté importante était la description des ressources logicielles et de leur intervention dans la tâche. Dans le cas de ressources logicielles qui sont des composants pour lesquels on dispose d'un contrat explicite d'interface, par exemple des services Web, cette difficulté est moindre. A la suite de ce travail, j'ai encadré un travail de stage sur la description de traitements de généralisation implémentés (stage de Pascal Michaux) puis le travail de thèse de Yann Abd-el-Kader sur ce même domaine. Avant de décrire ces travaux je présente des notions sur les services Web. Ceux-ci ont pris de l'importance à la fin de ma thèse et ont fortement influencé la proposition de Yann Abd-el-Kader.

2.1.3 Les services

Une difficulté dans TAGE est la description de briques logicielles réutilisables par les plans, c'est-à-dire les 'mécanismes' dans le modèle TAGE. Depuis ce travail, une évolution technologique considérable est le développement des architectures orientées services et des services Web. Ces services sont des ressources logicielles conçues comme des briques fonctionnelles réutilisables. Comme définition d'un service nous reprenons celle proposée par le consortium W3C « A service is an abstract resource that represents a capability of performing tasks that represent a coherent functionality. » Cette notion n'est pas spécifique aux services Web mais ces derniers en sont une implémentation phare. C'est pourquoi nous étudions souvent les services Web. Néanmoins de nombreuses propositions disent explicitement ne pas se limiter à ces services.

Les sections qui suivent présentent des travaux existants importants dans notre contexte. Notre proposition actuelle, qui vise à tirer partie de ces travaux, sera présentée elle dans la section suivante (2.1.4).

2.1.3.1 Généralités sur les services Web

Une architecture à base de services Web présente de nombreux avantages pour la mise en ligne de processus de manipulation de données géographiques. Un service Web, au sens large du terme, est un ensemble de fonctions qui peuvent être invoquées à distance par le biais d'échanges de messages transitant via un protocole Web comme le protocole HTTP¹ en utilisant le format XML pour l'échange d'information

¹ Le consortium W3C a restreint la définition d'un service Web en imposant le protocole SOAP pour les échanges, le modèle WSDL pour la description et le modèle UDDI pour se répertorier dans un catalogue. Ces préconisations vont dans le sens d'affaiblir le couplage entre les services:

- SOAP renforce l'interopérabilité en typant davantage les objets échangés, les services ne sont pas tenus de connaître leurs schémas de données à l'avance.
- SOAP permet de préciser la nature de certaines erreurs (par exemple liée aux structures des paramètres d'entrée). Ainsi l'origine d'une défaillance est mieux comprise et mieux gérée : faut-il refaire le message de requête ou s'adresser à un service miroir.
- WSDL est un contrat d'interface. L'utilisation d'un nouveau service peut reposer sur ce document, sans davantage de connaissance sur le service.
- UDDI est un modèle de catalogage : ainsi les services n'ont pas besoin d'être connus à l'avance, ils peuvent être 'découverts' au fur et à mesure des besoins grâce à des annuaires de services qui les indexent selon le modèle UDDI.

Mais ces restrictions ne sont pas toujours suivies à l'heure actuelle car la performance prime sur ces considérations et SOAP implique parfois un encodage lourd.

structurée. Cela permet de s'affranchir des problèmes d'hétérogénéités de plate-formes : un programme client peut interroger un service Web même s'ils tournent tout deux sur des plate-forme différentes. C'est un avantage important, par exemple dans le cas de programmes java qui utilisent des versions de bibliothèques java différentes (ce qui devient de plus en plus répandu).

Dans la plupart des cas, un service est mis en ligne pour que des programmes –non connus à l'avance du concepteur du service- puissent l'invoquer. Pour cela, un service fournit au préalable **son contrat d'interface**: les fonctions proposées et la communication à mettre en place avec le service pour effectuer ces fonctions. Le consortium W3C préconise l'utilisation du langage Web Service Description Language (WSDL) pour la description de services web. Grâce à l'utilisation d'un format standard pour décrire les contrats d'interface de services Web, il est possible de conduire facilement de nombreuses tâches comme découvrir des services pertinents dans un catalogue, explorer un service et le tester.

Une tâche importante liée aux services Web est leur **composition**. Différents modèles existent pour composer des services Web le plus automatiquement possible. Une approche très répandue consiste à écrire des scripts qui décrivent un enchaînement d'appels et de tests, et à exécuter ces scripts par un moteur dédié au format de script en question. C'est le cas de BEPL. Le moteur BEPL a été couplé avec une machine virtuelle java ce qui fait qu'il est possible d'ajouter des instructions java dans un script BEPL. Cela permet de décrire davantage de processus sans être soumis à la nécessité de faire réaliser tous les traitements, sauf les tests, par des services web. Ce mode de composition centralisé est appelé orchestration. Le moteur d'exécution du script se comporte en chef d'orchestre : tous les flux d'information passent par lui. Ce mode est très en vogue car il permet de décrire de nouveaux services sous la forme de composition de services existants et donc de gérer le processus complexe comme un service Web. Un autre mode de composition est appelé la chorégraphie. Cela consiste à décrire des bribes d'un processus en se concentrant sur les interactions possibles entre des services deux à deux. La réalisation du processus n'est donc pas contrôlée ni même conforme à un modèle comme dans l'orchestration. Cela revient à donner à chaque service qui doit être impliqué dans le processus des capacités d'agents pour qu'il remplisse correctement son rôle. Actuellement, les modèles de composition de services supportent généralement ces deux modes de composition.

Enfin, une dernière tâche liée aux services Web est leur **catalogage** de façon à permettre leur découverte sur le Web. Il existe plusieurs modèles de description de services Web dédiés au catalogage. Le modèle préconisé par le W3C est UDDI. Ce modèle référence un service Web par plusieurs descripteurs : catégorie, fournisseur, modèle.

Les modèles mentionnés précédemment se concentrent sur des **interopérabilités syntaxiques** (un service doit être alimenté avec des données au bon format) et sur des découvertes par mot-clé (les métadonnées de catalogage ne portent pas de sémantique formelle). **D'autres modèles sont proposés pour permettre la prise en compte de la sémantique dans les services Web**, c'est-à-dire automatiser la découverte, l'utilisation et la composition de services Web sur des critères plus avancés que la syntaxe des paramètres et les mots-clés des descriptions. Ces propositions s'appuient sur les modèles du Web sémantique présentés en section 0 de ce mémoire.

Une première proposition pour fonder des services web sémantiques a été **OWL-S** (Ontology Web Language for Services) (OWL 2004). OWL-S fonde la description d'un service sur trois facettes :

- Le *Model* : ce que fait le service de façon détaillée (exhaustive) et comment ces fonctions sont portées par des échanges entre le client et le service (et éventuellement d'autres services participants). Par exemple le *Model* d'un service fictif d'appariement pourrait signaler que ce service fournit des liens de correspondance entre un jeu de référence et un jeu comparé qui sont des liens $[n,m]$ entre des objets d'un jeu et des objets de l'autre jeu. Le *Model* préciserait que, pour obtenir ces liens, l'utilisateur doit fournir au service les jeux en question avec leurs schémas, que le service lui renvoie alors un appariement de schémas et que l'utilisateur doit valider cet appariement de schémas puis demander un appariement des données.
- Le *Profil* : ce que fait le service à première vue. Les informations du Profil sont destinées au catalogage, elles sont obtenues en sélectionnant des éléments de la facette Model (les fonctions que le fournisseur du service veut voir apparaître dans les catalogues, c'est-à-dire sur lesquelles il fonde sa publicité). Sur l'exemple précédent le Profil serait constitué du terme appariement et éventuellement des types d'objets pour lesquels l'appariement est proposé.
- Le *Grounding* : quels échanges programmatiques sont nécessaires pour mettre en œuvre les fonctions décrites dans le Model (conformément au processus décrit également dans le modèle). Il s'agit de l'url du service, du protocole à utiliser (ex : SOAP) et des schémas XML des messages à échanger.

Le *Model* est décrit sous la forme d'un processus, constitué de l'interaction du client avec le service. Ce processus peut générer une nouvelle information, cela est décrit dans les input et output du processus. Par exemple, un service d'appariement génèrera des liens de correspondance entre un jeu comparé et un jeu de référence. Ce processus peut aussi transformer le monde, cela est décrit dans les préconditions et effets. Par exemple un service d'équarissage de bâtiments aura un effet 'les bâtiments ont tous une géométrie rectangulaire' et comme précondition 'certains bâtiments n'ont pas une forme rectangulaire'. L'élément result permet de décrire sous quelle condition d'entrée le service produit une sortie donnée ou un effet donné. Par exemple, un service d'équarissage peut avoir un result 'les bâtiments sont inchangés' conditionné par un seuil de granularité des bâtiments. Les auteurs du modèle soulignent la difficulté de

proposer un modèle générique pour écrire des conditions (c'est-à-dire des formules logiques). Ils proposent de permettre divers formats d'écriture de formules logiques et d'associer à une expression (une formule logique écrite sous forme de texte) le format correspondant.

Une proposition plus récente du W3C consiste à **annoter les fichiers WSDL, les descriptions UDDI et les schémas XML dans des ontologies** (W3C Working Group 2007). L'annotation porte sur les opérations et sur les types d'objets manipulés. Par exemple, la définition du type d'un paramètre d'un service peut être suivie immédiatement d'une référence à un concept dans une ontologie auquel correspond le domaine de valeur de ce paramètre. Cette proposition est moins complète qu'OWL-S mais elle est également plus simple.

2.1.3.2 Les services géographiques

Dans le domaine de l'information géographique, un acteur majeur des services web est **le consortium OGC**. Celui-ci **a proposé des contrats types pour certaines fonctionnalités comme fournir des cartes, fournir des objets** ou fournir des champs. Il s'agit des spécifications appelées Web Map Service (WMS), Web Feature Service (WFS), Web Coverage Service (WCS). Ces contrats types correspondent un peu à des éléments qui pourraient figurer dans les taxonomies de services proposées dans le contexte des services web sémantiques. Ils facilitent des tâches élémentaires comme la découverte et l'utilisation d'un service web géographique. Ils facilitent aussi la mise en place des services. En effet, des programmes génériques sont développés pour remplir des fonctions de serveur (c'est-à-dire pour fournir les services) ou pour remplir des fonctions de client (c'est-à-dire pour interroger les services). Il s'agit par exemple des logiciels MapServer ou Geoserver (pour mettre en place les services WMS et WFS). Ces logiciels permettent à des personnes qui possèdent des données, la mise en place de serveurs de données ou de serveurs de cartes (objets, couches stylisées, champs, ...).

Les services OGC comportent tous une description qui peut être demandée en envoyant une requête '**GetCapabilities**' au service lui-même. Une traduction des GetCapabilities en format WSDL est en cours. Pour ce qui est du catalogage, le consortium ISO/OGC propose le modèle ISO19119 pour décrire des services pour le catalogage. Ce modèle est valable pour tous les services mais peut être enrichi avec des éléments du modèle ISO19115 dédié à l'information géographique (données et services). De plus, l'OGC a proposé une spécification d'interface pour le service de catalogage lui-même qui permet à un catalogue de communiquer avec un autre catalogue. Il s'agit du modèle CSW (Cataloguing Service on the Web). Il existe des logiciels pour installer des catalogues (de données ou de services) conformes aux spécifications CSW, comme MDWeb ou Geonetwork).

Une spécification de service OGC particulièrement intéressante est la spécification Web Processing Service (WPS) qui concerne la mise en ligne de traitements sur des données géographiques. Elle correspond à une fonctionnalité bien plus générique que les services OGC précédents : manipuler des données géographiques.

Cette spécification présente d'une part un intérêt similaire de celui des services précédents. Les spécifications WPS peuvent aider des fournisseurs de traitements, disponibles initialement sous forme de programmes desktop, à les mettre en ligne. Pour cela, il faut des logiciels comparables à MapServer et Geoserver qui puissent être installés et configurés pour donner accès non pas à des entrepôts de données mais à des traitements. Des logiciels existent déjà pour mettre en place des services WPS comme le logiciel de la communauté scientifique 52North (<http://52north.org/maven/project-sites/wps/52n-wps-webapp/>) ou le logiciel pyWPS très utilisé. Un client a aussi été proposé par 52North sous forme d'un plugin jump (Foerster and Schäffer 2007), d'un plugin udig, et d'une API client java.

Cette spécification présente d'autre part l'intérêt de répondre de façon anticipée à un problème plus complexe : **le problème de la découverte et de la composition de services de traitement de données géographiques**. Ces spécifications WPS proposent en effet un premier niveau de description générique qui sera utile pour la découverte et l'exploration de traitements WPS. Elles spécifient aussi une gestion possible des entrées et sorties sous forme de ressources externes (url) ou de ressources stockées localement sur le serveur, ce qui facilite la composition. Enfin, ces spécifications incitent les communautés scientifiques constituées autour de traitements spécifiques comme le traitement d'image à proposer des spécialisations de cette description générique. C'est le cas par exemple de la plate-forme de services de généralisation webgen initiée par l'Université de Zürich et transformée en WPS par l'Ordnance Survey (décrite en section 2.1.3.5). De telles spécifications plus avancées permettront la composition de traitements. A cet égard, le consortium OGC préconise de décrire des compositions de services WPS comme de nouveaux services WPS, c'est-à-dire en orchestration.

La spécification WPS comporte trois méthodes décrites ci-dessous.

1) La première est la **méthode GetCapabilities** commune à tous les services OGC et qui correspond à une description des méthodes du service (autres que GetCapabilities) donc ici les méthodes DescribeProcess et ExecuteProcess. La réponse GetCapabilities d'un WPS doit également comporter une section 'ProcessOfferings' qui décrit les traitements proposés sur des données géographiques.

Ci-dessous est reproduit l'exemple de la réponse à une requête GetCapabilities d'un WPS qui effectue un buffer.

```
<?xml version="1.0" encoding="UTF-8" ?>
```

```

<Capabilities [...]
  xsi:schemaLocation="http://www.opengeospatial.net/wps ..wpsGetCapabilities.xsd">
- <ows:ServiceIdentification>
  <ows:Title>AAFC GDAS-based WPS server</ows:Title>
  <ows:Abstract>AAFC GDAS-based WPS server developed for the OGC WPSie.</ows:Abstract>
- <ows:Keywords>
  <ows:Keyword>WPS</ows:Keyword>
  <ows:Keyword>AAFC</ows:Keyword>
  <ows:Keyword>geospatial</ows:Keyword>
  <ows:Keyword>geoprocessing</ows:Keyword>
  </ows:Keywords>
  <ows:ServiceType>WPS</ows:ServiceType>
  [...]
- <ows:OperationsMetadata>
- <ows:Operation name="GetCapabilities">
  [...]
  </ows:Operation>
- <ows:Operation name="DescribeProcess">
  [...]
  </ows:Operation>
- <ows:Operation name="Execute">
  [...]
  </ows:Operation>
  </ows:OperationsMetadata>
- <ProcessOfferings>
- <Process processVersion="1.0">
  <ows:Identifier>buffer</ows:Identifier>
  <ows:Title>Buffer a polygon feature</ows:Title>
  <ows:Abstract>Buffer the polygon coordinates found in one GML stream by a given buffer
  distance, and output the results in GML.</ows:Abstract>
  <ows:Metadata xlink:title="buffer" />
  <ows:Metadata xlink:title="polygon" />
  </Process>
  </ProcessOfferings>
</Capabilities>

```

2) La deuxième méthode est **DescribeProcess**. Celle-ci consiste à décrire des opérations fournies par le WPS. Les principaux paramètres d'entrée de cette méthode sont donc les identifiants de ces opérations.

Nous reprenons ici la réponse 'Describe Process' du service précédent appelée sur le processus de buffer.

```

<?xml version="1.0" encoding="UTF-8" ?>
- <!-- This example describes a buffer command that accepts polygon coordinates in GML, and
  used a buffer distance in meters to produce a buffered polygon feature, which is output in GML, in
  either UTF-8 or base64 encoding. The polygon can be returned directly as output, or stored by the
  service as a web-accessible resource. Ongoing processing status reports are not available.
  -->

```

```

-<ProcessDescriptions [...]>
- <ProcessDescription processVersion="2" storeSupported="true" statusSupported="false">
  <ows:Identifier>Buffer</ows:Identifier>
  <ows:Title>Create a buffer around a polygon.</ows:Title>
  <ows:Abstract>Create a buffer around a single polygon. Accepts the polygon as GML and
  provides GML output for the buffered feature.</ows:Abstract>
  <ows:Metadata xlink:title="spatial" />
  <ows:Metadata xlink:title="geometry" />
  <ows:Metadata xlink:title="buffer" />
  <ows:Metadata xlink:title="GML" />
- <DataInputs>
- <Input>
  <ows:Identifier>InputPolygon</ows:Identifier>
  <ows:Title>Polygon to be buffered</ows:Title>
  <ows:Abstract>URI to a set of GML that describes the polygon.</ows:Abstract>
  - <ComplexData defaultFormat="text/XML" defaultEncoding="base64"
  defaultSchema="http://foo.bar/gml/3.1.0/polygon.xsd">
- <SupportedComplexData>
  <Format>text/XML</Format>
  <Encoding>UTF-8</Encoding>
  <Schema>http://foo.bar/gml/3.1.0/polygon.xsd</Schema>
</SupportedComplexData>
</ComplexData>
  <MinimumOccurs>1</MinimumOccurs>
</Input>
- <Input>
  <ows:Identifier>BufferDistance</ows:Identifier>
  <ows:Title>Buffer Distance</ows:Title>
  <ows:Abstract>URI to a GML resource file</ows:Abstract>
- <LiteralData>
  <SupportedUOMs defaultUOM="meters" />
  <ows:AnyValue />
</LiteralData>
  <MinimumOccurs>1</MinimumOccurs>
</Input>
</DataInputs>
- <ProcessOutputs>
- <Output>
  <ows:Identifier>BufferedPolygon</ows:Identifier>
  <ows:Title>Buffered Polygon</ows:Title>
  <ows:Abstract>GML stream describing the buffered polygon feature.</ows:Abstract>
  - <ComplexOutput defaultFormat="text/XML" defaultEncoding="base64"
  defaultSchema="http://foo.bar/gml/3.1.0/polygon.xsd">
- <SupportedComplexData>
  <Format>text/XML</Format>
  <Encoding>UTF-8</Encoding>
  <Schema>http://foo.bar/gml/3.1.0/polygon.xsd</Schema>
</SupportedComplexData>

```

```
</ComplexOutput>
</Output>
</ProcessOutputs>
</ProcessDescription>
</ProcessDescriptions>
```

3) La troisième méthode est **ExecuteProcess** qui vise à invoquer une opération particulière. Ses arguments sont le nom de l'opération, les valeurs de ses paramètres ainsi que des instructions spécifiques comme le fait de stocker le résultat sur le serveur –dans ce cas il pourra être utilisé en argument d'une autre exécution–.

Nous reproduisons ci-dessous la requête d'exécution d'un buffer.

```
http://foo.bar/foo? request=Execute& service=WPS& version=1.0.0& language=en-CA&
Identifiant=Buffer&
DataInputs=Object=@xlink:href=http%3A%2F%2Ffoo.bar%2Ffoo;BufferDistance=10&
ResponseDocument=BufferedPolygon& StoreExecuteResponse=true
```

2.1.3.3 OPERA : une plate-forme sémantique pour des services géographiques

OWL-S a été utilisé par (Lemmens 2006) pour l'enchaînement de services Web géographiques. Dans sa thèse, il introduit trois types d'ontologies sur lesquelles s'appuieront les descriptions de services et l'utilisation de ces descriptions :

- les ontologies des concepts du monde réel (feature concept ontology),
- les ontologies des concepts des bases de données (feature symbol ontology)
- les ontologies des opérations géographiques (geo-operation ontology).

Il s'agit de types d'ontologies pour lesquels il propose à chaque fois des exemples. Ces types d'ontologies doivent comporter des liens entre elles. Un concept d'une base de données (par exemple un FeatureType particulier) doit être rattaché à un concept du monde réel (par exemple la catégorie Bâtiment dans une ontologie de ce domaine). Les données géographiques impliquées dans une opération doivent être rattachées à des concepts de bases de données. L'auteur propose une ontologie d'opérations appelée OPERA qui comporte deux parties : OPERA-R dédiée aux opérations atomiques (le R signifie référence) et OPERA-D dédiée aux opérations complexes construites sur les opérations atomiques (le D signifie dérivé). OPERA décrit des types d'opérations, un peu comme les spécifications OGC décrivent des types de services (une instance est un service implémenté conforme à ces spécifications d'interface). Il s'attache à décrire plus spécifiquement des opérations qui manipulent des données vecteur, 'feature processing opérations'. La classification de ces opérations dans OPERA s'appuie sur les classes des objets manipulés.

Ces ontologies permettent d'ajouter de la sémantique dans la description d'un service en faisant le lien les opérations qui composent le service et les types d'opérations décrits dans une ontologie, ce qui implique un lien entre les paramètres (et leurs types) des opérations composant le service et les paramètres (et leurs types) des types

d'opérations correspondantes dans OPERA. De plus, la description d'une composition de services s'appuie également sur OPERA qui comporte des éléments dédiés à l'écriture de workflow.

La Figure 41 illustre l'utilisation d'OPERA pour décrire une opération de géocodage du service gazettier de l'Alexandria Digital Library. L'ontologie de service permet d'annoter les services et les opérations alors que l'ontologie des symboles permet d'annoter les paramètres. Le gazettier de l'exemple est décrit comme une instance du type de service de haut niveau *serv:GeoService*. L'opération en question est également décrite comme une instance de *opera:LocSpat*. Un paramètre d'entrée de l'opération est décrit comme étant une instance du type de symbole *symbol:GF_LocationAttributeType*.

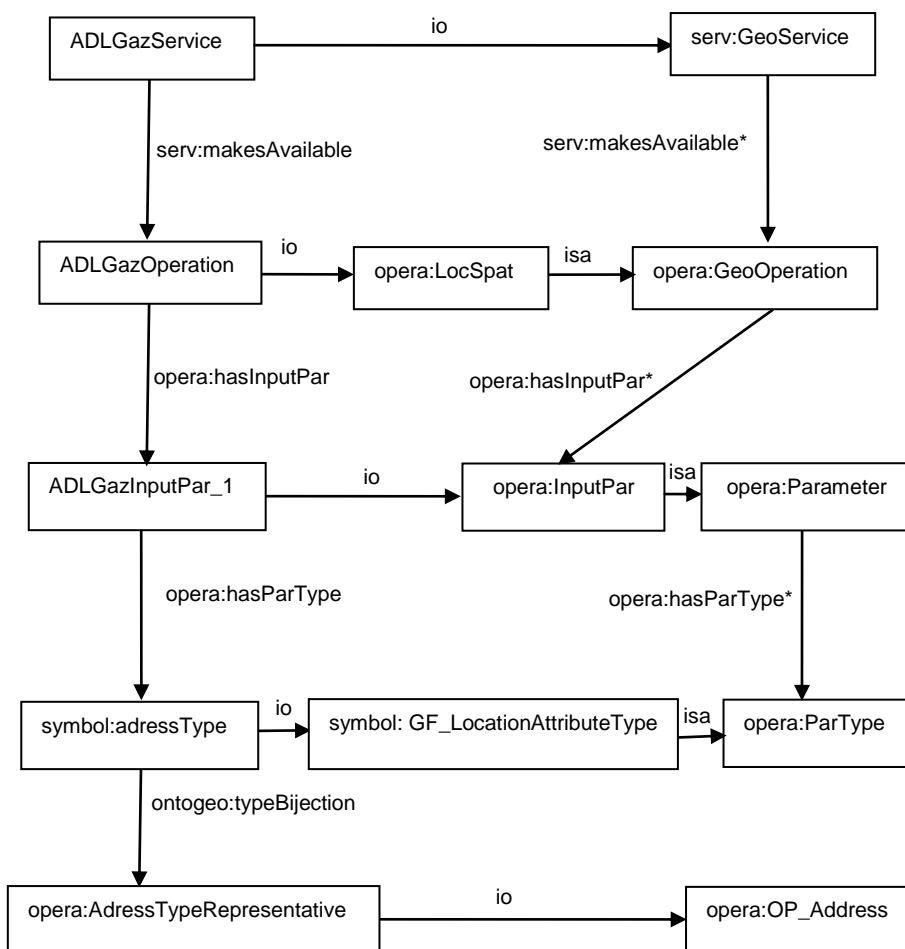


Figure 41. Description partielle du service gazettier de l'Alexandria Digital Library à l'aide du modèle de (Lemmens 2006).

Ainsi, la découverte de services peut s'appuyer sur des capacités de raisonnement automatique du formalisme OWL. Cela est très utile lors de la composition automatique :

le moteur vérifie que le résultat d'un service peut être envoyé à un autre service, c'est-à-dire qu'il aligne en quelque sorte les modèles d'entrée et de sortie des divers services. Dans le cas où les modèles ne correspondent pas, il recherche un service de traduction du concept de sortie du service courant de la chaîne dans le concept d'entrée requis par le service suivant.

2.1.3.4 MOBIDIC: une plate-forme pour le développement de services adaptables à des contextes mobiles

Une évolution sensible dans la problématique de la diffusion d'information géographique est l'**adaptativité croissante des interfaces de diffusion aux contextes d'utilisation**. Dans certaines situations, le destinataire de l'information n'est pas en mesure de mobiliser de nombreuses capacités pour utiliser cette information. C'est au diffuseur de 'faire le chemin' vers l'usage et de lui fournir une information la plus facile à assimiler, voire de 'pousser'¹ l'information vers lui à son insu. Cela peut se produire parce que l'utilisateur doit mobiliser, dans des délais éventuellement réduits, de nombreuses capacités de raisonnement (ou de conduite) et qu'il est souhaitable de ne pas le solliciter davantage. C'est par exemple le cas de la conduite automobile mais aussi de la gestion de crise (Friedmannova et al. 2006) ou encore d'une prise de décision d'un décideur politique (Ormeling and Kraak 2008). L'information géographique dont il a besoin doit alors lui être communiquée de la façon la plus immédiatement intégrable dans sa prise de décision. Le cas le plus avancé d'adaptation au contexte est ce qu'on appelle la cartographie ubiquitaire (Morita 2005)(Gartner et al. 2007). Le but des développeurs de logiciels et de matériels en informatique ubiquitaire est de rendre la plus permanente et la plus transparente possible la relation entre l'homme et des programmes qui vont l'aider dans ses activités. L'adaptation porte alors sur tout ce qui intervient dans la communication entre l'utilisateur et le programme, c'est-à-dire l'utilisateur lui-même, son activité, l'environnement physique, les réseaux de positionnement et de communication (couverture réseau et débits) ou encore les autres utilisateurs. Par exemple, (Laucius et al. 2006) analysent la couverture et le débit du réseau de communication sur un paysage parcouru par un utilisateur mobile pour déterminer un envoi d'objets géographiques sur un terminal nomade. Si l'utilisateur est susceptible d'emprunter un chemin peu couvert par le réseau ou encore un chemin où l'information géographique est trop dense pour le débit disponible, le serveur lui envoie par anticipation les informations importantes relatives au trajet. Ces informations sont conservées dans un 'sac à dos' et sont détruites lorsqu'elles deviennent inutiles, soit après que l'utilisateur a parcouru le

¹ Le terme 'pousser' fait référence ici à l'alternative 'push/pull' pour diffuser une information : en mode pull l'utilisateur émet une requête alors qu'en mode push l'utilisateur n'émet pas de requête et c'est le diffuseur qui prend l'initiative de lui adresser une information.

chemin, soit parce qu'il a choisi un autre chemin. Notons, qu'avec l'informatique ubiquitaire, de nombreux travaux portent également sur la notion d'interface ambiante, c'est-à-dire d'interface dans laquelle l'utilisateur n'exprime pas son besoin car des capteurs le font à sa place. La définition d'un contexte, l'acquisition du contexte courant et l'adaptation d'une réponse à un contexte sont donc des enjeux importants pour la diffusion d'information géographique. Pourtant, il n'existe pas encore une définition consensuelle et détaillée des éléments qui composent un contexte, des éléments pertinents dans une application donnée et de la façon de les prendre en compte. Nous reviendrons dans la section 2.2.1 de ce mémoire sur les éléments de contexte pris en compte dans la définition des styles d'une carte comme l'heure du jour ou l'âge du lecteur.

(Lopez-Velasco et al. 2009) proposent **une plate-forme pour permettre d'une part aux développeurs de services web localisés ou d'autres composants logiciels d'intégrer dans leurs services des éléments de contexte** (comme la taille de l'écran du dispositif nomade ou de sa mémoire) et **d'autres part à des concepteurs d'applications à base de tels services de découvrir et assembler ces services**. Il s'agit de la plate-forme MOBIDIC.

Le premier aspect de cette proposition repose sur un middleware installé sur les clients qui détecte le contexte d'utilisation et l'inclut dans les requêtes vers les services, en reprenant une proposition initiale de (Viana 2007). Le contexte est organisé en quatre sections : l'utilisateur, la localisation et la météorologie, la date et les entités informatiques. Ce middleware est aussi en charge d'afficher les réponses des services. Il est comparable aux environnements de développements comme celui de l'iphone ou android, à la différence que le middleware MOBIDIC a des missions plus ciblées. Ce modèle permet à des développeurs de méthode d'une part d'être conscients de la pertinence de ces facteurs dans un contexte mobile et d'autre part de les prendre en compte dans leur service si nécessaire. Par exemple un service qui fournit une carte pourra considérer qu'un paramètre est la taille de l'écran et pourra adapter sa réponse à ce paramètre.

Le deuxième aspect de cette proposition repose sur :

- Un modèle dédié de description de services (services web ou composants logiciels) : le modèle WSRM pour Web Service Representation Model
- Des bases de connaissances et des moteurs spécifiques qui fournissent des fonctions de découverte et de composition des services décrits dans ce modèle dédié.

Le modèle WSRM reprend les éléments des standards WSDL (méthodes proposées, etc) et UDDI (fournisseur du service) et y ajoute les éléments suivants :

- Le domaine d'application du service, comme le e-learning. Cet élément est documenté en référence à une base de connaissance des domaines.

- Le contexte d'utilisation du service, c'est-à-dire les prérequis côté client et utilisateur pour que le service soit utilisé (positionnement de l'utilisateur, droits d'accès, caractéristiques matérielles du client, etc). Cet élément est documenté en référence à une base de connaissance des contextes.

Les descriptions en WSRM des services sont stockées dans la base de connaissance des services et les orchestrations possibles sont stockées dans une dernière base de connaissances : la base de connaissances des applications. De plus, le langage utilisé pour implanter WSRM est le langage AROM (Associer Relations et Objets pour Modéliser) qui est un langage de représentation de connaissances par objet (Page et al. 2000). AROM présente des capacités d'inférences des logiques de description comme OWL (voir section 1.2.2.1).

La composition de services pour réaliser une application donnée s'appuie ensuite sur ces bases de connaissances et sur les capacités du langage AROM. L'originalité de cette proposition réside dans la prise en compte du contexte pour construire une composition adaptée. L'adaptation peut consister à spécifier les paramètres des services ou à ajouter des 'procédures' dans la composition globale.

2.1.3.5 WebGen : des services pour faire avancer la recherche au sein d'une communauté

Une communauté très active dans le domaine des services web géographiques est celle des chercheurs en généralisation. Ceux-ci se sont penchés relativement tôt sur la technique des **services Web vue comme une solution pour faire progresser la recherche en généralisation en permettant aux chercheurs de comparer leurs propositions** (Edwardes et al. 2005). La plupart de ces travaux ont pris place dans le cadre du projet WebGen initié par l'Université de Zürich. Le prototype WebGen a été ensuite repris pour être mis en conformité avec les spécifications WPS. Les résultats actuels sont des services Web fournissant des opérations de généralisation et surtout **des modèles communs pour que ces services soient le plus interopérables possibles** et soient décrits de façon précise et non ambiguë dans cette communauté. Ces modèles communs sont décrits dans la suite de cette section.

Les premiers modèles portent sur **les structures manipulées** (pour éviter les traducteurs lorsque cela est possible, et pour construire des descriptions partageables). Le standard ISO/OGC GML est utilisé pour les structures primitives comme un point ou une ligne. Les services de généralisation travaillent parfois sur des structures plus complexes comme les graphes. Pour ces services, des standards ad hoc sont proposés pour décrire les structures de données en question, comme le standard proposé par (Regnauld 2005) pour décrire les graphes. (Foerster et al. 2008) proposent une liste des modèles de données standard utilisés pour les WPS de généralisation.

D'autres modèles portent **sur les services et sur les opérations**. Pour décrire les opérations, une taxonomie d'opérations est proposée (Foerster et al. 2007). Une autre proposition apparaît dans (Burghard et al. 2005). Les auteurs proposent plusieurs classifications des services de généralisation dont une qui consiste à préciser la phase d'un processus global de généralisation dans laquelle se situe le service (prétraitement, transformation ou séquençement). Des services particulièrement étudiés pour la généralisation sont ceux qui permettent d'enrichir un jeu de données existant avec des objets de plus haut niveau comme les villes, les ronds points ou les îlots. En effet, lors de la généralisation, il faut généralement conserver les informations qui correspondent à un niveau de détail plus grand que celui de la représentation courante et il s'agit généralement d'objets complexes qui ne sont pas explicitement représentés dans la base mais apparaissent seulement comme leurs parties.

Enfin, concernant **la spécification des buts** d'un service de généralisation, des travaux portent sur la formalisation standard de **contraintes en cartographie** (Burghardt et al. 2007).

2.1.4 Le catalogage et le partage de ressources logicielles

Le début de la deuxième partie de ce mémoire a présenté une proposition globale élaborée dans ma thèse qui vise à faciliter l'accès d'utilisateurs à des processus de manipulation de données géographiques grâce à des modèles formels de ces processus. Une lacune importante était l'absence de brique logicielle. L'expérimentation portait donc sur la partie intentionnelle –aider l'utilisateur à préciser son besoin- et non sur la partie opérationnelle –générer l'application-. Depuis, les services sont une technologie prometteuse pour dépasser cette lacune. La partie précédente a présenté la notion de service et des démarches qui me semblent importantes dans le domaine des services géographiques. Cette partie-ci présente enfin comment ma proposition initiale a évolué pour prendre en compte les services.

2.1.4.1 MDT : un modèle de métadonnées pour le catalogage de ressources logicielles

Concernant **la description de traitements fournis par des ressources logicielles**, un premier travail de stage (Michaux 2003) a étudié les besoins de telles métadonnées et proposé un premier modèle dédié aux traitements de généralisation. Ce stage était co-encadré par moi-même, qui apportais l'expertise sur la problématique générale de décrire des traitements, et par Sébastien Mustière qui apportait son expertise du domaine de la généralisation. Le stagiaire a repris des fiches de description de traitements de généralisation élaborée dans le cadre du projet européen AGENT (Barrault et al. 2001) pour synthétiser les outils disponibles parmi les participants au projet et a animé des réunions avec des experts en généralisation pour affiner avec eux un modèle de description. Ces fiches AGENT, et avant elles des descriptions produites dans le contexte de l'OEEPE, représentaient un travail important de normalisation du vocabulaire en généralisation. Nous avons constaté **l'importance de pouvoir spécialiser les descriptions** dans certains cas –les mesures demandent des descripteurs spécifiques par exemple- et également **l'importance des illustrations pour communiquer sans ambiguïté la fonction d'un traitement cartographique**. Suite à ces réunions, le stagiaire a établi un glossaire de termes et a proposé un modèle de description en deux niveaux : le premier niveau est générique et doit s'appliquer à tout type de traitement, un niveau plus spécifique correspond aux traitements de généralisation. Ces deux niveaux sont illustrés sur la Figure 42 où les éléments entourés sont spécifiques au domaine de la généralisation.

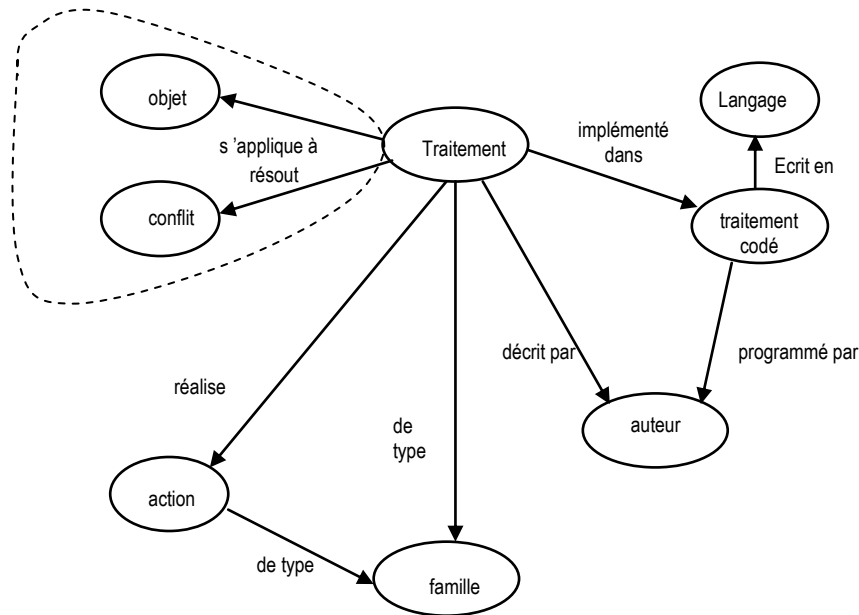


Figure 42. Modèle de description de traitements proposé par (Michaux 2003). Les descripteurs entourés sont dédiés aux traitements de généralisation.

Nous reprenons sur la Figure 43 le glossaire proposé par (Michaux 2003).

Traitement : Un traitement est une manipulation de l'information qui a une entrée et une sortie (disponible ou non, complexe ou non,...).

Traitement codé : désigne l'implémentation d'un traitement

Objet : Objet géographique représentant une entité et sur lequel on utilise des traitements de généralisation.

Conflit : Problème posé par le changement d'échelle auquel le traitement de généralisation doit répondre.

Famille : Grande classe d'opérations de généralisation correspondant à un but général. (simplifier, ...)

Action : Opération élémentaire réalisée par un traitement. Une action peut correspondre à un traitement mais un traitement peut réaliser plusieurs actions.

Auteur : Individu étant à l'origine d'un traitement ou ayant implémenté un traitement.

Figure 43. Glossaire des éléments du modèle de métadonnées proposé par (Michaux 2003).

La Figure 44 reprend une copie d'écran du prototype de catalogue développé par Pascal Michaux.

Navigation: famille | auteur | conflit | langage | action | traitement codé | objet | traitement | Sommaire

traitement : accordéon

Diagram: objet, famille, action, conflit, Traitement, auteur, traitement codé, langage

Informations sur la ressource	
conflit résolu	empatement
est implémenté dans	code de accordéon
objets traités	série de virages
famille du traitement	caricature

principe
Etale un virage ou une série en décalant chaque point intermédiaire, dans la direction principale

avantage
Algorithme particulièrement adapté aux séries de virages routiers en épingles à cheveux

inconvenient
Si l'axe de virage n'est pas reciligne l'écartement peu être abhèrent ou inadapté pour résoudre le problème. Si l'empatement du virage se situe plutot au sommet du virage, l'application de cet algorithme pour résoudre ce problème risque de fortement décaler les virages. Il vaut mieux utiliser un autre algorithme comme celui de la baudruche.

article
<http://quasimodo:8080/essai/doc/generalisation/accordion.doc>

Image

Figure 44. Copie d'écran du catalogue de traitements à base de TopicMaps développé par (Michaux 2003).

A la suite de ce travail de stage, une thèse a été conduite sur ce même sujet, la **thèse de Yann Abd-el-Kader dirigée par Marinette Revenu** professeur de l'Université de Caen et que j'ai encadrée au COGIT. L'objectif de cette thèse était de proposer un **modèle pour le catalogage de traitements fournis par des ressources logicielles**, ce qui correspondait aux mécanismes de TAGE. Dans le cadre de sa thèse, Yann Abd-el-Kader a communiqué un questionnaire à des personnels de la direction technique de l'IGN leur demandant quelles questions ils aimeraient poser à un moteur de recherche de traitements. L'objectif était d'acquérir des besoins et également d'identifier des termes de vocabulaire récurrents pour les reprendre dans le modèle de métadonnées.

Le résultat de ce questionnaire a été d'une part qu'il existait des besoins variés liés à la mise en place d'une base de métadonnées sur les traitements logiciels et d'autre part que de nombreux termes étaient ambigus ou flous, comme les termes *processus*, *traitement*, *opération*, *mesure*, *algorithme* et *méthode*.

Outre l'absence de vocabulaire précis pour parler des traitements, une autre difficulté à laquelle nous avons été confrontés pendant ce travail de thèse a été **l'absence de modèle de ressources pour le domaine des ressources logicielles en général**. Aussi, Yann Abd-el-Kader a proposé de distinguer trois catégories de ressources d'intérêt :

- les algorithmes vus comme des modèles abstraits de traitement,
- les traitements codés désignant tout ensemble d'instructions d'un langage informatique, interprétable par une machine'
- les ensembles de traitements codés qui sont effectivement échangeable, par exemple une classe java ou un jar.

Un choix fort de ce travail de thèse a été de reprendre dans la mesure du possible la terminologie des services web pour décrire les traitements proposés par des ressources logicielles, qui ne sont pas nécessairement des services Web. Ce choix est motivé par le succès des architectures orienté-services : la notion de service semble essentielle pour accéder à des ressources logicielles et pour construire des applications à partir de ressources existantes. Ce choix est également motivé par la perspective de développer ultérieurement des services web en s'aidant des métadonnées construites.

Ainsi, le modèle **de description MDT s'inspire du modèle OWL-S** décrit précédemment. Il associe à une RessourceTraitement les facettes suivantes (voir Figure 45):

- la Fonction est la description de ce que fait le traitement, elle reprend par exemple les éléments du modèle de (Michaux 2003) pour les traitements de généralisation
- le Fonctionnement est la description du fonctionnement interne
- le ModeEmploi est la façon d'accéder à la ressource
- la facette Identification regroupe des éléments qui servent ensemble à identifier sans ambiguïté la ressource décrite
- la facette Evaluation décrit des utilisations de la Ressource et des performances constatées.

Par rapport à la description d'un service Web, les éléments réutilisés sont ceux de *OWLS-Model*. Par définition, *MDT-Fonction* correspond à *OWLS-Model*. Le fonctionnement interne d'un service atomique n'est pas décrit dans OWL-S mais le fonctionnement d'un service complexe est décrit dans le *Model*. Le mode d'emploi d'un service Web est décrit dans le *Model* et le *Grounding* de OWL-S. Le *Model* comporte la description à un niveau abstrait et le *Grounding* décrit comment les échanges décrits dans le *Model* s'implantent dans un protocole Web. Ainsi dans le modèle MDT les deux éléments nouveaux sont essentiellement les suivants :

- **l'implantation du mode d'emploi d'un traitement** qui n'est pas nécessairement un Service Web,
- **le fonctionnement interne d'un traitement atomique.**
- **la facette Evaluation correspondant à des propriétés de performance** (propriétés qu'il est possible d'ajouter dans le *Profil* d'un service en OWL-S).

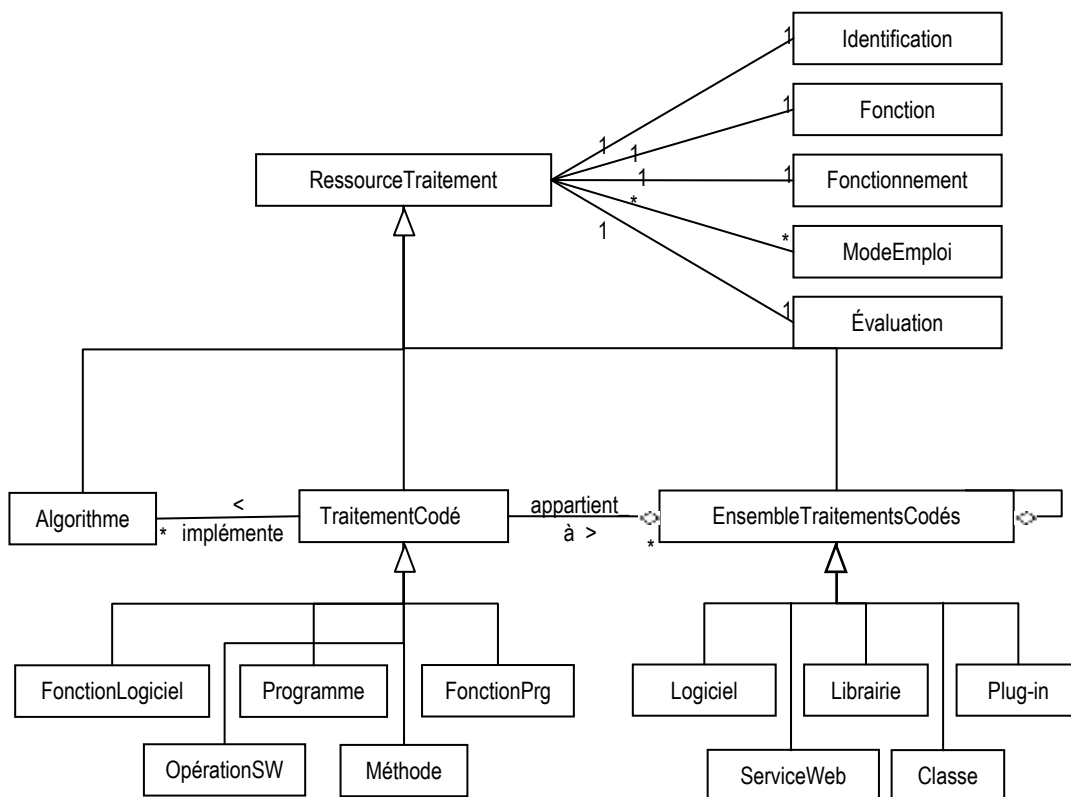


Figure 45. Modèle de métadonnées MDT pour décrire des traitements fournis par des ressources logicielles.

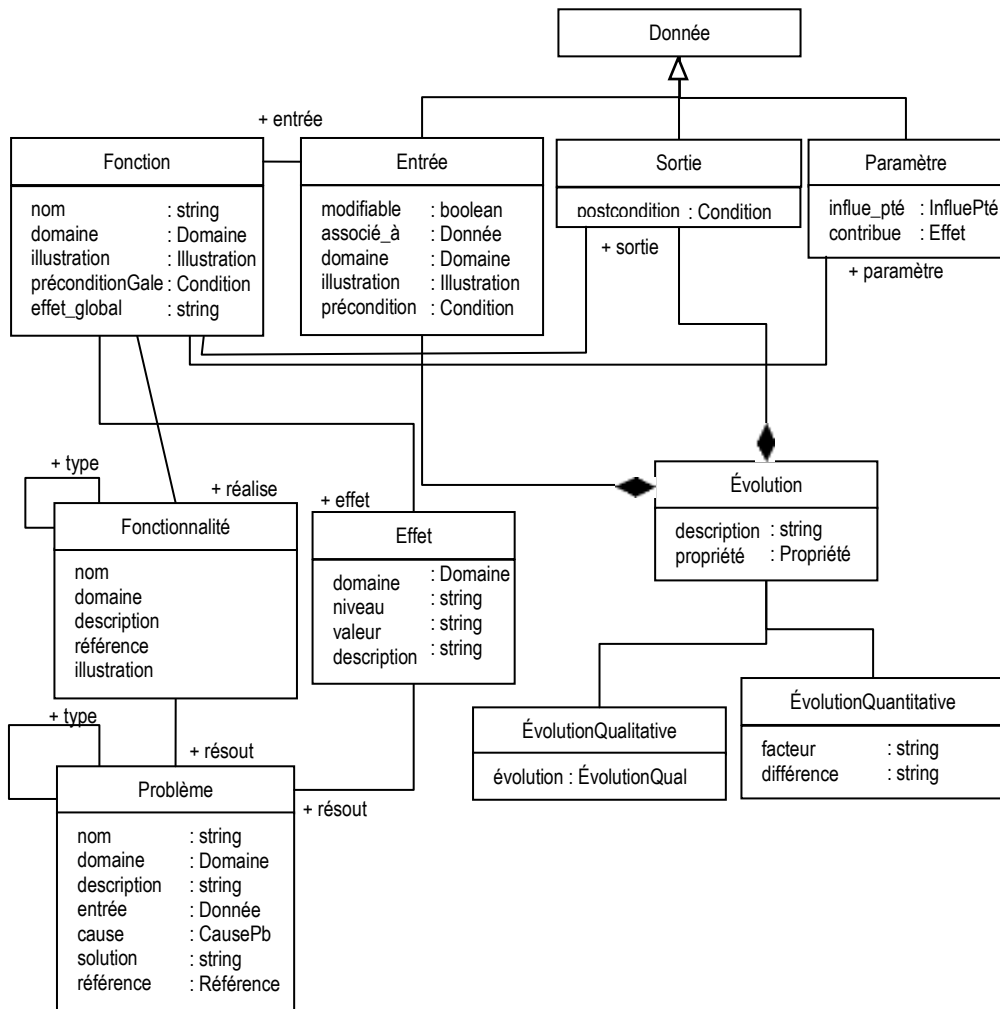


Figure 46. Modèle MDT pour la description détaillée d'une Fonction.

Une différence clé entre MDT et OWL-S se situe dans la facette ModeEmploi. En effet la facette Grounding d'un service Web décrit des échanges de messages XML qui conduisent à la fonction voulue. Ici les interactions possibles avec les ressources logicielles sont autres. (Abd-el-Kader 2006) propose un modèle générique qui veut décrire toutes les interactions possibles avec une ressource logicielle pour accéder à un traitement supporté par cette ressource. Ces interactions peuvent se faire via une interface graphique ou programmatique.

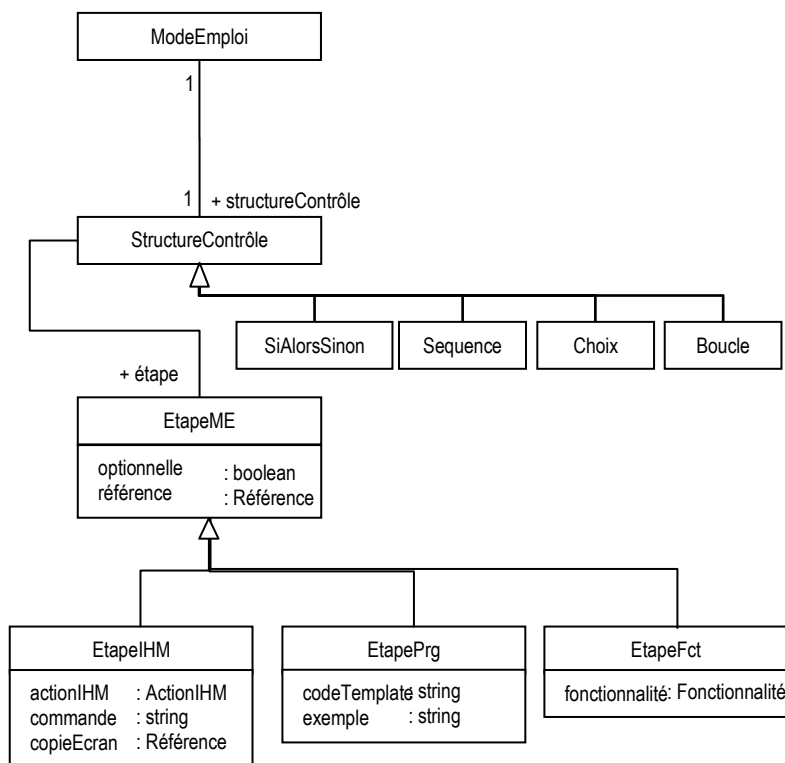


Figure 47. Détail de la facette ModeEmploi du modèle MDT qui décrit comment accéder à un traitement codé encapsulé dans une ressource logicielle.

Le deuxième volet de sa proposition consiste en un **système expert qui exploite des connaissances contextuelles pour faciliter la consultation de la base**. Ce système peut adapter un Mode d'emploi à un contexte. Le contexte est constitué des données d'entrée, de la sortie attendue, de l'expertise de l'utilisateur et de son environnement de travail. L'utilisateur saisit d'abord le contexte dans lequel il veut faire appel au traitement. Le système évalue l'écart entre ce contexte utilisateur et le contexte requis par le Mode d'emploi et déclenche des règles d'adaptation du mode d'emploi pour combler cet écart.

2.1.4.2 Synthèse des enseignements liés à TAGE et MDT

Cette section résume les enseignements des travaux qui précèdent (TAGE et MDT) avant la présentation d'une proposition intégrant ces acquis (MFM en section 2.1.4.3)

Les descripteurs importants pour faciliter le partage de ressources logicielles géographiques sont ceux qui rendent compte de l'utilité du logiciel dans divers contextes –les fonctions qu'il peut fournir et dans quelles conditions-. Les 'conditions' comprennent par exemple des propriétés spécifiques des données manipulées (taille du jeu, valeurs manquantes, granularité des objets, type de paysage). Il est important de souligner que la plupart des fonctions géographiques ne sont pas simplement disponibles dans certains contextes et indisponibles dans d'autres. Il est plus exact de dire que la fonction fournie n'est pas tout à fait la même en fonction du contexte. Ces descripteurs peuvent se résumer par le concept de *services rendus par la ressource*.

Nous avons aussi appris que **lorsque le traitement est cartographique, le recours à des illustrations est indispensable pour les experts** pour évaluer la fonction fournie par le traitement. Ce dernier enseignement est lié également à la proposition de thèse de (Hubert 2003) dirigée par Anne Nicole de l'Université de Caen et encadrée par Anne Ruas au laboratoire COGIT. Frédéric Hubert a proposé une interface permettant à un opérateur de paramétrer un traitement de généralisation de bâtiments. La difficulté réside dans le fait que cet opérateur ne connaît pas le fonctionnement du traitement et ne peut donc pas comprendre les paramètres et les valeurs qu'il doit leur donner. Pour l'aider à déterminer les bonnes valeurs de ces paramètres, l'application lui présente des exemples de bâtiments généralisés avec ce traitement et fait varier les valeurs des paramètres ainsi que les bâtiments. En effet, le traitement présente des comportements différents selon les propriétés des bâtiments (forme et granularité). Dans notre contexte, pour permettre à l'auteur de la description d'un traitement de rendre compte des variations de comportement de son traitement en fonction de variation de propriétés de l'entrée comporte plusieurs difficultés. Il faut lui permettre de décrire les propriétés pertinentes du contexte alors qu'elles peuvent se situer à différents niveaux d'abstraction. Il faut aussi lui permettre de 'paramétrer' la fonction (ses effets) par ces propriétés. Cela pose des défis non seulement pour le modèle de description mais aussi pour l'acquisition des valeurs des descripteurs sur un traitement donné.

Un autre critère déterminant pour l'accès aux traitements –ou le partage des ressources logicielles- dans le domaine des données géographiques est **le modèle de données sous-jacent**. En effet, les modèles délimitent souvent des familles de fonctions et des domaines d'expertise. Le terme modèle est ici très générique, il peut renvoyer ici aux

éléments de la réalité représentés (les bâtiments, les réseaux), ou encore à un mode de représentation (données 3D, MNT, données vecteur).

Un apport important de ces travaux a également été d'**explorer le lien entre ressource et fonction** et la complexité de ce lien.

Nous avons aussi été confrontés à la nécessité de **faire un compromis entre les exigences fonctionnelles que nous prêtons au modèle de métadonnées et les efforts de formalisation que les auteurs de description sont prêts à accepter**.

Par ailleurs, il est apparu qu'une grande difficulté consistait à **mobiliser les auteurs pressentis de métadonnées**, c'est-à-dire les personnes possédant l'expertise nécessaires pour connaître les fonctions fournies par les logiciels. Ce problème est récurrent dans le domaine des métadonnées. Dans le cas de description de traitements, nous voulons préciser que les auteurs pertinents ne sont pas des experts en Services Web mais plutôt les développeurs et les utilisateurs de ressources logicielles (Bucher et Jolivet 2008). Les modèles TAGE et MDT présentent de toute évidence des obstacles à leur appropriation par ces auteurs :

- TAGE est basé sur une conceptualisation qui n'est pas familière aux chercheurs et aux développeurs. Pour expliquer TAGE, je devais traduire chaque concept important par un terme plus familier ('une tâche est comme une activité dans un diagramme d'activité', 'un rôle est une variable décrite par la fonction qu'elle a dans... l'activité', ...). De plus, la représentation des structures de décomposition dans TAGE est un peu l'équivalent de la notation polonaise inversée dans les calculatrices ; elle n'est pas intuitive et demande d'aller à l'encontre d'une façon plus naturelle de construire un processus qui est le diagramme d'activités. C'est une façon plus naturelle pour la génération d'ingénieur qui a été formée à SADT, cela ne le sera peut-être plus par la suite.

- MDT est un modèle extrêmement structuré dont le concepteur a visé l'opérationnalité avant tout. Le problème est que cette structure est imposée car elle n'a pu s'appuyer sur une structure existante ou naturelle. Elle demande donc un effort conséquent aux auteurs pour la comprendre et l'adopter. Cet effort est d'autant plus lourd lorsqu'il existe, pour certains champs, des modèles pour les décrire. Par exemple, les types de données en entrée et en sortie doivent être décrits selon un formalisme spécifique MDT. Un objet manipulé ne peut être en entrée et en sortie –ce qui correspond à une pratique recommandée mais pas à la réalité–.

-

Dans la conception d'un nouveau modèle qui capitalise sur ces premières expériences, l'accent a donc été mis davantage sur la création des métadonnées que sur leur exploitation par un système intelligent.

Un autre enseignement est **qu'il est parfois difficile pour l'auteur des descriptions de relier les descripteurs pertinentes (fonctions, modèles) aux ressources qui leur sont familières** (Bucher et Jolivet 2008). Distinguer dans une application des fonctions

et les nommer n'est pas toujours simple selon la structure de cette application. Il nous a également semblé que la notion de service est particulièrement complexe à comprendre parce qu'elle est liée à la fois à la notion de ressource logicielle et à la notion de contexte d'utilisation. C'est pourquoi nous proposons d'impliquer dans la description plusieurs acteurs : les développeurs et les utilisateurs. Par ailleurs, comme nous l'avons mentionné plus haut, il peut être difficile d'évaluer la valeur de certains descripteurs quand bien même la signification de ces descripteurs est comprise, c'est le cas des préconditions et des effets.

Concernant la modélisation elle-même, il ressort de l'état de l'art et de ces premiers travaux que la description d'un processus comporte d'une part la description d'un domaine, les objets manipulés, et d'autre part la description de comportements, les manipulations elles-mêmes. Un modèle de description peut se focaliser sur un aspect ou l'autre. Pour construire plus généralement des descriptions partageables de comportements, il faut les appuyer sur une description partageable du domaine. Ainsi dans le domaine des services Web, des ontologies de domaines comme le tourisme sont développées pour accompagner la description des comportements des services Web. En IA, les chercheurs préconisent de conserver une indépendance entre la description de ce qui relève de la nature des objets manipulés, c'est-à-dire les ontologies de domaines, et les connaissances de manipulation.

Globalement, il semble qu'un certain consensus désigne les connaissances de contrôle comme les plus délicates à modéliser, que ce soit synchroniser des actions ou organiser une négociation entre plusieurs agents. De notre propre expérience, exprimer le lien entre une description d'un processus et celle d'un autre processus plus spécifique est aussi délicat, ainsi que construire et manipuler les descriptions des rôles du comportement (Bucher 2003). Enfin, ces modèles sont tous complexes et d'un abord difficile. Pour autant ils sont importants. L'expert qui décrit un processus respecte rarement les enseignements que nous venons de lister. Il est donc nécessaire d'utiliser un cadre formel pour décrire des processus.

Enfin, il existe plusieurs **correspondances entre MDT et TAGE** qui sont rappelées sur la Figure 48. Tout d'abord, par définition des objectifs de MDT, le ModeEmploi correspond au Mécanisme des tâches primitives de TAGE. La facette Fonction dans MDT comporte les informations groupées sous l'appellation Contexte dans les tâches de TAGE. Elle propose aussi, conformément à OWL-S, de décrire de façon distincte le type générique de données d'une entrée ou sortie et des restrictions éventuelles (pas de valeur manquante, pas de bâtiments isolés, ...) dans des préconditions et postconditions. Cela conduit à des descriptions plus simples et lisibles que le choix fait dans TAGE de grouper toutes les informations relatives aux valeurs que peuvent prendre les entrées et sorties dans la définition du domaine de valeur d'un rôle.

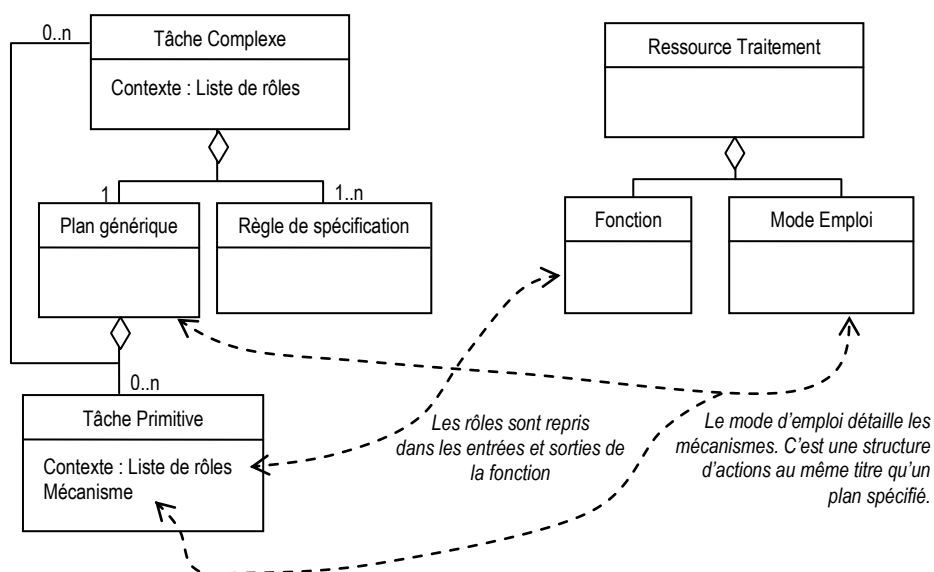


Figure 48. Principales correspondances entre les modèles TAGE (à gauche) et MDT (à droite).

2.1.4.3 MFM : Un modèle de métadonnées orienté-services pour le catalogage de ressources logicielles et le déploiement de services web pertinents et utilisables

Cette section présente **le modèle que je propose actuellement pour construire des métadonnées à propos du traitement de données géographiques (ressources logicielles et savoir-faire associé)**. En pratique, j'ai pris en compte les complémentarités des modèles TAGE et MDT ainsi que d'autres modèles pertinents comme le modèle de (Balley 2006) pour décrire la structure de données géographiques ou le modèle de (Gesbert 2006) pour décrire les spécifications de bases de données IGN. J'ai aussi tenu compte des exigences imposées par les considérations listées ci-avant. Ce modèle s'appelle MFM pour Métadonnées Fonctions-Modèles de ressources logicielles. Il est présenté dans la suite de cette section en précisant ses évolutions par rapport à TAGE et MDT.

Il existe en réalité deux versions de ce modèle :

- une version principale présentée dans cette section et dont l'exploitation actuelle est présentée en section 2.1.4.4.
- une version documentaire, dédiée à la constitution d'un vocabulaire commun de description des fonctions et présentée en section 2.1.4.5.

Le **modèle général** est présenté sur la **Figure 49**. Son originalité est l'introduction du concept d'Activité pour faire le lien entre le concept de Fonction (service rendu) et les ressources logicielles. Intuitivement, ce modèle permet d'associer à des ressources logicielles, par exemple des bibliothèques Geoxygene, des métadonnées sous la forme d'Activité qui décrivent comment effectuer la Fonction en interagissant avec la ressource. Les entrées et sorties de la fonction sont décrites comme des variables dont le domaine de valeur peut correspondre à une Structure de données (ou Modèle de données). Une telle structure est éventuellement associée à une ou plusieurs ressources logicielles qui sont une implémentation de son schéma logique.

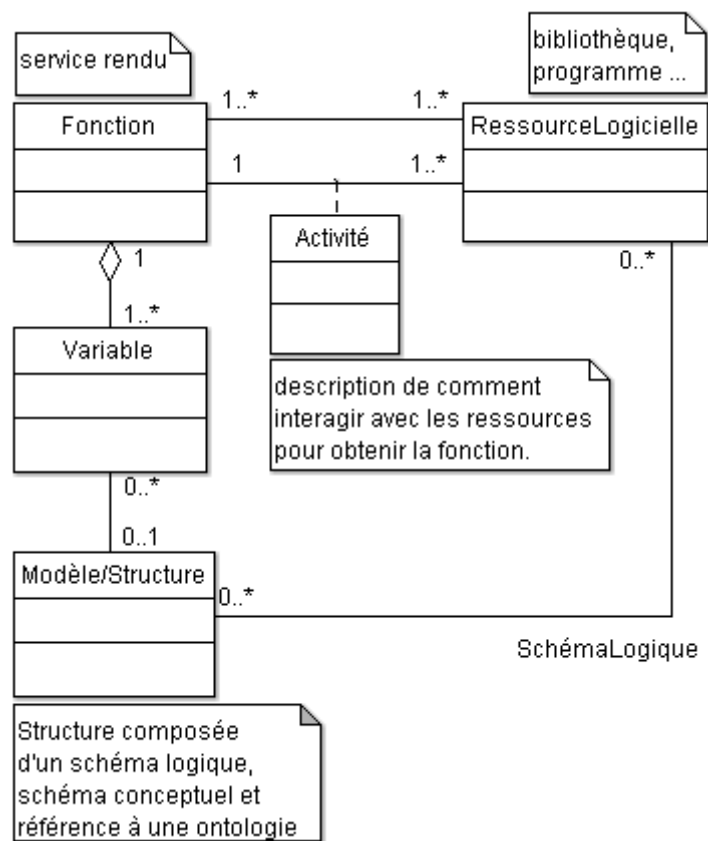


Figure 49. Diagramme principal de classes de MFM, modèle de métadonnées de fonctions et de modèles pour des ressources logicielles.

Comme annoncé plus haut, la notion de **Fonction** demeure un descripteur essentiel pour évaluer la pertinence d'une ressource dans un contexte donné. Les éléments dédiés à la description de la notion de Fonction sont illustrés sur la Figure 50. Pour décrire les diverses entrées et sorties j'ai repris le modèle ISO19119 qui préconise d'associer à un service une liste de Paramètres dont chacun à une direction qui vaut 'in',

'ou' ou 'inout'. Cela se rapproche de la modélisation du contexte dans TAGE. Par contre, je n'ai gardé ni la dénomination 'paramètre' d'ISO19119 ni la dénomination 'rôle' de TAGE. Il semble que le nom 'Variable' est mieux compris

Le Modèle/Structure correspond à la notion de structure formalisée dans la thèse de Sandrine Balley. Nous avons fait évoluer ce modèle pour tenir compte de plusieurs faits :

- Un élément de schéma conceptuel ne correspond pas forcément à un concept d'une ontologie mais plutôt à un élément d'une ontologie (concept ou attribut).
- Le lien élément de schéma et élément d'ontologie est un lien [n,m] décrit par des règles de sélection et de représentation conformément au modèle proposé par (Gesbert 2006) et repris par (Abadie 2009) pour décrire des spécifications d'une base de données géographiques.
- Enfin, dans certains cas on ne dispose pas d'ontologie ou on ne dispose pas de schéma conceptuel. Cela n'est pas forcément souhaitable mais dans ces deux cas, notre rôle n'est pas de contraindre les développeurs à introduire ces niveaux dans leurs modèles, nous cherchons uniquement à leur permettre de décrire leurs ressources pour mieux les partager.

Par rapport aux choix faits dans MDT, j'ai décidé de n'introduire une formalisation que lorsque cela était nécessaire et dans un premier temps surtout pour gérer les ambiguïtés dans les descriptions (éviter que des ressources soient mal identifiées et que des mots clés soient mal interprétés). Concernant la description des Fonctions et des Modèles, il est extrêmement important d'éviter les ambiguïtés dans la sémantique d'une Fonction. L'idéal serait donc de disposer d'une ontologie des fonctionnalités d'intérêt (comme filtrer, apparier, simplifier) et d'une ontologie des modèles d'intérêt (graphe, alignement, etc). Malheureusement il n'existe pas de telle ontologie qui couvre le domaine des fonctions et modèles SIG. Cela avait été relevé dans mon travail de thèse (Bucher 2002). Comme cela est le cas pour l'indexation des services Web commerciaux, nous proposons d'associer à une fonction non pas une catégorie dans une classification figée mais plutôt un ensemble de catégories pouvant renvoyer à diverses classifications. Une fonction peut être définie comme fonction de référence (auquel cas l'uri de son référentiel est documenté). Elle peut aussi reprendre une fonction de référence c'est-à-dire hériter de ses variables et autres propriétés. Cela est renseigné dans le champs 'referencesSemantiques' qui est une liste d'identifiants de Fonction de la forme 'uriReferentiel:nomFonction'.

Pour un lecteur humain, des descripteurs importants de la fonction sont son nom, ses références Sémantiques (c'est-à-dire les classes de fonction auxquelles elle correspond dans des classifications), sa description, ses effets et ses variables. Un effet est une description d'un changement de l'état du monde opéré par l'application de la fonction.

Pour un lecteur plus formel, des descripteurs importants sont les variables, les préconditions et les postconditions. Les préconditions et postconditions viennent

certaines ressources logicielles (par exemple les classes java, la famille des produits Geoconcept, etc). Le seul élément indispensable à ce niveau est l'identification de la ressource dont il est question.

La maîtrise du **lien Fonction – Ressource Logicielle** est importante pour que les auteurs comprennent quelles Fonctions ils peuvent associer aux ressources logicielles qu'ils connaissent. Le lien Fonction – Ressource Logicielle correspond dans MDT au mode d'emploi de la ressource pour atteindre la fonction. Et il correspond dans TAGE soit au mécanisme d'une tâche primitive soit à la méthode de résolution d'une tâche (plan générique et connaissances stratégiques associées). En effet ces connaissances dans TAGE font le lien entre les ressources logicielles citées dans les mécanismes et les 'contexte' des tâches. Que ce soit dans MDT ou TAGE, les modèles proposés sont relativement complexes à comprendre et à documenter. Et dans les deux cas, il s'agit de 'structures d'actions' plus ou moins génériques. Dans le modèle actuel nous avons donc représenté le lien Fonction – Ressource Logicielle comme une structure d'actions plus ou moins générique et nous avons aussi cherché à simplifier la façon dont est représentée cette structure d'actions. Nous proposons de la représenter par un nouveau concept, **l'Activité**. En effet, ce concept est relativement familier et compréhensible à l'IGN où les diagrammes SADT ont longtemps utilisés pour la modélisation puis les diagrammes d'activités UML (Bucher et al. 2005). Le concept d'Activité sert à décrire la structure d'un processus global sous forme d'un diagramme d'activités, en reprenant les stéréotypes UML2 du diagramme d'activité (Nœuds, Arêtes, ..).

Il a été nécessaire de ne **pas reprendre les stéréotypes UML2 mais de s'en inspirer pour aboutir à un modèle plus simple (car moins ambitieux)**. Une première simplification consiste à éliminer les connaissances complexes liées à l'implémentation, comme les envois de messages ou le stockage des objets en attente de consommation par une activité. Globalement, la gestion du temps n'a pas été traitée ce qui nous permet d'ôter du modèle des notions de synchronisation qui complexifient grandement les connaissances de contrôle. La simplification suivante consiste à réduire le nombre de classes utilisées. Dans MFM, deux classes sont gardées pour décrire les nœuds des diagrammes d'activités : les Contrôles et les Activités. Je ne garde pas le concept d'Action. Dans UML2, une Action est une unité de travail exécutable sans interaction forte de la part de l'utilisateur. Cela peut consister à invoquer une Activité dans un contexte spécifique. Cette distinction revient souvent dans la littérature. Dans les modèles à base de tâches, on distingue ainsi les tâches primitives des tâches complexes. Dans notre modèle, une unité de travail exécutable demeure une Activité ou un Contrôle. Nous justifions ce choix par notre interprétation du concept Activité. Une activité est un ensemble de réalisations passées ou possibles. Un intérêt immédiat est de faciliter le travail de l'auteur d'une description : il n'a pas besoin de choisir si un processus spécifique est une Action ou une Activité. Cette interprétation est également

importante parce que c'est sur elle que nous baserons notre interprétation de la relation de spécialisation entre activités.

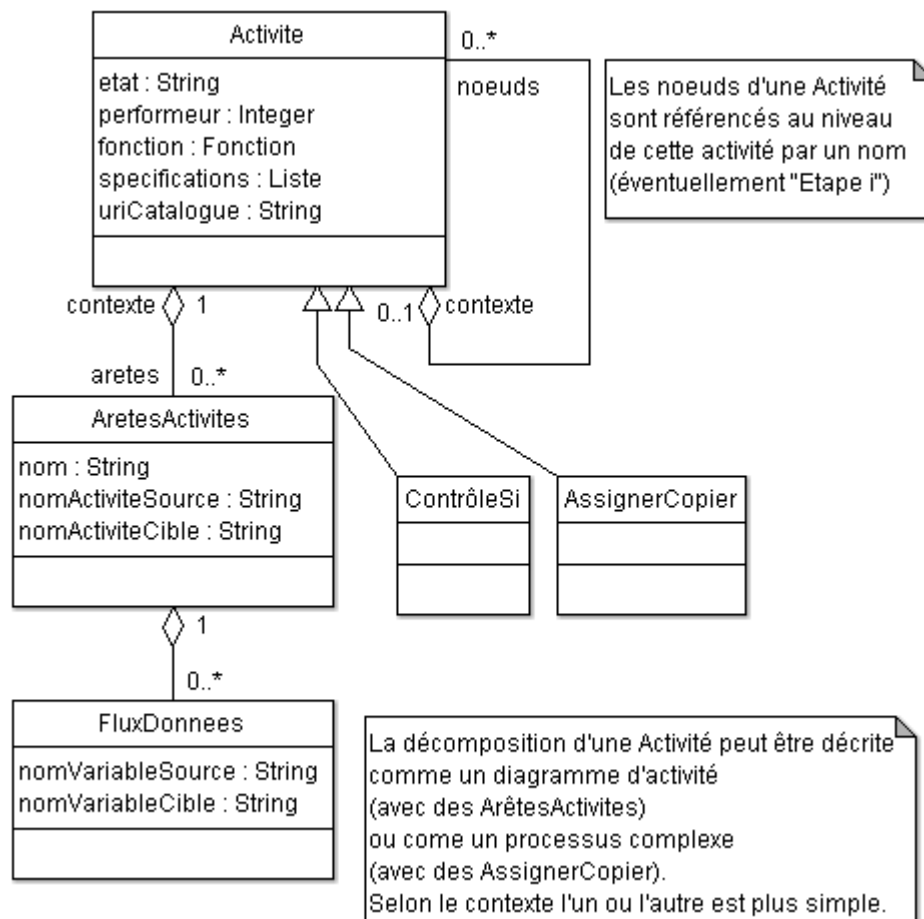


Figure 51. Modèle de description d'une activité.

La classe **Activité** est dédiée à permettre la spécialisation et l'exécution d'un processus qui permet d'obtenir une **Fonction** désirée à partir de **Ressources**. Un champ important est le champ 'estPrédéfini' que nous avons introduit pour distinguer les gabarits d'activité qui ne peut être modifié qu'en étant spécialisés d'activités en cours de construction (par exemple un gabarit en cours d'édition ou un historique de session). Cela correspond à la distinction Tâche élémentaire (Activité prédéfinie) et Tâche principale (en cours de construction) de (Lopez-Velasco et al. 2009). La classe générique *Activité* est destinée à être spécialisée pour représenter les boucles de contrôle, les noeuds début et fin d'un diagramme d'activités et enfin des activités particulières comme invoquer une méthode java. Nous indiquons sa structure en Figure 51.

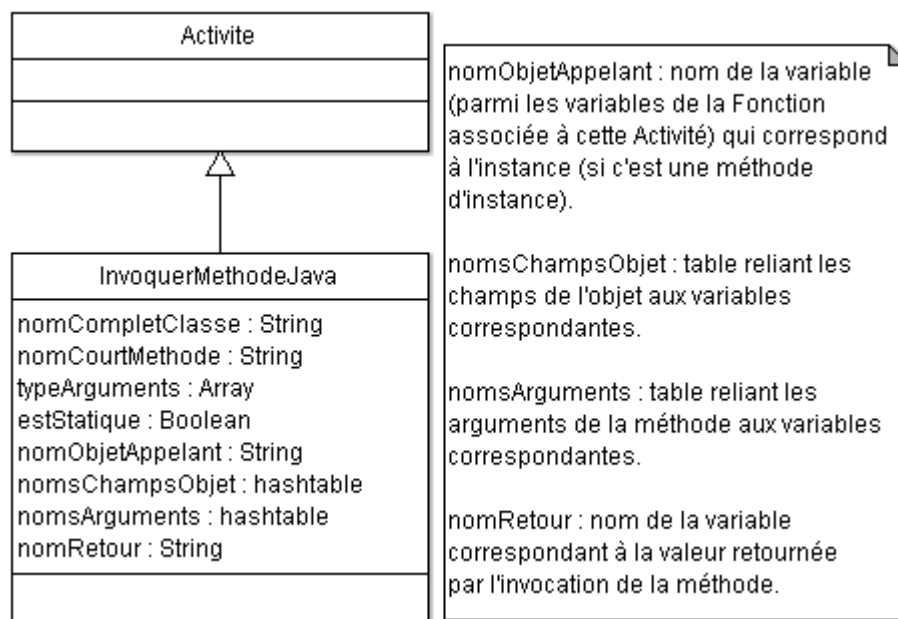


Figure 52. Modèle de description d'une activité et spécialisation pour décrire les types d'activité qui correspondent à l'invocation d'une méthode java.

La spécialisation `InvoquerMethodeJava` permet essentiellement de faire automatiquement le lien entre les variables d'une fonction et les éléments java nécessaires à l'appel de la méthode : l'objet appelant si la méthode n'est pas statique, les champs de cet objet ou les champs de la classe si la méthode est statique, les arguments et l'objet renvoyé.

Cette approche rejoint en définitive celle de (Becam 2002) qui a proposé une solution à base de métadonnées et de pilotes pour faire interopérer des modèles urbains. L'auteur utilisait déjà la métaphore de l'orchestre pour décrire son objectif, terme repris ensuite dans le domaine des Service Web. Les modèles urbains sont des procédures dérivant des objets spécifiques de la ville, ils sont décrits à un niveau logique et physique. (Becam 2002) s'est concentré sur le traitement de tous les problèmes d'interopérabilité alors que je considère que les traitements à chaîner éventuellement sont implémentés dans la même plate-forme. Enfin, (Becam 2002) a soigné particulièrement la synchronisation des modèles.

2.1.4.4 Découverte et exploration à distance de traitements implémentés dans GeOxygène fondées sur MFM

Le modèle MFM a été utilisé pour implémenter des prototypes de catalogues de traitements. Un premier prototype vise à faciliter, à un utilisateur sur le Web, l'accès à

des traitements qui peuvent s'exécuter sur un serveur. L'utilisateur découvre, spécifie et déclenche à distance des traitements à l'aide de leurs descriptions. L'objectif second de ce prototype était de sensibiliser les experts à l'utilité des métadonnées pour partager leurs ressources. Une première version a été proposée dans le travail de stage de Brice Aramini. Le stagiaire a proposé une interface Web permettant d'invoquer à distance un traitement exécuté sur le serveur dans la plate-forme Geoxygene et de visualiser le résultat au format SVG (Aramini 2006). Ce prototype a été repris et enrichi par la suite (Bucher ICC2007). J'ai distingué dans les missions du serveur deux phases de consultation.

- 1) La première phase est la découverte. Elle vise à identifier les traitements potentiellement d'intérêt. Elle est supportée par les mots-clés attachés aux fonctions et aux modèles. Elle s'appuie donc sur une indexation des ressources par les fonctions et les modèles. Dans le prototype, l'utilisateur navigue dans une hiérarchie de fonctions comme illustré sur la Figure 53.

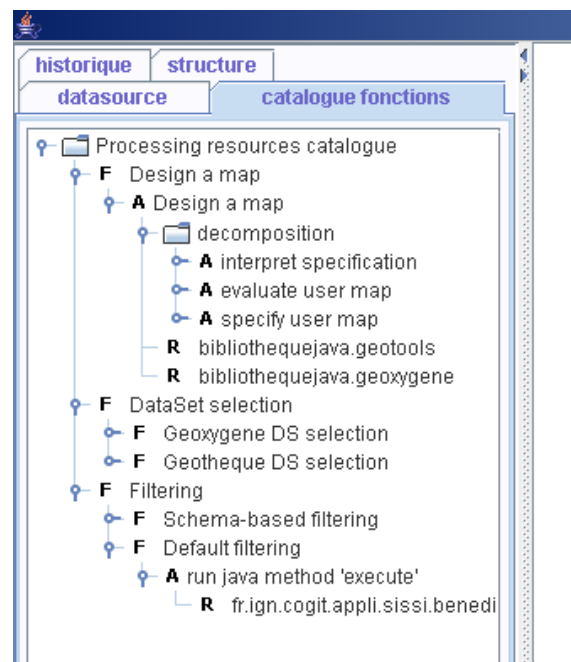


Figure 53. Catalogue indexant les ressources (nœuds étiquetés 'R') par les les fonctions offertes (nœuds étiquetés 'F'). La liaison entre un nœud ressource et un nœud fonction repose sur des activités.

- 2) La deuxième phase est l'exploration. C'est une phase plus délicate. Elle doit permettre à l'utilisateur d'évaluer la pertinence dans son contexte des ressources qui ont été découvertes et de comparer les pertinences de plusieurs ressources.

Nous avons proposé la décomposition suivante de la pertinence d'un traitement (Bucher ICC 2007):

- Est-ce que la fonction réalisée correspond aux attentes de l'utilisateur ? il s'agit là d'une notion d'efficacité du traitement. Nous détaillons ci-après comment cette fonction est supportée.
- Quel est le coût pour réaliser cette fonction à l'aide des ressources? Il s'agit là davantage d'une notion d'efficience. Pour évaluer ce deuxième aspect, le coût, je propose que l'utilisateur s'appuie sur les éléments de description de l'objet Activité qui fait le lien entre la fonction fournie et la ressource logicielle et évalue le nombre de sous activités, les logiciels nécessaires, la qualité des descriptions.

Pour évaluer l'efficacité d'un traitement dans un contexte donné, notre prototype comporte la possibilité pour l'utilisateur de charger des données sur le serveur, de paramétrer le traitement, de l'appliquer aux données et de visualiser le résultat sur une portion du jeu (pour des raisons de performance du serveur et du client, les données complètes ne sont pas traitées mais seulement un extrait). Le concept de carte d'activité a été ajouté au modèle pour désigner une représentation cartographique dédiée non pas à représenter un espace géographique mais à rendre compte des effets d'une Activité. Pour visualiser les effets du traitement, il est utile de pouvoir regrouper dans des classes graphiques les objets détruits, créés ou modifiés. Cela n'est possible que si le traitement possède une interface programmatique permettant de récupérer les identifiants des objets détruits, créés ou modifiés. Par ailleurs, pour comparer la pertinence de deux traitements sélectionnés, l'utilisateur peut les appliquer à tour de rôle sur les données initiales et visualiser les résultats sur deux cartes voisines. Le modèle permet la gestion de l'historique de la session et les retours à des états antérieurs du jeu en cours de traitement pour appliquer un autre traitement ou un autre paramétrage. Cela est illustré sur la Figure 54.

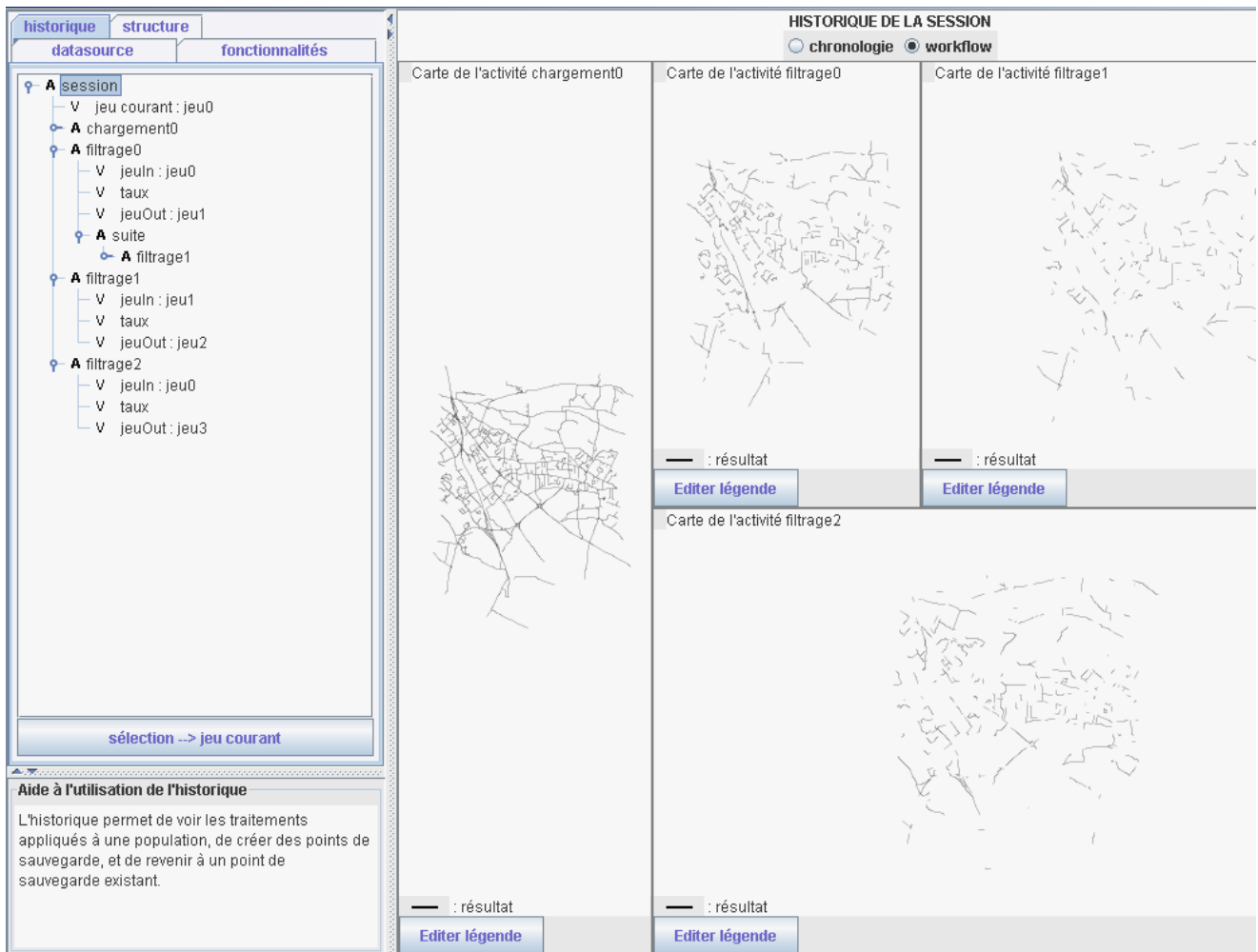


Figure 54. Utilisation du catalogue pour comparer le résultat de deux séquences différentes d'activités.

Ce prototype Web a été porté dans une interface graphique commune aux développeurs sous Geoxygene (à l'époque, OpenJUMP), sous forme d'une application locale. L'objectif poursuivi est de lui coupler une interface d'édition pour que le développeur Geoxygene qui consulte le catalogue soit d'autant plus incité à documenter ses ressources dans notre modèle qu'il les voit immédiatement apparaître dans le catalogue. Dans le nouveau prototype, les métadonnées sont organisées en deux types d'enregistrements, stockés dans des fichiers XML : les catalogues de traitements (qui correspondent à des listes d'Activités) et les référentiels de fonctions, illustrés sur la Figure 55.

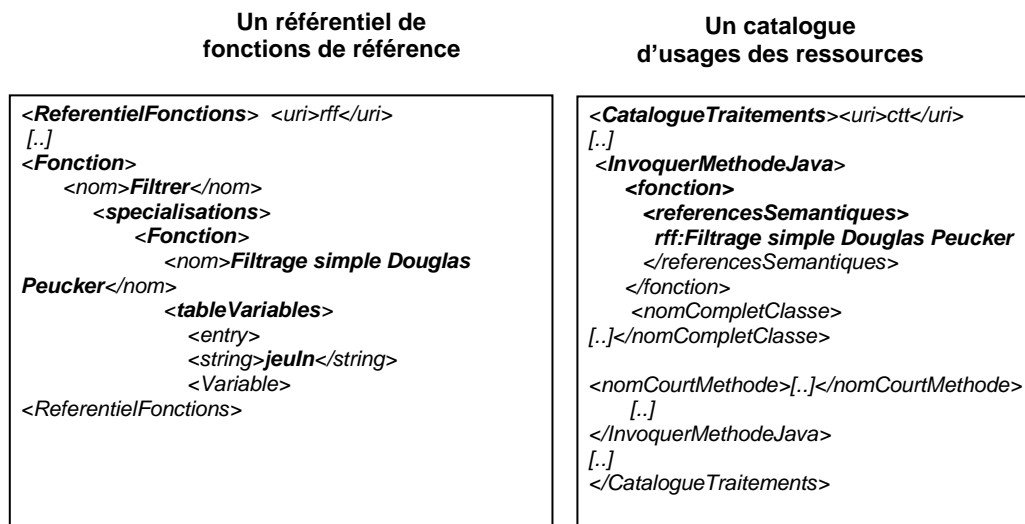


Figure 55. Organisation de métadonnées MFM en deux types de ressources : les référentiels de fonction de référence et les catalogues d'usages de ressources. Un catalogue d'usage de ressources comporte des Activités dont les Fonctions peuvent être renseignées en faisant référence à une fonction dans un référentiel de Fonctions.

L'utilisateur consulte ces métadonnées XML via une interface graphique qui lui permet de consulter les traitements décrits (c'est-à-dire les Activités). Lorsqu'il a choisi un traitement, il peut le spécifier et l'appliquer à ses données comme illustré sur la Figure 56.

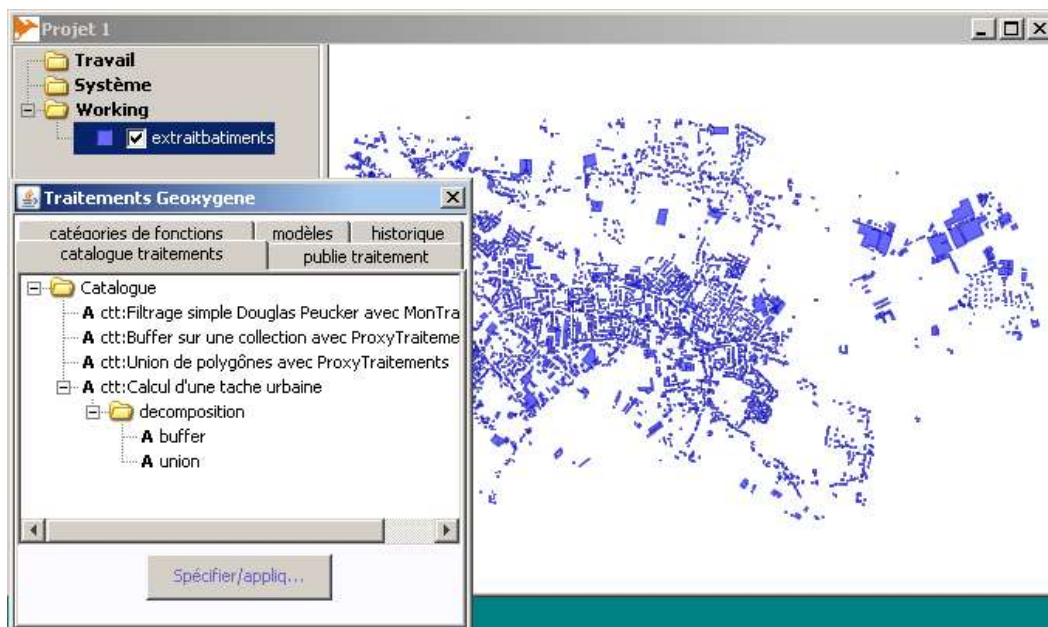


Figure 56. Prototype dans Geoxygene permettant de consulter la base de traitements disponibles.

Le résultat apparaît alors sous forme d'une nouvelle couche comme illustré sur la Figure 57.

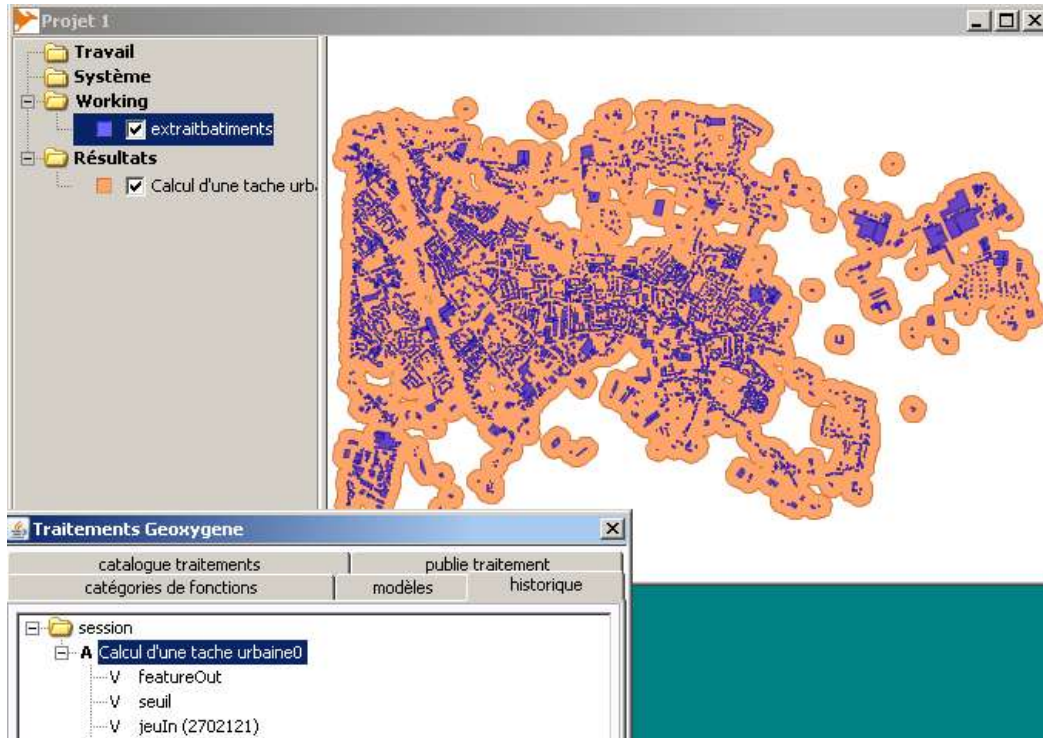


Figure 57. Visualisation de l'application d'un traitement : le calcul d'une tâche urbaine.

2.1.4.5 Une version documentaire de MFM dédiée à la constitution d'un vocabulaire de description

Un catalogue Web documentaire a été développé qui est destiné à l'échange d'information entre des humains et à la construction de vocabulaire de référence pour parler des traitements. L'objet de ce catalogue est de référencer les traitements implémentés (dans Geoxygene ou ailleurs) ainsi que les algorithmes non implémentés disponibles au laboratoire COGIT. Le but du catalogue est que les descriptions soient consultables et éditables par des humains. Le modèle sous-jacent a donc été extrêmement simplifié par rapport au modèle MFM. Les objets *Activité* qui font le lien entre les *Fonctions* et les *Ressources Logicielles* ont été réduits à une description en texte libre car le but n'est pas d'automatiser l'invocation mais de découvrir l'existence de ressources d'intérêt. Le modèle est présenté en Figure 58.

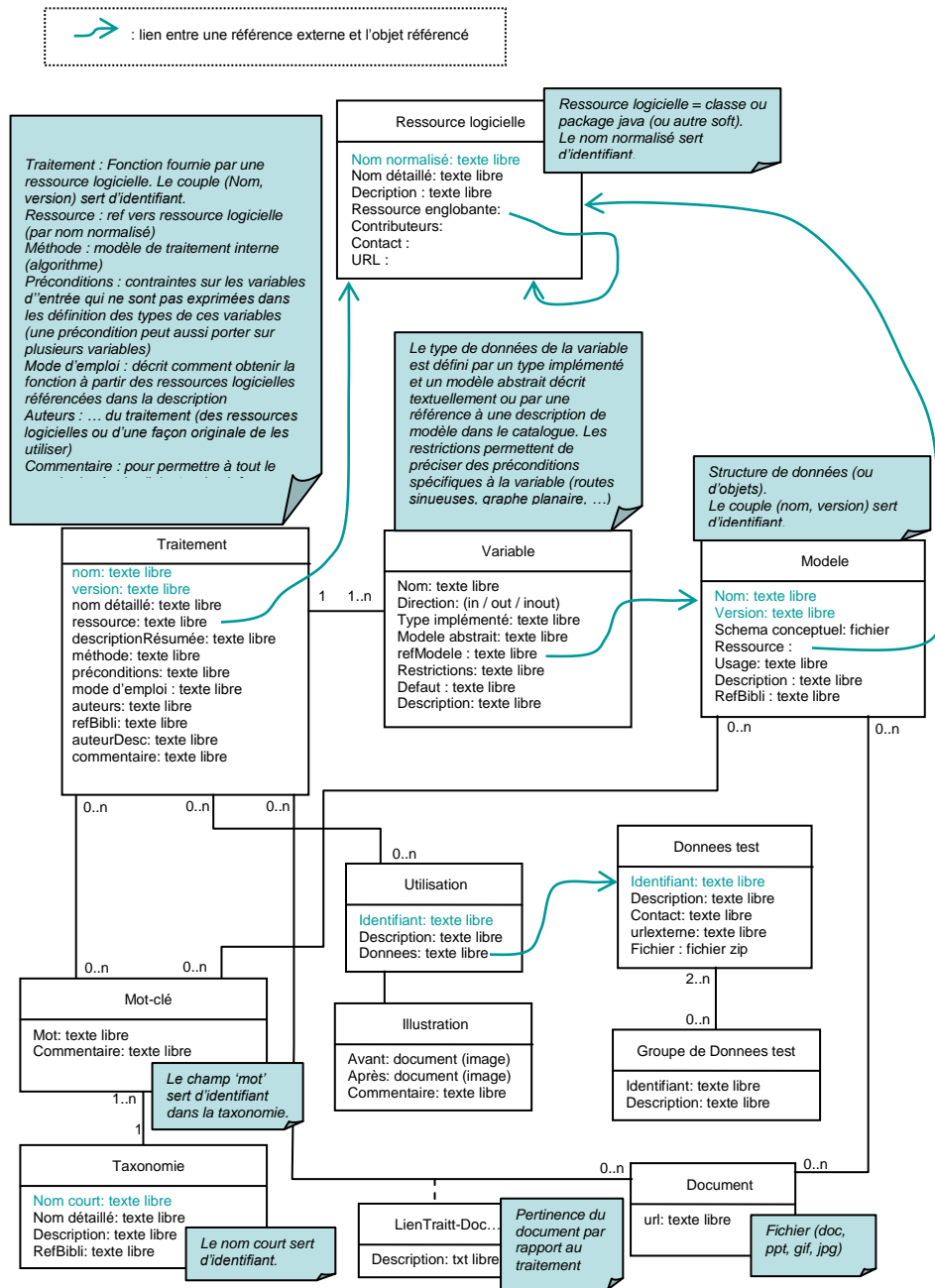


Figure 58. Version documentaire de MFM utilisée dans le catalogue documentaire Web.

L'accent a été mis sur la description des *Fonctions* et des *Modèles* et sur la constitution de vocabulaires de référence pour remplir ces descriptions. Ce catalogue permet donc de publier des taxonomies.

<p>Accueil</p> <ul style="list-style-type: none"> ◦ Retour <p>Listes</p> <ul style="list-style-type: none"> ◦ Traitements <ul style="list-style-type: none"> ◦ Tous ◦ Nouveau ◦ Modèles <ul style="list-style-type: none"> ◦ Tous ◦ Nouveau ◦ Mots clés <p>Support</p> <ul style="list-style-type: none"> ◦ Forum ◦ Documentation 	<p>Mots clés (chaque mot est défini dans une taxonomie)</p>										
<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 20%; padding: 5px;">Nom court</td> <td style="padding: 5px;">MustiereGen</td> </tr> <tr> <td style="padding: 5px;">Nom détaillé</td> <td style="padding: 5px;">Taxonomie des opérations de généralisation proposée par Seb</td> </tr> <tr> <td style="padding: 5px;">Description</td> <td style="padding: 5px;">Taxonomie organisant les opérations de généralisation en fonction du but poursuivi et comportant 3 niveaux : - les grandes familles d'opération - les opérations - les actions</td> </tr> <tr> <td style="padding: 5px;">Références bibliographiques</td> <td style="padding: 5px;">Mustière Sébastien, Apprentissage supervisé pour la généralisation cartographique, Thèse en informatique - intelligence artificielle de l'Université Paris 6, 2001 http://recherche.ign.fr/labos/cogit/pdf/THESES/MUSTIERE/These_Mustiere_2001.pdf</td> </tr> <tr> <td style="padding: 5px;">Contenu</td> <td style="padding: 5px;"> <ul style="list-style-type: none"> • simplifier, Famille d'opérations dont le but est d'éliminer de l'information, éditer, supprimer • caricaturer, famille d'opérations dont le but est d'exagérer des caractères, éditer, supprimer • harmoniser, famille d'opérations dont le but est de gommer des différences, éditer, supprimer • généralisation de modèle, opérations visant à rendre les objets similaires, éditer, supprimer • sélection-élimination, opérations visant à diminuer le nb d'objets, éditer, supprimer • sélection, opération qui sélectionne les objet à garder lors d'une opération de type sélection-élimination, éditer, supprimer • élimination, opération qui identifie les objets à éliminer lors d'une opération de sélection-élimination, éditer, supprimer • amplification, exagérer la forme et la taille, éditer, supprimer • lissage-filtrage, opérations visant à atténuer ou éliminer les parties d'un objet, éditer, supprimer • lissage, opération visant l'atténuation des parties d'un objet, éditer, supprimer • filtrage, opération visant l'élimination de parties d'un objet, éditer, supprimer • amplification, opération visant l'exagération de la forme ou de la taille, éditer, supprimer </td> </tr> </table>	Nom court	MustiereGen	Nom détaillé	Taxonomie des opérations de généralisation proposée par Seb	Description	Taxonomie organisant les opérations de généralisation en fonction du but poursuivi et comportant 3 niveaux : - les grandes familles d'opération - les opérations - les actions	Références bibliographiques	Mustière Sébastien, Apprentissage supervisé pour la généralisation cartographique, Thèse en informatique - intelligence artificielle de l'Université Paris 6, 2001 http://recherche.ign.fr/labos/cogit/pdf/THESES/MUSTIERE/These_Mustiere_2001.pdf	Contenu	<ul style="list-style-type: none"> • simplifier, Famille d'opérations dont le but est d'éliminer de l'information, éditer, supprimer • caricaturer, famille d'opérations dont le but est d'exagérer des caractères, éditer, supprimer • harmoniser, famille d'opérations dont le but est de gommer des différences, éditer, supprimer • généralisation de modèle, opérations visant à rendre les objets similaires, éditer, supprimer • sélection-élimination, opérations visant à diminuer le nb d'objets, éditer, supprimer • sélection, opération qui sélectionne les objet à garder lors d'une opération de type sélection-élimination, éditer, supprimer • élimination, opération qui identifie les objets à éliminer lors d'une opération de sélection-élimination, éditer, supprimer • amplification, exagérer la forme et la taille, éditer, supprimer • lissage-filtrage, opérations visant à atténuer ou éliminer les parties d'un objet, éditer, supprimer • lissage, opération visant l'atténuation des parties d'un objet, éditer, supprimer • filtrage, opération visant l'élimination de parties d'un objet, éditer, supprimer • amplification, opération visant l'exagération de la forme ou de la taille, éditer, supprimer 	
Nom court	MustiereGen										
Nom détaillé	Taxonomie des opérations de généralisation proposée par Seb										
Description	Taxonomie organisant les opérations de généralisation en fonction du but poursuivi et comportant 3 niveaux : - les grandes familles d'opération - les opérations - les actions										
Références bibliographiques	Mustière Sébastien, Apprentissage supervisé pour la généralisation cartographique, Thèse en informatique - intelligence artificielle de l'Université Paris 6, 2001 http://recherche.ign.fr/labos/cogit/pdf/THESES/MUSTIERE/These_Mustiere_2001.pdf										
Contenu	<ul style="list-style-type: none"> • simplifier, Famille d'opérations dont le but est d'éliminer de l'information, éditer, supprimer • caricaturer, famille d'opérations dont le but est d'exagérer des caractères, éditer, supprimer • harmoniser, famille d'opérations dont le but est de gommer des différences, éditer, supprimer • généralisation de modèle, opérations visant à rendre les objets similaires, éditer, supprimer • sélection-élimination, opérations visant à diminuer le nb d'objets, éditer, supprimer • sélection, opération qui sélectionne les objet à garder lors d'une opération de type sélection-élimination, éditer, supprimer • élimination, opération qui identifie les objets à éliminer lors d'une opération de sélection-élimination, éditer, supprimer • amplification, exagérer la forme et la taille, éditer, supprimer • lissage-filtrage, opérations visant à atténuer ou éliminer les parties d'un objet, éditer, supprimer • lissage, opération visant l'atténuation des parties d'un objet, éditer, supprimer • filtrage, opération visant l'élimination de parties d'un objet, éditer, supprimer • amplification, opération visant l'exagération de la forme ou de la taille, éditer, supprimer 										

Figure 59. Publication d'une taxonomie dans le catalogue MFM.

De plus, ce catalogue permet de décrire des algorithmes non implémentés comme illustré sur la Figure 60 ou encore d'intégrer dans la description d'algorithmes des illustrations d'utilisations comme illustré sur la Figure 61.

Catalogue de traitements et modèles

DetectInflexPlaz 1

[\(Editer\)](#) , [Retour à la liste des traitements](#)

Ressource pas d'implémentation java

Mots clés Caractérisation de ligne (AnalyseSpatialeGeom),

Description résumée Détecte sur une ligne les changements de direction principaux (points d'inflexion).
Ne modifie pas la ligne en entrée.

Méthode Le but de cet algorithme est de détecter les changement de direction sur un ligne (points d'inflexion). Pour éviter de détecter toutes les micro-inflexions, un lissage gaussien est d'abord appliqué sur la ligne, et les changements de direction sont détectés sur la ligne lissée. La force du lissage gaussien (paramètre sigma) est un paramètre de l'algorithme de détection. Plus la valeur de ce paramètre est forte, moins on détecte de points d'inflexion.

Auteurs du traitement Corinne Plazanet

Références bibliographiques
Plazanet C. 1996. Enrichissement des bases de données géographiques : Analyse de la Géométrie des objets linéaires pour la généralisation cartographique. Application aux routes. Thèse de doctorat, Université de Marne-la-Vallée, France, 1996.
Plazanet C., Affholder J-G., Fritsch E. 1995. The Importance of Geometric Modelling in Linear Feature Generalization. Cartography and Geographic Information Systems, vol.22, n°4, 1995, pp.291-305

Auteurs de cette description Cécile Duchêne

Documents associés [\(Editer\)](#)

Utilisations, tests [\(Editer\)](#)

Utilisation DetectInflexPlaz : Utilisé dans la thèse de Corinne Plazanet pour segmenter une ligne en portions homogènes du point de vue de la sinuosité. Utilisé pour isoler les virages dans une série de virages, par deux algorithmes de généralisation maintenant utilisés en production: accordéon (issu aussi de la thèse de corinne Plazanet, qui écarte les virages d'une série) et schématisation (qui supprime des virages au sein d'une série). Dans ces cas, la détection est maintenant faite sur une portion de ligne issue d'une

Figure 60. Description d'un algorithme dans le catalogue MFM.

Le choix du prochain bâtiment à supprimer découle d'un calcul relatif à chaque bâtiment, une moy l'encroûtement du bâtiment par rapport aux objets urbains (tronçons routiers et autres bâtiments)

- le nombre de directions dans lequel le bâtiment est encroûté,
- la moyenne de ses encroûtements,
- le nombre d'encroûtements bâti-bâti par rapport aux encroûtements bâti-route,
- la taille relative du bâtiment par rapport au plus grand bâtiment de l'ilot.

Le calcul des encroûtements des bâtiments nécessite au préalable une triangulation de l'ilot, les

- les centroïdes des bâtiments,
- les points des tronçons routiers les plus proches des bâtiments.

Cette triangulation permet de calculer pour chaque bâtiment des segments de proximité renseignés segments de proximité sont ensuite projetés sur des directions d'encroûtements pré-déterminées

Auteurs du traitement J. Renard (d'après J. Gaffuri)

Références bibliographiques:

Gaffuri J., 2008, 'Généralisation automatique pour la prise en compte de thèmes champs: le modèle dirigée par Anne Ruas, co-encadrée par Cécile Duchêne

Ruas A., 1999, 'Modèle de généralisation de données géographiques à base de contraintes et d'au dirigée par Jacques Désarménien

Auteurs de cette illustration J. Renard

Utilisations, tests ([Editer](#))

Test 1 : J. Renard (validation par J. Gaffuri et C. Duchêne) Support du test: plate-forme de générali
Données utilisées: Echantillon couvrant la ville de Bayonne (54)
Illustrations (avant-après):



Figure 61. Description d'un algorithme intégrant des illustrations de tests d'utilisation.

Des jeux tests sont en cours d'introduction pour faciliter la documentation des métadonnées, sur le modèle de ce qui a été proposé dans le travail de thèse de (Abd-el-Kader 2006).

2.2 L'aide à l'accès à des cartes topographiques sur mesure en ligne

Cette dernière sous-partie du mémoire présente l'aide à la conduite d'un processus spécifique : la rédaction d'une carte, et plus précisément la construction de la légende. Des outils de rédaction de cartes et des données sont à la disposition de tout utilisateur du Web. La manipulation de ces outils peut être améliorée en leur associant une expertise cartographique ainsi qu'une aide à la création artistique. Cette section décrit plusieurs propositions menées dans mon action de recherche dans cette direction. Ces propositions s'appuient sur les travaux menés au COGIT sur la couleur (Chesneau 2006)(Chesneau et al 2005)(Buard and Ruas 2007).

Une première partie situe notre approche générale dans le contexte puis nous détaillons des travaux spécifiques.

2.2.1 Contexte

2.2.1.1 Les nouveaux rédacteurs de cartes

L'apparition d'outils cartographiques bureautiques avait permis **d'étendre l'activité de création d'une carte en dehors des agences classiques de cartographie.** Les outils de cartographie sur le Web, bien plus largement accessibles, ont amplifié ce mouvement. La diffusion d'information géographique est maintenant souvent le fait de créateurs de cartes qui n'ont pas acquis les notions élémentaires de cartographie comme l'importance du choix du niveau de détail pour décider de la représentation de l'information et le choix du style de dessin (Gartner and Peterson 2007). Ces créateurs utilisent généralement des interfaces programmatiques. Les premières API cartographiques (comme celles de Google, ViaMichelin, Géoportail) ont toutes offert le même service : permettre à tout auteur de site Web d'inclure dans son site une carte construite à la volée en interrogeant un serveur de données. L'auteur intègre dans son site une carte qui est maintenue à jour par le fournisseur d'information géographique de façon transparente pour lui. **Le critère minimal de spécification de la carte voulue est la zone d'intérêt ainsi que le type de fond de carte (image ou symbolisé).**

Un niveau supplémentaire dans la création de carte sur le web consiste à **ajouter des couches (dites couches utilisateurs)** dans sa carte, qui seront superposées au fond de carte mentionné précédemment. Il s'agit typiquement des mashups réalisés via l'API de Google. Les couches ajoutées au fond de carte sont généralement des couches

vecteurs munies d'un style de dessin. Tout objet peut apparaître sur une carte Web pourvu qu'il soit muni de coordonnées accessibles en ligne. Par exemple, wikinear permet de visualiser sur son téléphone mobile les 5 articles wikipedia les plus proches. Un autre exemple est la possibilité de visualiser sur une carte la localisation en temps réelle de personnes, pourvu que celles-ci aient activé la fonction Google Latitude sur leurs téléphones mobiles.

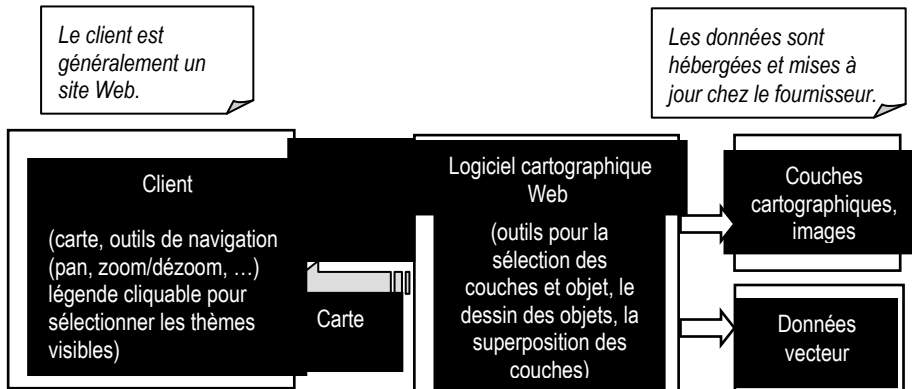


Figure 62. Fonctions et composants d'un serveur de dessin cartographique.

Enfin, un autre niveau dans la création de carte sur le Web est la **personnalisation des styles utilisés pour le dessin**. Elle est proposée par exemple par l'assistant de navigation Tomtom qui a mis en place un site¹ pour créer et échanger des styles. Cette fonctionnalité est également proposée par l'entreprise CloudMade² pour le dessin des données OpenStreetMap. Techniquement, les spécifications OGC permettent de définir des styles utilisateurs, grâce au modèle Styled Layer Descriptor (OGC 2007), et de demander à un serveur de dessin de les interpréter pour dessiner une carte. Des chercheurs proposent d'enrichir ce modèle pour ajouter des styles multivariables (Sykora et al. 2007). Cela a donné naissance au développement à l'ETH de Zürich d'un logiciel serveur Web cartographique, destiné à améliorer les capacités de dessin des serveurs WMS actuels : QGIS MapServer (Hugentobler and Iosifescu 2009). D'autres spécifications OGC s'attachent à permettre le dessin de cartes thématiques efficaces dans des architectures distribuées, Geographic Linkage Services (OGC 2009). Ces spécifications ont fait l'objet d'implémentations, par exemple par la société GeoClip. Pour l'heure, GeoClip est dédié à des professionnels (logiciel payant) mais ce logiciel démontre la possibilité de créer simplement des cartes choroplèthes à partir de données

¹ <http://www.colorschemeeditor.com>

² <http://cloudmade.com/>

statistiques, en utilisant des référentiels géographiques existants. Il nous semble que ces possibilités seront bientôt à la portée du plus grand nombre étant donné la vitalité de la communauté de logiciels opensource géographiques et la disponibilité en ligne de nombreuses données statistiques. Un exemple de fonctions disponibles en ligne pour aider à la rédaction de cartes thématiques est le ColorBrewer (Harrower and Brewer 2004). Il permet au rédacteur d'une carte thématique de spécifier le nombre de classes qu'il souhaite cartographier ainsi que le type de relation entre ces classes. En retour il propose des styles adaptés. L'utilisateur peut tester les styles sur des données du site et également faire apparaître des couches topographiques pour un premier test de lisibilité de la carte.

Cela étant, ces nouveaux créateurs de cartes sont limités dans la mesure où ils ne sont pas amenés à faire tous les choix nécessaires à la conception d'une carte sur mesure, qui répondrait du mieux possible à leur besoin. L'usage du langage cartographique n'est pas entièrement intuitif. Des auteurs comme (Monmonnier 1991) ont souligné les pièges de ce langage. En particulier, l'apprenti cartographe peut croire avoir réalisé une carte efficace car, connaissant le message qu'il veut faire passer sur cette carte, il se concentre sur ce message ce qui biaise sa lecture de la carte. Il ne voit alors pas les messages parasites éventuellement portés par les signes de sa carte. Ainsi le nouveau créateur a besoin d'un apport d'une expertise cartographique. Parallèlement à la multiplication d'outils de cartographie en ligne, l'enseignement de la cartographie n'a pas cru. (Forrest 1999) souligne même que l'enseignement de la cartographie a diminué chez les étudiants en géographie, du moins dans le Royaume Uni.

Si les cartes ont pris une importance quantitative croissante sur le Web, il semble que la cartographie n'a pas encore le niveau qualitatif qu'elle devrait avoir sur le Web. (Ormeling and Kraak 2008) soulignent la nécessité de faire évoluer la théorie cartographique pour que les interfaces cartographiques soient à la hauteur de ce que nous sommes en droit d'attendre d'elles, c'est-à-dire qu'elles exploitent le potentiel de la représentation cartographique. Certains auteurs pointent justement le risque d'évaluer l'efficacité des cartes Web selon des critères traditionnels qui ne sont pas adaptés à ces nouvelles cartes. D'autres critères doivent être proposés tenant compte des nouveaux usages de la carte sur le Web (Cartwright 2008)(Kent 2009). (Nivala et al. 2007) ont conduit une évaluation de cartes Web (MapQuest, Google Maps, Microsoft Maps and Directions) selon plusieurs méthodes :

- une évaluation centrée utilisateurs où des utilisateurs devaient effectuer des tâches prédéfinies selon ces interfaces
- une évaluation par des experts cartographes
- une évaluation par des experts en interaction homme-machine.

Les experts cartographes sont les évaluateurs les plus sévères. Nous retenons que l'évaluation globale conclut à des problèmes d'utilisabilité des cartes causés en particuliers par des mauvais choix de styles. Par exemple, MapQuest utilise une couleur similaire pour les hôpitaux et les zones commerciales, ou encore dans les cartes MSN

les autoroutes sont en bleu ce qui fait penser à l'hydrographie et non au thème routier. Au colloque HyperUrbain2 qui se tenait à Paris en 2009, le chercheur Boris Beaude a fait un panorama très large sur l'usage des cartes dans les sites Web2.0 et a conclu sa présentation par 'il nous faut maintenant de meilleures clés pour nos cartes'. De fait, de nombreuses cartes Web sont spectaculaires mais supportent peu de tâches d'analyse : par exemple les cartes des projets Urban mobs (<http://www.urbanmobs.fr/fr/>) ou Senseable au MIT (<http://senseable.mit.edu/>) qui montrent l'activité des téléphones portables, ou encore les cartes d'émotions produites par Christian Nold, ou encore la carte des messages twitter mentionnant la grippe aviaire, ... la liste est longue.

2.2.1.2 La pertinence des approches centrées utilisateurs pour la conception d'interfaces cartographiques

Un constat qui fonde nos travaux sur la carte sur mesure est que la conception d'une carte efficace demande une phase d'analyse du besoin auquel cette carte veut répondre. Or les architectures actuelles de diffusion de carte comme les API du Web2.0 ou celles de l'ISO/OGC ne supportent pas suffisamment cette expression de besoin. L'importance d'intégrer l'utilisateur dans le cycle de la conception d'un produit a été largement acceptée et renvoie à la théorie 'user centred design'. Nous avons appris et mis en pratique cette théorie dans le cadre de notre participation au projet européen SPIRIT (cf section 1.2.1.3). (Nivala et al. 2007) recommandent une approche centrée utilisateur pour améliorer l'utilisabilité des cartes sur le Web. (Jakobsson 2002)(Stanek et al. 2007) (Tsou and Curran 2008) mettent en œuvre de telles approches dans leurs travaux. (Tsou and Curran 2008) ont une démarche centrée utilisateur très formelle qui reprend une proposition spécifique à la conception d'interfaces Web. Nous pensons que le manque d'expression explicite du besoin auquel la carte doit répondre dans les architectures actuelles encourage la création de représentations cartographiques qui ne répondent pas à des objectifs et qui, de ce fait, n'exploitent pas le plein potentiel du langage cartographique. Durant le processus de conception de la carte, le créateur de la carte se pose actuellement la question des flux de données alimentant la carte et du symbole choisi pour faire apparaître les flux sur la carte. Or il doit d'abord décider si cette carte servira par exemple à s'orienter, à identifier un itinéraire, à comprendre un phénomène ou à communiquer une analyse de l'espace (ex : des zones à risque). Cette analyse du besoin doit précéder et contraindre la définition du bon niveau de détail de la carte, l'acquisition des données thématiques éventuelles, la définition des classes visuelles, l'agrégation des données et leur restructuration, la définition des styles associés et enfin l'évaluation du résultat (Moisuc 2007)(Bucher et al. 2007).

Le rédacteur d'une carte novice a donc besoin d'être aidé de deux façons :

- **dans l'analyse des objectifs de la carte**
- **dans le choix des données et des styles pour atteindre ces objectifs.**

2.2.1.3 Critères d'adaptation d'une carte topographique à un contexte

Dans la suite nous nous concentrons sur les cartes topographiques. Cette section détaille les critères connus pour **l'adaptation d'une carte topographique à un contexte**.

Les styles d'une carte topographique doivent être choisis en fonction des supports attendus pour la carte (écran, projection vidéo, impression papier,...) car les relations entre couleurs sont rarement conservées lors d'un changement de support. Dans leur outil destiné à proposer des gammes de couleurs efficaces pour cartes thématiques, le ColorBrewer, (Harrower and Brewer 2003) précisent si ces couleurs peuvent être correctement vidéo projetées ou imprimées. (Jenny et al. 2008) rappellent que le choix des polices est important pour les cartes écran et qu'il convient d'éviter les polices à empâtement et de favoriser des polices avec des formes simples et ouvertes. Les auteurs ajoutent qu'il est possible d'améliorer la lisibilité d'une carte fortement contrastée sur un écran en ajoutant de l'anti-aliasing¹. Il faut également prévoir les transferts envisagés d'un support à un autre. Une piste prometteuse est la typologie de transformations de cartes de (Moellering 2007). L'auteur propose d'abord une typologie de cartes réelles ou virtuelles construite sur deux critères croisés. Le premier critère est le fait que la carte soit visible ou non. Le deuxième est le fait que la carte soit matérielle ou non. Cela conduit à quatre types de cartes :

- les cartes réelles qui sont matérielles et directement visibles,
- les cartes virtuelles de type I qui sont directement visibles et immatérielle (cartes projetées, certains types de cartes mentales,...)
- les cartes virtuelles de type II qui sont matérielles mais ne sont pas directement visibles (gazetiers papiers, anaglyphes, DVD, ...)
- les cartes virtuelles de type III qui sont immatérielles et non visibles directement (certaines cartes mentales, bases de données, mnt, ...).

Il associe à cette typologie de types de carte une typologie de 16 transformations de carte (qui comprend donc les transferts de support). Selon lui, cette typologie des transferts de support est efficace pour décrire l'évolution d'une carte au cours de ses usages et il l'utiliserait depuis des années avec ses étudiants pour étudier les conservations ou modifications des propriétés d'une carte (couleurs,...) lors de ses usages.

Comme nous l'avons mentionné précédemment, le contexte de la cartographie ubiquitaire a conduit à de nombreux travaux sur l'adaptation des cartes topographiques au contexte. Les éléments de contexte importants pour l'affichage d'une carte

¹ L'anti-aliasing consiste à corriger des dégradations d'une image affichée sur un écran qui se produisent sur les zones à fort contraste. Cette technique introduit une zone intermédiaire moins contrastée avec ses deux voisines et évite l'effet

topographique dans un contexte nomade sont: la localisation de l'utilisateur, son activité, la saison et l'heure de la journée (Jakobsson 2002). Les deux derniers éléments sont importants car ils influencent les conditions physiques de lecture de la carte. Les propriétés de l'utilisateur lui-même sont son âge, sa nationalité, son sexe et sa profession (Jakobsson 2002)(Kurokawa and Ota 2007)(Wakabayashi 2007). Ces propriétés influenceront notamment sur le choix des styles sur une carte.

L'adaptation porte sur la sélection des objets d'intérêt, les couleurs, les pictogrammes et la taille des toponymes. **L'adaptation des couleurs** est également nécessaire, pour renforcer le message de la carte ou pour s'adapter à des lecteurs qui perçoivent différemment les couleurs. 'Some colors are definitely liked more than others, and preferences vary from culture to culture. Color preferences also change with age. For example, young children tend to like pure warm colors, such as high-chroma red, followed by blue, green, and the other spectral hues. [...] Studies of adult color preferences in the United States – which probably reflect Western culture- suggest that blue, green, and red are generally considered 'pleasant.' Blue is liked most, while orange and yellow are rated significantly lower. We appear to like greenish-yellows the least. These preferences also depend on the value and chroma level, with lower-chroma 'pastel' hues preferred over the full-chroma in a few instances.[...]' (Robinson 1995). L'auteur ajoute que les préférences concernent aussi les combinaisons de couleurs et que les américains seraient très sensibles aux contrastes de valeurs. (Edsall 2007) étudie comment la culture des lecteurs attendus de la carte doit être prise en compte lors de la conception de la carte en particulier pour le choix des couleurs. L'auteur souligne également la difficulté méthodologique que cela implique ; d'une part nous ne disposons pas de variable indiquant la culture (la nationalité n'est pas satisfaisante) et d'autre part, dans un monde de globalisation, les cultures évoluent vite. (Kurokawa and Ota 2007) utilisent un graphe de proximité entre des mots et des couleurs pour adapter automatiquement les couleurs d'une carte à des mots clés décrivant un contexte. Les auteurs ont présenté un prototype lors de la conférence ICC de Moscou. La carte affichée change de couleurs au fur et à mesure que le créateur de la carte ajoute des mots clés dans la zone utilisée pour spécifier le contexte attendu de lecture de la carte.

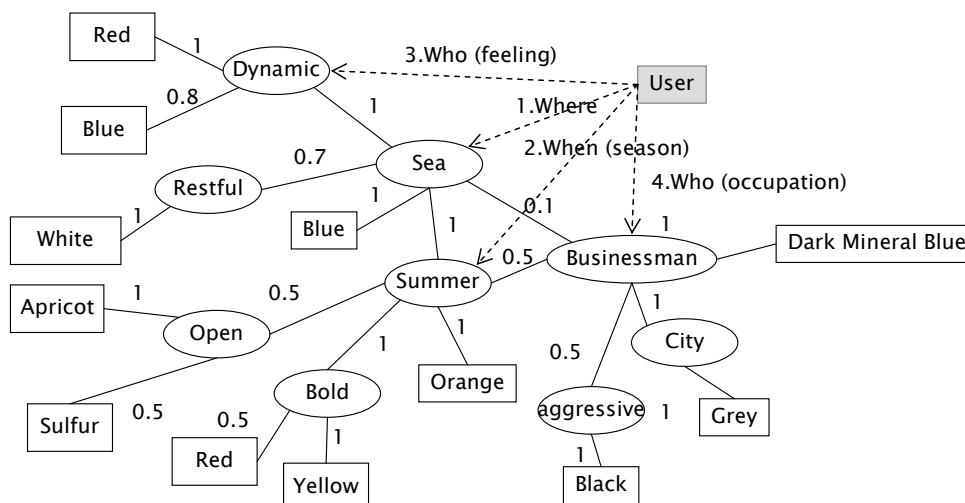


Figure 63. Graphe de proximité entre des concepts et des couleurs proposé par (Kurokawa and Ota 2007) pour adapter les couleurs d'une carte à des mots clé décrivant un contexte.

2.2.2 Modèle général des travaux sur la carte à la carte

Plusieurs travaux de mon action de recherche ont été dédiés à l'aide à la conception de carte sur mesure sur le Web. Globalement ces travaux visent à comprendre le besoin de carte, à proposer une carte en réponse à ce besoin et à évaluer et améliorer cette proposition. Elles s'appuient sur diverses connaissances que nous avons essayé de mutualiser entre ces travaux. Un modèle particulièrement important est le modèle de spécification de carte qui rend explicites les intentions associées à la carte. Plus précisément, tous ces travaux sont partis d'un modèle commun mais ont chacun évolué ensuite indépendamment en s'autorisant à modifier le modèle pour les besoins de l'application étudiée.

2.2.2.1 Modèle SISSI de spécification d'une carte

Le modèle SISSI (du nom de l'action de recherche Servir de l'Information Spatiale Sur Internet) de spécification d'une carte comporte trois niveaux.

Le premier niveau est celui des objectifs de la carte. Il comporte plusieurs propriétés. Une première propriété est la catégorie. Il n'existe pas à l'heure actuelle de modèles de catégorie de cartes. Nous distinguons trois types de catégories :

- le type de tâche (communication, analyse, inventaire, orientation, ...)
- le domaine (risque, navigation, ...)

- le contexte de production.

La catégorie d'une carte peut donner des informations utiles tout au long du processus de rédaction. Par exemple, le domaine d'une carte donne des indications sur d'éventuelles conventions de styles en vigueur dans ce domaine (géologie, risque, ...). Le type de tâche donne des indications sur le type de contenu et le type de lisibilité de ce contenu. Pour le guidage il est essentiel de pouvoir identifier à première vue les grands axes et de lire facilement les identifiants des routes. La spécification des objectifs d'une carte doit aussi comporter des informations sur le contexte d'usage. Le contexte est décrit dans la littérature selon les critères suivantes : les activités des lecteurs, les profils des lecteurs, les conditions externes (saison, météorologie,...) dans le cas de cartes sur terminaux nomades et le support de la carte. Les profils des lecteurs doivent être étudiés selon les critères suivants : 'handicaps', âge, nationalité, sexe, culture. Nous ajoutons dans les objectifs d'une carte la notion d'impact émotionnel recherché. L'importance de ce facteur sera présentée ultérieurement. La section suivante détaillera notre proposition pour structurer cet élément dans notre modèle.

Le deuxième niveau est celui du contenu en information et du type de lisibilité voulu pour ce contenu. La théorie cartographique nous apprend que la représentation cartographique peut donner à voir ou à lire des regroupements d'objets –thèmes- et des différences ou similarités entre groupes d'objets –les relations-. Nous reprenons pour ce niveau le modèle de (Chesneau 2006) inspiré des cartographes comme Jacques Bertin (Bertin 1983). Nous introduisons par ailleurs la notion de niveau de lecture. Trois niveaux de lecture spécifiques sont utilisés pour le moment :

- Le niveau 'première vue' qui regroupe les informations qui doivent être 'vues immédiatement'. Ce niveau est particulièrement important dans les cartes de communication et dans les cartes d'aide à la prise de décision en contexte de gestion de crise ou de guidage de navigation.
- Le niveau 'lecture1' regroupe les informations qui doivent être lisibles. La représentation graphique des objets correspondant peut comporter des détails plus fins que celle des objets du niveau 'première vue'.
- Le niveau 'fond' correspond au fond de carte. Il regroupe des informations qui doivent être lisibles mais qui ont une moindre importance que le niveau 'lecture1'.

L'expression du niveau de détail ne doit pas se limiter aux échelles de référence. Il est important de spécifier pour les domaines représentés les niveaux de détails attendus. Il semble que ces informations pourront par exemple être traduites également en contraintes de préservations utiles dans un processus de généralisation de données.

Le dernier niveau est celui de la spécification du processus de dessin, que nous appelons l'implémentation. Il correspond aux FeatureStyle de l'OGC qui peuvent être pourvus d'annotation sémantique portées par le FeatureType concerné. Ce modèle est proche de la LigneDeLegende de (Chesneau 2006) mais permet de décrire dans un

même style la couche des points d'intérêt dans laquelle les pictogrammes varient en fonction de la nature du point d'intérêt (édifice religieux, terrain de sport, etc.).

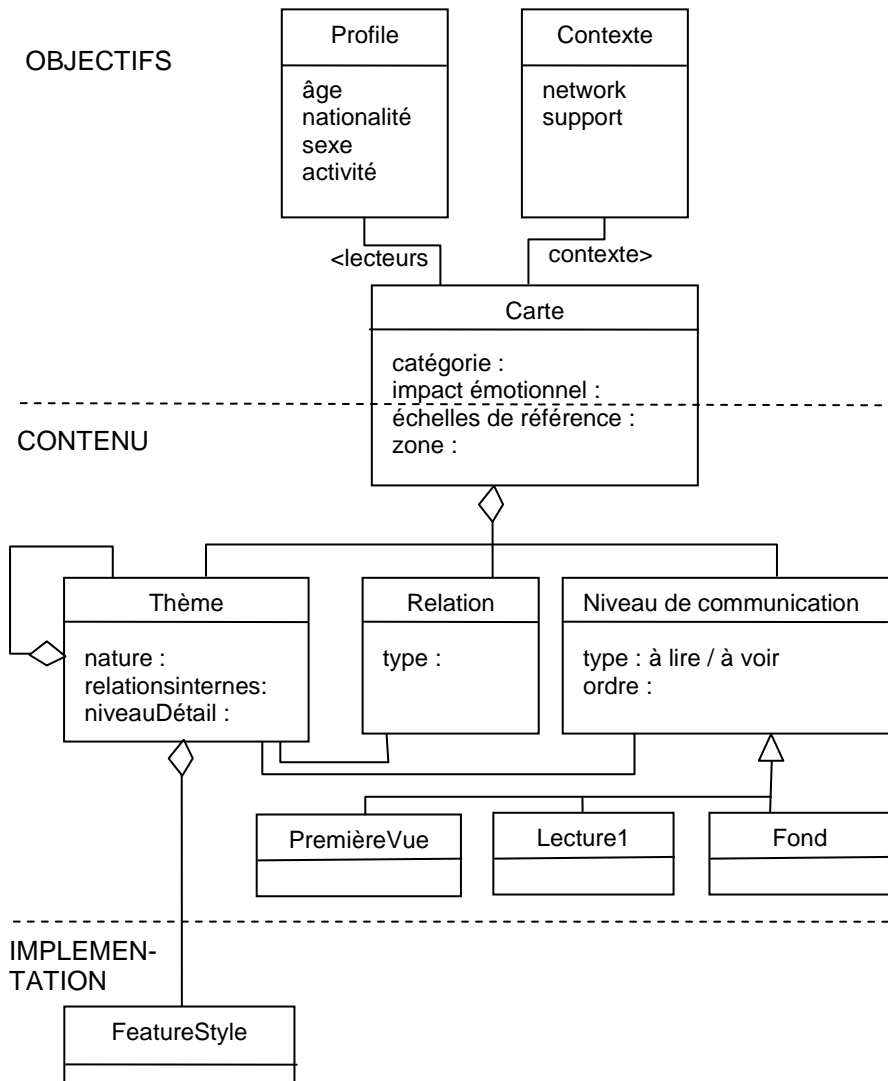


Figure 64. Modèle de spécification d'une carte.

2.2.2.2 La prise en compte de l'impact émotionnel

Cette section explique l'importance de l'impact émotionnel d'une carte, propriété qui a été introduite dans le modèle SSSI de spécification d'une carte puis décrit comment nous proposons de le prendre en compte dans le modèle de spécification.

Importance de l'impact émotionnel d'une carte

Un défi des SDIs du futur est d'être plus attractives pour les utilisateurs. Or les émotions suscitées par un produit ont une part importante dans son attractivité. Cela participe au succès des vues satellites ou photographiques et plus encore des mondes virtuels comme Google Earth. Ceux-ci sont très ancrés dans l'expérience sensible que le lecteur fait couramment du monde et lui offrent un contraste de point de vue en lui permettant de voler comme un oiseau à la surface de la terre. Google Earth procure à la fois de la surprise –surprise d'expérimenter une nouvelle relation à un espace géographique familier- et un sentiment de pouvoir –avoir la capacité de se déplacer librement dans cet espace-. L'usage des mondes virtuels pour faire prendre conscience des dégâts causés par l'homme sur la planète est également plus efficace auprès de la plupart des personnes que des cartes de synthèse.

La surprise, la joie, la peur font partie des huit émotions primaires du modèle de Robert Plutchik (joie, confiance, peur, surprise, tristesse, dégoût, colère, attente). La fierté (sentiment de pouvoir) est elle-même une émotion secondaire dans le modèle de Parrot ou encore une émotion complexe dans le modèle de Lojban. Les travaux en psychologie ont mis en évidence l'importance des émotions dans la cognition et dans l'adaptation de l'homme à son environnement. Une émotion est un changement dans l'état physiologique d'une personne en réaction à un processus de perception. Charles Darwin a décrit le rôle d'émotions comme la peur ou la colère dans la communication entre les êtres vivants et leur adaptation à leur environnement. De façon simplifiée, les émotions mettent en oeuvre une région spécifique du cerveau ce qui permet d'accélérer certains mécanismes (réaction spontanée, mémorisation facilitée) et de rendre les autres capacités cognitives plus disponibles à la prise de décision (théorie d'Herbert Simon). Les émotions sont maintenant intégrées à l'informatique, par exemple avec les travaux de (Cañamero 2001) qui visent à doter les programmes (essentiellement dans les robots) de mécanismes reproduisant les émotions.

Des pistes existent pour renforcer l'impact émotionnel des cartes. L'usage de vues cavalières dans les systèmes de navigation favorise le rapprochement entre la carte et la réalité. La superposition d'une couche image avec la possibilité de passer d'une couche image à une couche stylisée comme sur le géoportail permet à l'utilisateur de passer en continu d'une vue réaliste de la terre à une carte symbolique. Enfin, avec Street View,

Google propose un point de vue réaliste à l'intérieur de la carte. Nous nous intéressons à la possibilité de susciter des émotions, et essentiellement le plaisir esthétique, liées aux styles mêmes de la carte.

Dans le domaine des cartes, l'émotion privilégiée est le plaisir esthétique. (Tufte 2006) introduit son livre sur le graphisme de la façon suivante : "A colleague of Galileo, Federico Cesi wrote that Galileo's 38 hand drawn images of sunspots 'delight both by the wonder of the spectacle and the accuracy of expression.' That is beautiful evidence." **La conception d'une légende, comme de nombreux processus de conception, est à la fois une science et un art dans la mesure où elle fait appel à la science cartographique mais également aux capacités de créativité artistiques de son auteur** pour stimuler des émotions (McMaster and Shea 1992) (Krygier 1995). (Cartwright et al 2008) ont introduit le concept de 'affective map' pour souligner l'importance du facteur 'affect' dans la lecture d'une carte. (Kent 2009) se concentre sur un affect particulier qui est le sentiment d'être à un endroit particulier véhiculé par la carte. Selon lui, les cartes Web ont perdu en partie l'impact des cartes produites par les agences cartographiques parce que les sites de web mapping sont trop 'égalitaires' alors que les cartographes traditionnels créaient une représentation véhiculant le sentiment d'être à un endroit particulier. En définitive, **la part de créativité artistique du cartographe l'aidait à composer des éléments graphiques pour stimuler les sens du lecteur.** (Cartwright et al 2008) avancent que la cartographie s'est éloignée de l'art pour obtenir une meilleure légitimité scientifique avec une nouvelle discipline : la géovisualisation. Ils soulignent l'importance de rapprocher à nouveau cartographie et créativité artistique. Cela les a conduit à créer un nouveau groupe de travail dans l'Association International de Cartographie : « Art and Cartography ».

Prise en compte de l'impact émotionnel d'une carte dans le modèle SISSI

Pour structurer la propriété 'impact émotionnel' d'une carte, Catherine Domingues de l'équipe SISSI au COGIT a étudié l'existence de descripteurs objectifs qui rendraient compte de l'émotion suscitée par une carte de façon plus ou moins universelle et à l'instar de ce qui existe pour les couleurs.

Nous avons conduit **une expérience pour étudier les descripteurs employés pour désigner le ressenti face à une carte.** Nous avons présenté diverses cartes aux membres de laboratoire COGIT dont certaines produites pour l'expérience par nous-mêmes pour nous éloigner des cartes traditionnelles. Pendant cette expérience, nous avons recueilli les réactions spontanées des personnes, exprimées librement et collaborativement. Nous avons pu remarquer que les personnes de formation cartographe liaient fortement beauté et lisibilité de la carte. Si une carte était peu lisible elle ne pouvait être jugée belle –ou plutôt on ne pouvait dire que c'était une belle carte-. La carte est avant tout une carte avant d'être une œuvre graphique. Catherine Domingues a appliqué des méthodes d'analyse linguistique sur les expressions

recueillies. Elle a regroupé les adjectifs utilisés en groupe de termes proches et a choisi un représentant pour chaque groupe. **Elle a obtenu 7 termes : chaud, pastel, lumineux, sobre, riche, original, réaliste.**

Suite à cela, le travail de stage de mastère CarthaGéo de Pierre Philippe, encadré par Catherine Domingues a visé à **construire des exemples de cartes qui aient des valeurs différentes pour ces propriétés** (Philippe 2006) (Domingues and Bucher 2006). Ces cartes devaient être construites à partir de mêmes données mais en variant les couleurs de la légende. Pour ce faire, il a repris diverses gammes colorées provenant d'un **ouvrage sur l'harmonie des couleurs (Sawahata 2001)**. Pour chaque gamme colorée, il a construit deux légendes en cherchant à utiliser au mieux les couleurs de la gamme en fonction des critères cartographiques. Les légendes obtenues étaient de qualités cartographiques diverses car certaines gammes ne comportaient pas les couleurs conventionnelles nécessaires à certains thèmes de la carte. Chaque légende a alors été appliquée au dessin de données pour obtenir des cartes de mêmes tailles mais correspondant à deux échelles différentes. Pour renseigner la valeur de chaque propriété sur chaque carte, les cartes ont été évaluées par des membres du laboratoire en regard de la liste des propriétés. Nous avons ensuite considéré la valeur moyenne de chaque propriété dans tous les tests comme valeur de la propriété pour l'échantillon en question (voir Figure 65 et Figure 66).

Les résultats ont montré tout d'abord que, pour chaque testeur, **son goût était corrélé aux valeurs des propriétés**. Cela semble indiquer que ces propriétés évoluent (très globalement) de façon cohérente avec la satisfaction esthétique ressentie par le lecteur de la carte.

Ces résultats ont aussi montré que, **sur l'ensemble des échantillons qualifiés, certaines propriétés étaient corrélées comme les propriétés riche et sobre, originale et réaliste**. Nous avons aussi constaté que dans certains cas il n'y avait pas de consensus sur la valeur des propriétés. Nous en avons conclu que **les propriétés pertinentes pour qualifier une carte dépendaient de cette carte**. Selon la carte, il n'est pas toujours possible d'utiliser par exemple le terme 'sobre' (accompagné d'un modificateur comme 'très, plutôt, plutôt pas, pas du tout') pour qualifier la carte car les personnes ne s'accordent pas suffisamment sur cette valeur. Cela étant, lorsque la valeur moyenne est extrême (proche de 1 ou proche de 5) la propriété semble pertinente pour qualifier la carte dans la mesure où il y a un consensus sur sa valeur. Par ailleurs, les propriétés n'ont pas toutes le même comportement. Nous avons constaté que les propriétés 'pastel' et 'riche' étaient globalement moins consensuelles que les propriétés 'originales' et 'chaude'.



Chaud : 2,91
Pastel : 1,45
Lumineux : 3
Sobre : 2,18
Riche : 3,36
Originale : 2,18
Réaliste : 3.82



Chaud : 2,73
Pastel : 3,18
Lumineux : 3,45
Sobre : 2,55
Riche : 3,36
Originale : 2,73
Réaliste : 3.64



Chaud : 2,64
Pastel : 2,55
Lumineux : 2,73
Sobre : 2,55
Riche : 2,09
Originale : 3,73
Réaliste : 2,73

Figure 65. Exemples d'échantillons qualifiés. Les valeurs ont été calculées comme des moyennes ce qui explique la présence de valeurs non entières.



Chaud : 3
Pastel : 3,09
Lumineux : 3,18
Sobre : 1,91
Riche : 2,27
Originale : 4,64
Réaliste : 1,09



Chaud : 2,55
Pastel : 3,82
Lumineux : 3,82
Sobre : 2,18
Riche : 2
Originale : 3,91
Réaliste : 1,82



Chaud : 2,64
Pastel : 2,64
Lumineux : 2,27
Sobre : 3,18
Riche : 1,91
Originale : 3,91
Réaliste : 1,73

Figure 66. Exemples d'échantillons qualifiés. Les valeurs ont été calculées comme des moyennes ce qui explique la présence de valeurs non entières.

Enfin, un aspect particulièrement intéressant de ce travail est **qu'il est possible de retrouver sur certains échantillons les qualificatifs associés aux gammes de**

couleurs initiales. La Figure 67 montre que les mots clés associés aux gammes peuvent être repris pour les cartes.

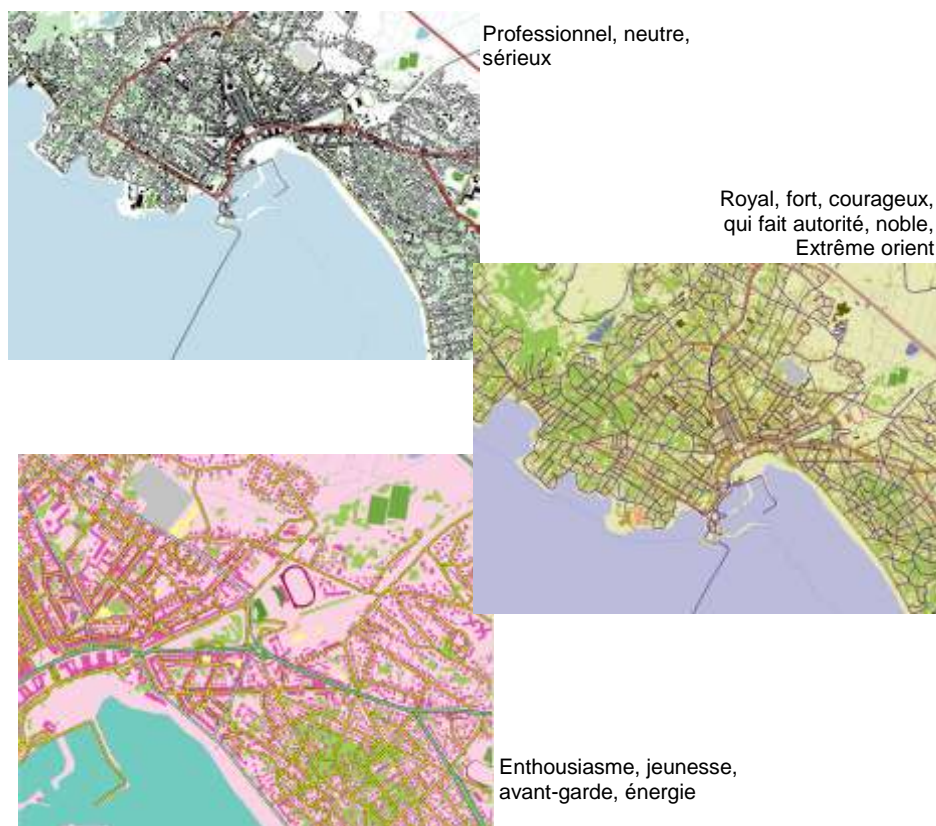


Figure 68. Cartes obtenues en utilisant des gammes de couleurs du livre sur l'harmonie des couleurs de (Sawatha 2001) et mots clés associés à ces gammes.

Cette expérience appelle plusieurs suites.

L'analyse statistique doit être améliorée pour identifier des propriétés consensuelles –pour une carte donnée-, des propriétés discriminantes, et enfin des propriétés subjectives et les profils de lecteurs associés. Cela demande d'abord d'augmenter le corpus d'échantillons. Des travaux ont été faits dans ce sens, en particulier le travail de stage de Jérémy Renard encadré par Laurence Jolivet (Renard 2008) pour construire des légendes prédéfinies à partir de cartes topographiques européennes. Cela demande aussi d'améliorer le protocole de test. Laurence Jolivet et Jérémy Renard ont ainsi modifié les qualificateurs utilisés et ont proposé des paires de qualificateurs opposés ce qui permet de réduire l'ambiguïté des termes (Renard 2008). Cela demande aussi d'améliorer la qualité cartographique des échantillons. Les testeurs ayant une forte culture cartographique semblent avoir du mal à s'exprimer sur des cartes de mauvaise

qualité cartographique de façon spontanée. Nous avons pu le constater lors d'un autre test conduit au laboratoire par Sidonie Christophe et Catherine Domingues. Il serait opportun par exemple de cerner les routes dans certaines cartes pour améliorer les contrastes et permettre au testeur de se concentrer sur ses autres émotions. Enfin, le recours à l'analyse par composantes principales tel que l'a expérimenté Laurence Jolivet sur les échantillons produits lors de stage de Jérémy Renard est une technique très pertinente pour comprendre les corrélations entre les propriétés et les propriétés discriminantes d'une carte à une autre.

Il semble également important **d'étudier le lien entre des propriétés concrètes de la carte (quantités et distribution de couleurs, relations entre objets colorés) qui conduisent aux valeurs obtenues pour les propriétés abstraites.** Chronologiquement, cela pourrait suivre l'amélioration de la base d'échantillons qualifiés.

Enfin, il est nécessaire de **mieux comprendre le vocabulaire utilisé pour parler d'une carte**, la formation des termes par son lecteur ou par une personne qui cherche une carte –qui parle de la carte qu'il voudrait avoir-. Il s'agit là d'un travail de linguiste qui a été entrepris par Catherine Domingues. Elle a entrepris une démarche plus globale de construction d'un corpus de textes sur les cartes qui permettent, via des techniques d'analyse linguistique, de mieux comprendre la perception et l'usage des cartes.

2.2.3 Des services cartographiques additionnels dans les architectures ISO/OGC de Web mapping

Les travaux de Laurence Jolivet ont porté sur l'amélioration des interfaces programmatiques de cartographie sur le Web. Sa proposition peut être vue plus largement comme un ensemble d'APIs que l'IGN pourrait mettre à la disposition de créateurs de sites cartographiques.

2.2.3.1 Approche méthodologique

D'un point de vue méthodologique, nous avons d'abord proposé un modèle logique du processus de conception d'une carte sur mesure (Bucher et al. 2007). Nous avons ensuite comparé les exigences fonctionnelles de ce processus et les fonctions supportées par les services actuels. Cela nous a permis de proposer des services pour combler les lacunes fonctionnelles.

Un premier volet de cette proposition est donc un modèle logique du processus de conception d'une carte sur mesure, dans le contexte moderne des architectures distribuées de données, couches stylisées et services de traitements. Un premier modèle calqué des modèles classiques de conception de carte sur mesure a été révisé après une première implémentation, dans la suite je présente le modèle issu de cette première itération. Une première étape est l'analyse du besoin c'est-à-dire l'expression des usages attendus de la carte et leur traduction en des éléments clés comme la zone, l'échelle, les domaines de la réalité. Il faut ensuite confronter ce besoin aux capacités de dessin cartographiques existantes. Cela se faisait classiquement en deux étapes : acquisition des données puis restructuration des données (pour produire des classes graphiques conformément aux modalités de dessin des outils de cartographie) et définition des styles. Cette décomposition en deux étapes doit être revue. Actuellement, une carte se construit souvent en combinant des couches stylisées. De plus, il est opportun de tenir compte des possibilités d'obtenir une nouvelle couche stylisée à partir d'une couche existante en faisant varier les données (si plusieurs représentations sont disponibles) ou les styles. (Brewer and Buttenfield 2007) ont ainsi construit une table de décision qui supporte ce genre d'approche. Nous proposons dans notre modèle de grouper dans une même étape l'acquisition des données et celle des couches stylisées et de tenir compte des couches stylisées potentiellement dérivables en modifiant les données par un traitement tout automatique ou en modifiant le style. Cette proposition s'inspire donc grandement de celle de (Brewer and Buttenfield 2007). De la sorte, le dessin se fera en exploitant au mieux les ressources dont on dispose que ce soit des cartes existantes ou des données, ou des données pourvues de styles, ou des légendes prédéfinies correspondant à un certain modèle générique de données. Selon le cas, la conception d'un produit sur mesure nécessitera une dérivation de données ou une définition d'un nouveau style. Suite à cette deuxième étape, le rédacteur a fait un choix de couches cartographiques à acquérir, de données à acquérir, à transformer éventuellement et de styles pour ces données. Il peut alors acquérir ces ressources et dessiner sa carte. Cela peut nécessiter d'intégrer les données ainsi que les couches cartographiques. Concernant ce deuxième point, il faut comprendre qu'on ne peut se contenter de reprendre les légendes de deux cartes pour obtenir une nouvelle carte. D'une part les données peuvent être incohérentes et nécessiter un recalage par exemple. D'autre part de nouveaux problèmes de lisibilité peuvent surgir et nécessiter une généralisation inter-thèmes par exemple. Enfin, les styles initiaux devront être repensés en fonction des objets géographiques apparaissant sur la carte superposée. Si deux thèmes provenant de sources différentes ont des styles initiaux aux teintes proches, le lecteur de la superposition y verra une proximité de nature entre les thèmes. Le rédacteur doit également ensuite positionner les toponymes. Il lui faut enfin évaluer la carte. Même si la légende est correcte des problèmes de lisibilité peuvent surgir au niveau des données. Par exemple, un objet plongé dans un aplat de grande taille peut voir sa perception modifiée, il peut ne plus être lisible du tout. Il peut aussi être lu différemment si sa couleur est perçue comme différente de la couleur apparaissant dans la légende. C'est

particulièrement sensible dans le cas où une légende propose une relation d'ordre longue –qui correspondra donc à des symboles de même teinte mais d'intensités ou de tailles différentes-. Sur la carte finale, certaines entités auront pu glisser d'un niveau de cet ordre à un autre. L'évaluation consiste à détecter ces problèmes ainsi que d'autres problèmes comme le fait qu'une maison stylisée peut empiéter sur une route stylisée alors que les objets sont disjoints dans la réalité. Selon le problème, la réparation passera par une modification des données ou des styles.

Nous avons comparé les exigences fonctionnelles d'un processus de cartographie précédemment listées aux fonctions supportées par des services existants. La première étape n'est tout simplement pas supportée par des APIs. La deuxième étape est partiellement supportée. Les catalogues de serveurs de données et de serveurs de couches cartographiques (serveurs WCS) permettent à l'utilisateur de trouver les données dont il a besoin pour peu que le besoin soit exprimé dans les termes de requête de ces catalogues. L'acquisition des données et des couches est bien évidemment supportée ensuite par les serveurs de couches cartographiques (serveurs WMS) et d'objets (serveurs WFS). Des travaux existent sur les services de transformation des données, il s'agit en effet ici de la transformation de schéma pour créer des classes graphiques. Des services d'intégration de schémas et de données voient également le jour. Par contre, il n'y a pas de service d'intégration de couches cartographiques qui modifie les styles des couches initiales pour obtenir une superposition efficace¹. La proposition de styles est supportée par les serveurs WMS qui permettent à l'utilisateur de sélectionner des styles parmi les styles prédéfinis. La proposition de styles est également proposée par l'API geotools construite sur le ColorBrewer. Le dessin de la carte est pour l'instant géré sur un client mais sera géré à terme sur des serveurs dédiés (Web Portrayal Services). Des travaux existent sur les services de placement de toponymes. Enfin il existe également des travaux sur les services de généralisation.

¹ Ce problème fait l'objet du travail de thèse de Firas Bessadok qui débute au COGIT en automne 2009, encadré par Sébastien Mustière et dirigé par Anne Ruas

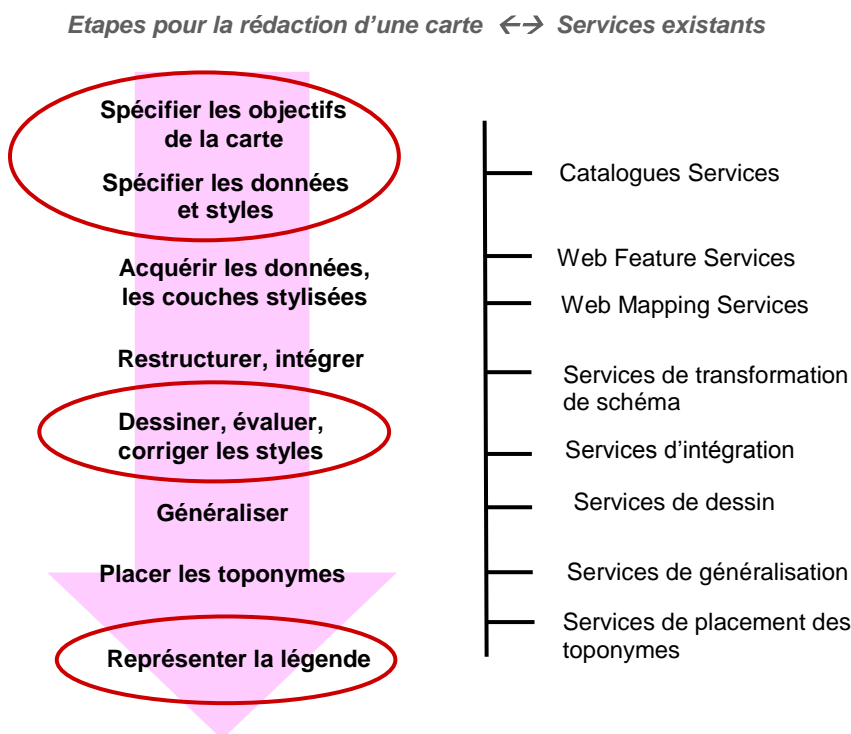


Figure 69. Etapes pour la rédaction d'une carte et services 'existants' de façon opérationnelle ou dans la littérature scientifique. Nous avons cerclé les étapes qui n'étaient pas supportées suffisamment par de tels services.

Ainsi, des services manquants sont ceux qui permettent de passer d'une expression quelconque du besoin en une spécification de carte à dessiner, ceux qui permettent d'intégrer des couches cartographiques en modifiant les styles pour qu'ils deviennent compatibles et enfin les services d'évaluation qui détectent les problèmes de lisibilité et recommandent des réparations (modification des styles ou des données).

2.2.3.2 Services proposés

Notre proposition repose sur les trois niveaux de spécification d'une carte précédemment présentés (section 2.2.2.1). Le modèle SISSI de spécification d'une carte possède un niveau très proche des usages et un niveau très proche de la carte dessinée. **Le processus vise à passer du niveau proche des usages au niveau le plus 'bas' où les spécifications sont interprétables automatiquement en un dessin grâce à des serveurs de dessin (comme MapServer ou GeoServer).**

Notre proposition comporte trois services:

- Un premier service d'aide à la spécification de contenu.
- Un service d'aide à la spécification de couches stylisées en réponse à une spécification d'objectifs et de contenu.
- Un service d'évaluation qui distingue :
 - o l'efficacité cartographique de la légende (c'est-à-dire la cohérence entre les styles proposés et les spécifications de contenu de la carte)
 - o l'efficacité cartographique de la carte dessinée (qui prend en compte les effets de contrastes simultanés et de déplacement des couleurs dans la carte du fait des données).

Le service d'aide à la spécification de contenu vise à aider un créateur à traduire des objectifs en contenu, c'est-à-dire à construire le 2^e niveau de spécification à partir du niveau le plus haut. Il ne s'agit pas d'une étape incontournable d'un processus de cartographie. Si l'utilisateur demande une carte de navigation routière, il est inutile d'analyser le contenu nécessaire de sa carte car une carte –ou une légende– répondant à son besoin existe déjà et est qualifiée de 'carte d'aide à la navigation routière'. Cependant ce service est utile dans deux cas :

- Il est utile pour aider l'utilisateur à trouver des données et des styles dans le cas où il n'existe pas de cartes ou légende prédéfinie. En effet, le service suivant pourra alors proposer des nouveaux styles mais il aura besoin des spécifications de contenu pour faire ces propositions.
- L'évaluation d'une carte doit prendre en compte les relations attendues entre les objets dessinés sur la carte (Chesneau 2006) (Buard et Ruas 2007). Ces informations sont spécifiées à cette étape.

Le service d'aide à la spécification de couches stylisées vise à aider un créateur de carte en définissant des couches stylisées pour sa carte (c'est-à-dire à préciser le niveau 'implémentation' en fonction des spécifications d'objectifs ou des spécifications de contenu de la carte). A ce stade, nous considérons que l'utilisateur ne se soucie pas de savoir si la carte existe déjà –éventuellement parce qu'un autre utilisateur a exprimé le même besoin avant lui- ou si elle est calculée en réponse à sa requête et de quelle façon elle est calculée. Globalement nous voulons donc l'aider à confronter son besoin aux possibilités existantes de dessin cartographiques (données, styles, composants des styles). Nos travaux se sont concentrés par la suite sur ce deuxième service.

Le service d'évaluation des légendes et des cartes repose sur les travaux de l'action de recherche en sémiologie. Laurence Jolivet a étudié la mise en place d'une interface programmatique modulaire au-dessus de l'application d'amélioration des valeurs de contraste développée par Elodie Buard dans la plate-forme Geoxygene à partir de la proposition initiale d'Elisabeth Chesneau (Chesneau 2006). Les travaux sur l'évaluation sont donc concentrés dans l'action de recherche en sémiologie.

2.2.3.3 Un répertoire de styles et de couches stylisées pour l'aide à la création de carte

Un prototype a été implémenté pour le deuxième service. Ce prototype comporte un **catalogue de couches stylisées et de styles prédéfinis**. Le catalogue devrait comporter également à terme des opérations permettant de dériver une nouvelle couche stylisée à partir de couches existantes. Son modèle est repris sur la Figure 70. Un style correspond à la spécification de « Comment dessiner les objets d'un certain type », il est donc spécifique à des éléments d'un schéma de données. Une couche stylisée est obtenue en appliquant un style à des données. Il s'agit de notions standard OGC (Styled Layer Document d'(OGC 2007)). Plus précisément les ressources principales du catalogue sont des combinaisons de styles et des combinaisons de couches stylisées. Cela correspond aux notions de légendes et de cartes. Comme nous l'avons mentionné en section 2.2.3.1, il est important de comprendre qu'un groupe de styles ou un groupe de couches stylisées n'aura pas pour propriété l'agrégation des propriétés de ses composants. C'est pourquoi nous introduisons explicitement dans le catalogue ces groupes : pour pouvoir documenter correctement leurs propriétés.

Les descripteurs des groupes correspondent aux critères pertinents de requête du catalogue c'est-à-dire aux éléments du modèle de spécification d'une carte (objectifs et contenus). **Dans la mesure du possible nous essayons de définir des groupes de styles (des légendes) et de les qualifier. Généralement les propriétés d'un groupe de couches stylisées (une carte) se déduisent des propriétés du groupe de styles correspondant. Mais ce n'est pas systématique.** Par exemple si un Groupe de Styles donne une importance visuelle particulière à la mer (c'est-à-dire possède un niveau de lecture du type 'premièrevue' avec le thème 'mer' dedans), il est possible que ce niveau de lecture ne figure pas dans les propriétés d'une carte correspondante si celle-ci correspond à une zone sur laquelle il n'y a pas de mer. Le descripteur 'type de zone' a été introduit pour préciser le type de zone pour lequel une légende était adaptée. En effet, il existe des types de zones qui influencent la lisibilité des contenus. En région montagneuse par exemple, les courbes de niveau sont nombreuses et resserrées, les routes sont sinueuses. En région urbaine dense les bâtiments et les rues sont extrêmement denses. Dans la construction des premiers éléments du catalogue, Laurence Jolivet s'est concentrée sur les types de zone suivants : montagneux, plaine-rural, plaine-urbain.

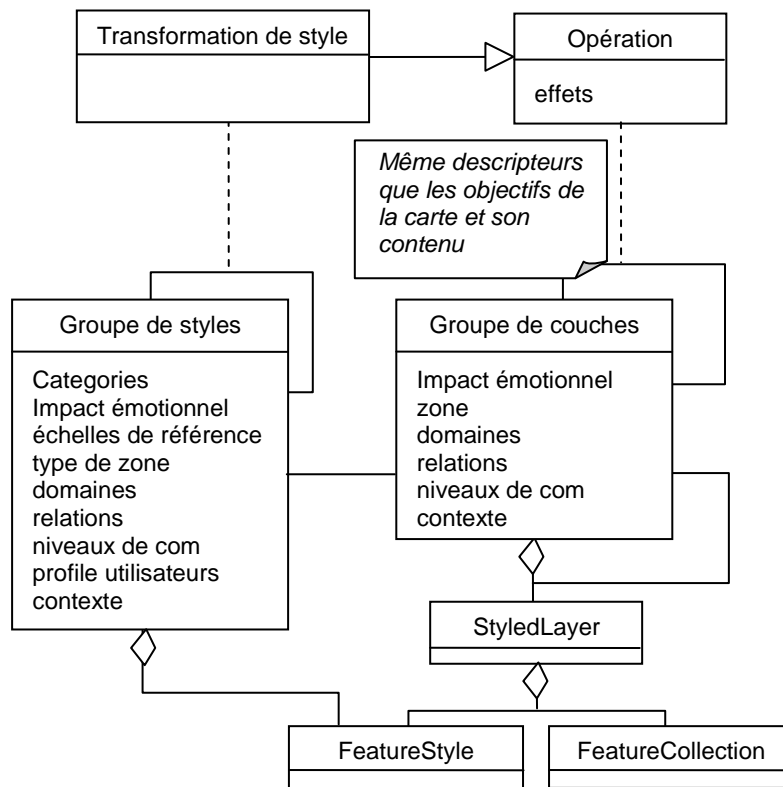


Figure 70. Modèle de notre catalogue de styles et de couches stylisées.

Les premiers éléments du catalogue ont été acquis pendant le stage de Jérémie Renard en extrapolant des légendes topographiques européennes au modèle de données topographiques IGN (simplifié). La Figure 72 montre des cartes obtenues avec les légendes Danoise et Estonienne. Le processus est schématisé sur la Figure 71. L'extrapolation des styles s'est concentrée sur la couleur et les éléments de légende étudiés par (Kent 2007) n'ont pas été pris en compte ici.

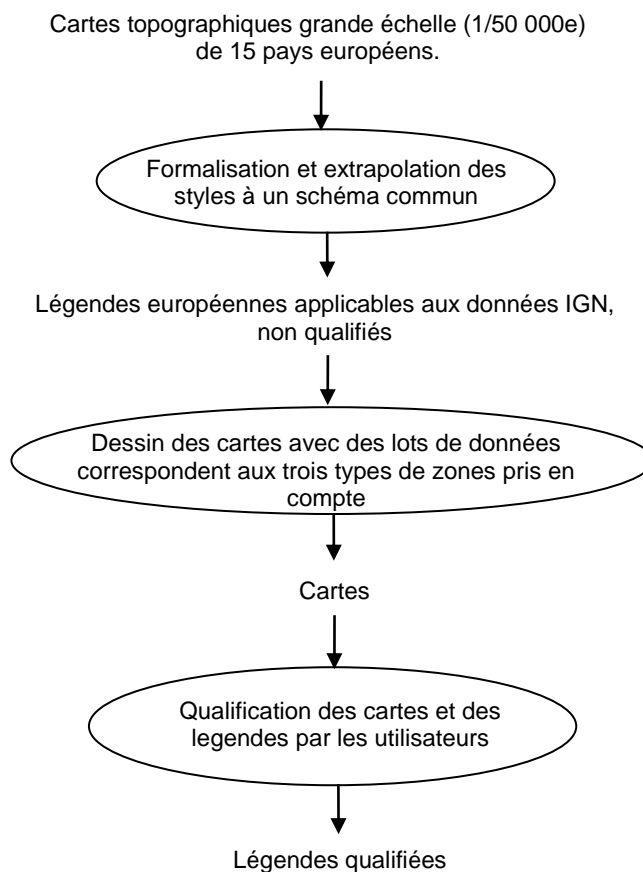


Figure 71. Processus d'acquisition d'éléments particuliers pour le catalogue à partir de cartes existantes : des groupes de styles applicables aux données IGN et qualifiés.

Les tests utilisateurs ont consisté à montrer à des membres du laboratoire les cartes obtenues en leur posant des questions relatives aux descripteurs de notre modèle (Renard 2008).

Nous retenons que **les descripteurs prennent des valeurs très différentes d'une légende à une autre**. Ainsi la notion de carte topographique ne renvoie pas à une légende standard mais à des choix dans chaque pays pour rendre compte de la réalité topographique du terrain.

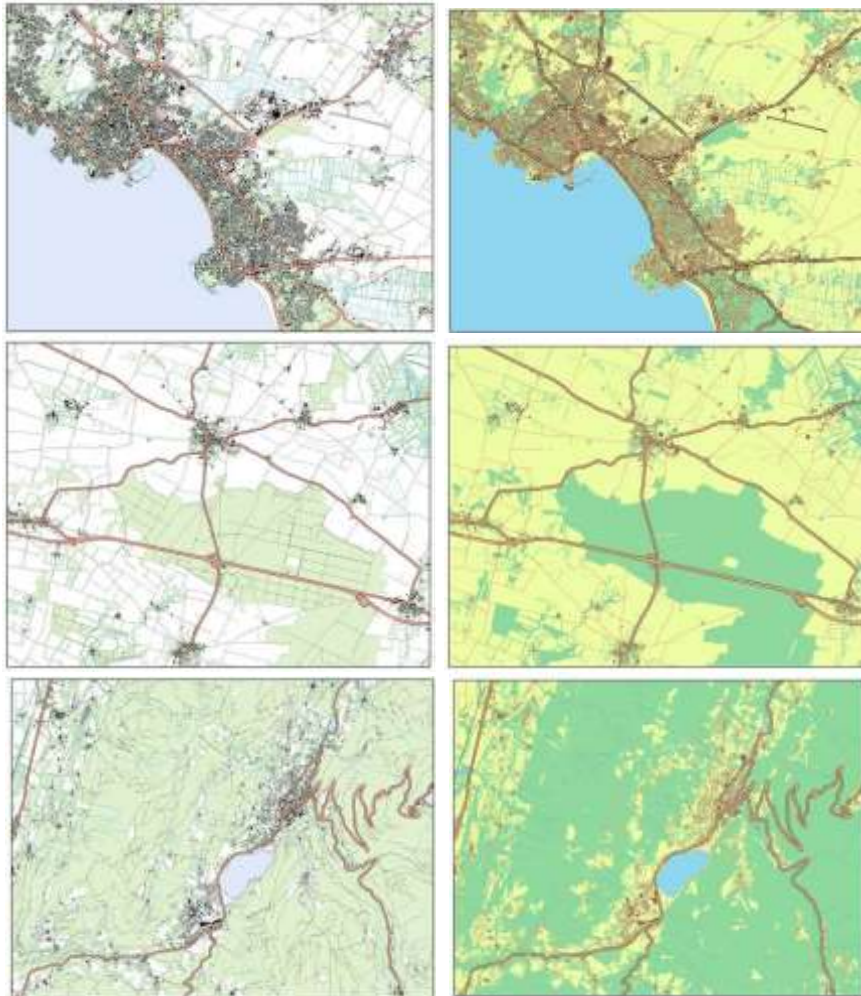


Figure 72. Cartes obtenues avec la légende Danoise (gauche) et avec la légende Estonienne (droite). La légende Danoise a été qualifiée de classique et précise alors que la légende Estonienne a été qualifiée d'originale et brouillonne.

D'une légende à l'autre, **les niveaux de lecture différent** comme le montre la Figure 73.



Figure 73. Exemple de différences de niveaux de lecture induites par un changement de légende sur de mêmes données topographiques.

A la suite de ce travail, Laurence Jolivet a implémenté un service de requête du catalogue. Plusieurs fonctionnalités sont proposées :

- requêter des groupes de styles en utilisant les descripteurs, par exemple 'une légende topographique qui mette en valeur le réseau routier',
- consulter les descripteurs d'une sélection de groupes de styles,
- requêter les groupes de styles en tenant compte d'une couche utilisateur,
- demander des propositions de couleurs pour une nouvelle couche.

La troisième fonction se fait en considérant que la couche utilisateur est un thème différent de ceux du fond de carte et en proposant les groupes de styles qui conduisent à la meilleure note de contraste de la légende lorsqu'on les combine avec le style de la couche utilisateur. Cette note de contraste est fournie par le module d'évaluation des contrastes dans les cartes conçu par (Buard et Ruas 2007).

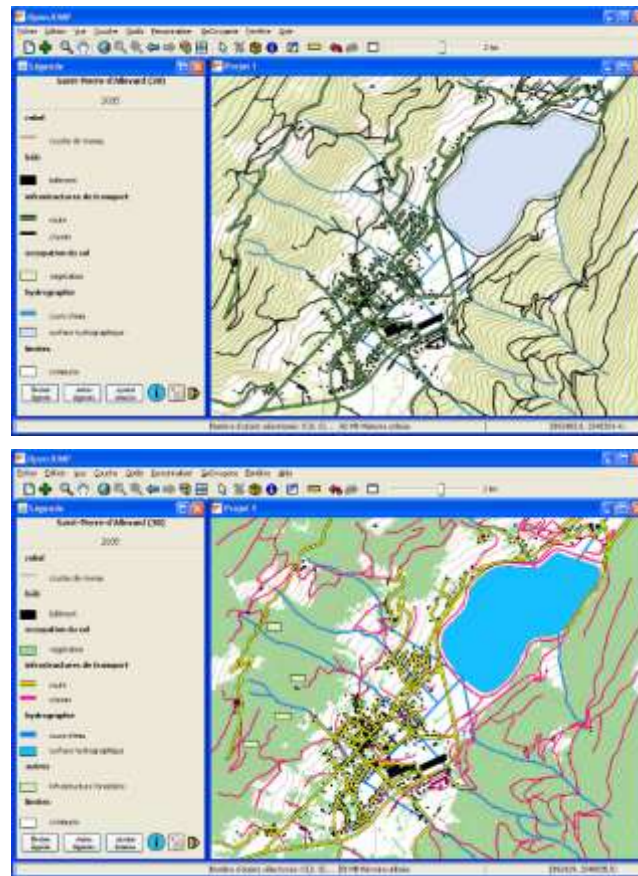


Figure 74. Deux propositions de légendes topographiques (dans le deuxième cas, la requête comporte une couche utilisateur).

2.2.4 Un dialogue pour la création collaborative des styles d'une carte

2.2.4.1 Motivation et travail préliminaire

Un autre travail sur l'aide à la conception de carte sur mesure s'est concentré sur d'autres impératifs.

Un premier impératif est **l'innovation**. A la fin de la deuxième guerre mondiale, l'américain Robinson a constaté le danger des cartes illisibles –voire des cartes qui induisent en erreur- et a affirmé, avec d'autres cartographes, la nécessité de définir des règles objectives en cartographie pour garantir une certaine lisibilité des cartes. Une autre solution aurait été de normaliser les styles cartographiques –quitte à proposer plusieurs normes selon les audiences de la carte pour tenir compte de leurs différences de perception des couleurs-. Cette solution n'a pas été adoptée, les cartographes ont préféré exprimer des contraintes objectives et scientifiques sur le résultat. Il y a deux explications à cela :

- La première est que c'est impossible. Typiquement, il n'y a pas assez de place dans la gamme des couleurs pour réserver une teinte pour chaque thème susceptible d'apparaître sur une carte en gardant des teintes suffisamment contrastées entre elles. La normalisation demeure possible pour des types de cartes pour lesquels les thèmes sont connus à l'avance comme dans le domaine militaire.
- La deuxième est que la normalisation enterrerait l'innovation. La carte est souvent une interface de communication et, dans le domaine de la communication, il est utile de se renouveler, d'innover pour s'améliorer.

Un deuxième impératif est de **permettre aux rédacteurs de cartes de s'impliquer dans le processus et de comprendre la pertinence des règles de sémiologie graphique**. Cet impératif est dicté par le rapport presque passionné qu'entretiennent de nombreuses personnes à l'objet carte. Les nouveaux créateurs de carte mentionnés en section 2.2.1.1 ne se satisferont pas tous de services leur faisant des propositions automatiques comme ceux présentés dans la section précédente. Certains seront plus intéressés, pensons-nous, par une assistance qui leur permet d'exprimer leur sensibilité et de faire des choix au niveau des styles –et non seulement des objectifs et du contenu de la carte-. (Weger 1999) rappelle que la création d'une carte peut répondre à un besoin d'estime. Il définit les fonctions d'estime associées à une carte comme « [Le] plaisir induit par la possession même de l'objet, que l'on peut attribuer notamment à l'esthétique, aux garanties ou à la notoriété qu'il confère. » La carte est une opportunité

pour son *rédacteur* d'exprimer sa personnalité ou une identité. Cette fonction est très importante sur le Web, lieu où de nombreux internautes livrent des informations très personnelles par exemple en tenant un blog (sorte de journal de bord) ou en publiant des photographies ou des vidéos personnelles. Dans certains forums également, le recours à des pseudos imagés illustre à quel point le Web est un lieu où les personnalités s'expriment et où les intervenants sont soucieux de transmettre une image d'eux en plus de leurs opinions. La conception d'une carte permet à son auteur d'exercer sa créativité, d'affirmer son individualité en se différenciant des autres, d'affirmer son appartenance à une communauté en empruntant ses codes, de se valoriser en réalisant une belle œuvre, une prouesse technique ou une œuvre originale. Ainsi, le nouveau créateur de cartes ne recherche pas toujours des solutions toute faites.

Notre approche pour répondre à ces impératifs est de concevoir **une application de dialogue pour l'aide à la conception de carte**. L'exemple de dialogue fictif qui suit vise à montrer les propriétés intéressantes du dialogue homme-homme que veut reproduire le dialogue homme machine :

- J'ai faim. Trouvons donc un restaurant.
- Il y a un snack près de la cathédrale.
-
- Tu penses que c'est trop loin ?
- Non mais je pensais à un vrai restaurant.
- ... je suis un peu fauché en ce moment, reportons à une autre fois.
- [Paumes levées verticalement] Laisse-moi choisir le restaurant et je t'invite.

Globalement cet exemple nous montre qu'un dialogue permet de résoudre un problème qu'une machine ne saurait pas résoudre. Ce dialogue est caractérisé par le fait que les deux interlocuteurs ont un but commun 'déjeuner ensemble'. Il laisse deviner l'importance des actes de communication entre les deux : phrases, silences, mouvements des mains, expressions du visage. Enfin, il illustre l'importance des stratégies pour se rapprocher du but à l'aide des actions de communication. Les stratégies sont importantes pour lever les ambiguïtés. Par exemple, devant le silence de l'un, l'autre essaye de comprendre la raison de ce silence en proposant une explication. Les stratégies sont importantes également pour gérer les conflits. Par exemple, un des interlocuteurs veut sortir du conflit en reportant le but. L'autre interlocuteur négocie en faisant évoluer le but. Pour faire une carte, l'utilisateur et la machine experte en cartographie auront besoin de se comprendre mutuellement, en particulier la machine devra comprendre ce que sont les goûts de l'utilisateur. Ils auront besoin de négocier si l'utilisateur veut par exemple beaucoup de couleurs parce qu'il n'anticipe pas sur l'effet que cela produira sur le lecteur. Enfin, la machine devra s'adapter à la façon de créer de l'utilisateur. Nous pensons pour toutes ces raisons que le dialogue homme-machine est une technique intéressante pour l'aide à la conception de cartes.

Une étude préliminaire a été entreprise par Catherine Domingues qui a repris l'architecture de l'application de dialogue proposée par Frédéric Hubert lors de sa thèse

au COGIT. Le travail de Frédéric Hubert visait à aider un utilisateur novice à paramétrer correctement un traitement de généralisation en se fondant sur des échantillons cartographiques. Les échantillons utilisés dans l'application de (Hubert 2003) étaient des exemples de bâtiments transformés par le traitement avec différentes valeurs des paramètres. Dans la proposition de (Domingues and Bucher 2006), cette proposition est extrapolée au cas du choix de couleurs dans une légende et les échantillons utilisés sont les exemples de cartes présentés dans la section 2.2.4.

Nous avons rapidement rencontré deux difficultés.

La première difficulté est de parcourir l'espace des solutions (c'est-à-dire l'espace des combinaisons de couleurs possibles). Ce parcours doit tenir compte de ce que le serveur apprend sur le besoin de l'utilisateur au fur et à mesure du dialogue. Ce parcours doit aussi, dans un premier temps, respecter la variété du champ des possibles et, dans un deuxième temps, restreindre le champ des possibilités et aider l'utilisateur à converger vers une solution. L'espace des couleurs possibles pour une légende ne se prête pas à une structuration simple comme l'espace des valeurs possibles des paramètres du traitement de généralisation qui est représenté comme un parallélépipède à trois dimensions dans la proposition de (Hubert 2003). Il est vite apparu que le parcours de cet espace de solution, de façon intelligente, était problématique. Dans la proposition de (Hubert 2003), une hypothèse de monotonie du comportement du traitement de généralisation selon chaque paramètre permettait d'explorer cet espace de façon pertinente avec les retours de l'utilisateur et également de façon à converger vers des valeurs de plus en plus précises pour chacun de ces paramètres. Explorer intelligemment l'espace des couleurs possibles pour une légende demande de nouvelles solutions.

La deuxième difficulté est de proposer des actes de communication pertinents à l'utilisateur. Un aspect de l'application Cartable proposée par (Hubert 2003) que nous souhaitons conserver était la possibilité pour l'utilisateur de réagir par des mots. Pour être plus précis, Cartable comporte un moteur de dialogue en langue naturelle dont la base de connaissances avait été enrichie pour comprendre des phrases comme 'je veux généraliser un bâtiment'. Nous n'avons pas cherché dans un premier temps à permettre à notre utilisateur d'exprimer des phrases. Mais le modèle de communication de Cartable repose par ailleurs sur une interprétation des expressions en langue naturelle en des opérations de restriction de l'espace des solutions acceptables : 'plus petit', 'moins détaillé' signifient qu'il faut augmenter ou diminuer la valeur d'un paramètre ou de l'autre. Cette interprétation s'appuie sur le fait que, dans le modèle de généralisation à base de contraintes, les paramètres sont en fait des seuils de valeurs de contraintes. Nous voulons à terme permettre à l'utilisateur de réagir avec des expressions comme 'plus moderne', 'plus lumineux', 'moins primaire' que le serveur de dialogue saurait interpréter pour affiner ses propositions de styles. A ce stade, la difficulté est de proposer des termes non ambigus et de savoir les interpréter. Il est nécessaire

d'acquérir davantage de connaissances sur l'expression d'un ressenti face à des exemples de cartes ou face à une carte en cours de conception pour savoir interpréter de telles expressions en des opérations sur les styles proposés. Cette nécessité a déjà été mentionnée à la suite des premiers résultats de qualification des échantillons car il s'agit du même travail globalement.

Ainsi le travail de dialogue pour la création de carte a été scindé en deux :

- un travail sur la compréhension du vocabulaire associé aux cartes conduit par Catherine Domingues
- un travail sur l'exploration des choix de couleur possibles pour une carte (sans même proposer l'usage de mots par l'utilisateur).

Pour ce deuxième travail, j'ai proposé un sujet de thèse : une méthode de conception coopérative de légende. Cette thèse a été menée par Sidonie Christophe sous la direction d'Anne Ruas, directrice du COGIT. Nous présentons ce travail dans la suite de cette section.

2.2.4.2 Une méthode de conception coopérative de légende

La proposition de (Christophe et al. 2007) consiste en une application de dialogue pour la conduite coopérative de méthodes de spécification de couleurs dans une légende. L'application possède des méthodes de choix de couleurs et sait conduire chacune de ces méthodes de façon coopérative. L'application sait aussi observer le dialogue et réagir lorsqu'elle détecte un problème. L'expertise du serveur de dialogue est utile d'une part pour déclencher et encourager un processus créatif et d'autre part pour évaluer l'efficacité de la carte en cours de conception.

L'expertise de l'application de dialogue est modélisée sous forme d'activités hormis la stratégie globale de dialogue qui est représentée procéduralement. Les activités permettent de modéliser plusieurs connaissances :

- les méthodes coopératives de choix de couleurs,
- l'évaluation d'une carte,
- la proposition de couleurs pour retoucher une carte
- les activités d'évaluation et de réparation du dialogue.

Cette modélisation a plusieurs avantages. Elle correspond à la nature des connaissances modélisées. Elle factorise simplement certaines connaissances comme par exemple la construction de proposition de légendes en fonction de couleurs choisies qui est une connaissance partagée par les deux méthodes coopératives modélisées. Cette modélisation a aussi l'avantage d'être flexible. Elle permettra d'étendre le système avec d'autres méthodes de choix de couleurs qui seront exploitables directement par le moteur de dialogue.

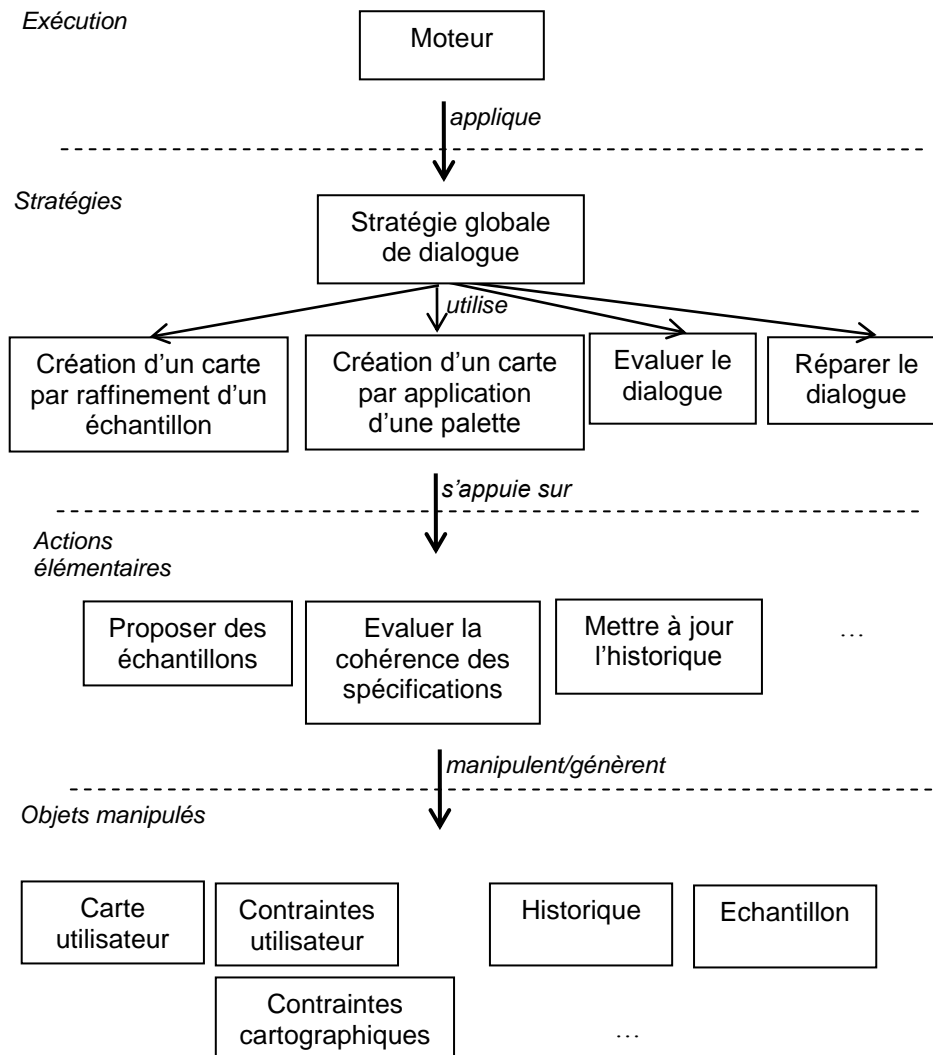


Figure 75. Un modèle à base d'activités permet de prendre compte les natures des différentes connaissances intervenant dans l'expertise du système.

Un apport important de ce travail est d'avoir proposé des méthodes efficaces de construction de styles innovants qui permettent de faire coopérer le système et l'utilisateur. L'approche globale consiste à permettre à l'utilisateur d'exprimer ses préférences à partir d'exemples puis d'extrapoler ce que ces exemples indiquent (des couleurs, des couleurs pour des thèmes, des combinaisons de couleurs, ...) à la carte en cours de conception. Il s'agit d'une méthode classique en aide à la conception (Christophe 2009).

Cette méthode de conception générique se décline en deux sous-cas correspondant à deux types d'exemples : les échantillons cartographiques qualifiés et les toiles de maître. Ces deux sous-cas pourront à l'avenir être fusionnés mais ils ont permis à Sidonie Christophe de se concentrer sur deux aspects différents du processus de conception coopérative.

Les échantillons sont en nombre suffisamment nombreux pour étudier le processus d'expression des préférences. Ce processus consiste ici à aider l'utilisateur à explorer la base des échantillons en lui fournissant à chaque fois des échantillons à la fois variés et qui tiennent compte de ce qu'il a pu exprimer auparavant. Il s'agit donc du problème d'exploration de l'espace des possibilités infinies mentionné dans la section précédente. La solution s'appuie ici sur la classification des échantillons. Une première classification s'appuie sur les qualificatifs. Elle permet à l'utilisateur d'exprimer ses préférences en termes d'impact émotionnel attendu. Une deuxième classification s'appuie sur les deux couleurs dominantes dans la légende de chaque échantillon (stage de Guillaume Ménégaux encadré par Sidonie Christophe). Cette deuxième classification est particulièrement intéressante car elle peut être faite à la volée. Autrement dit, lorsque l'utilisateur a exprimé ses préférences sur des premières planches, le système peut calculer les regroupements des échantillons correspondant aux préférences de façon à lui présenter des échantillons encore représentatifs de la variété des possibles. Par ailleurs, le système permet à l'utilisateur d'exprimer ses préférences à partir d'un échantillon en disant qu'il aime l'échantillon ou une couleur de l'échantillon ou une combinaison de couleurs.

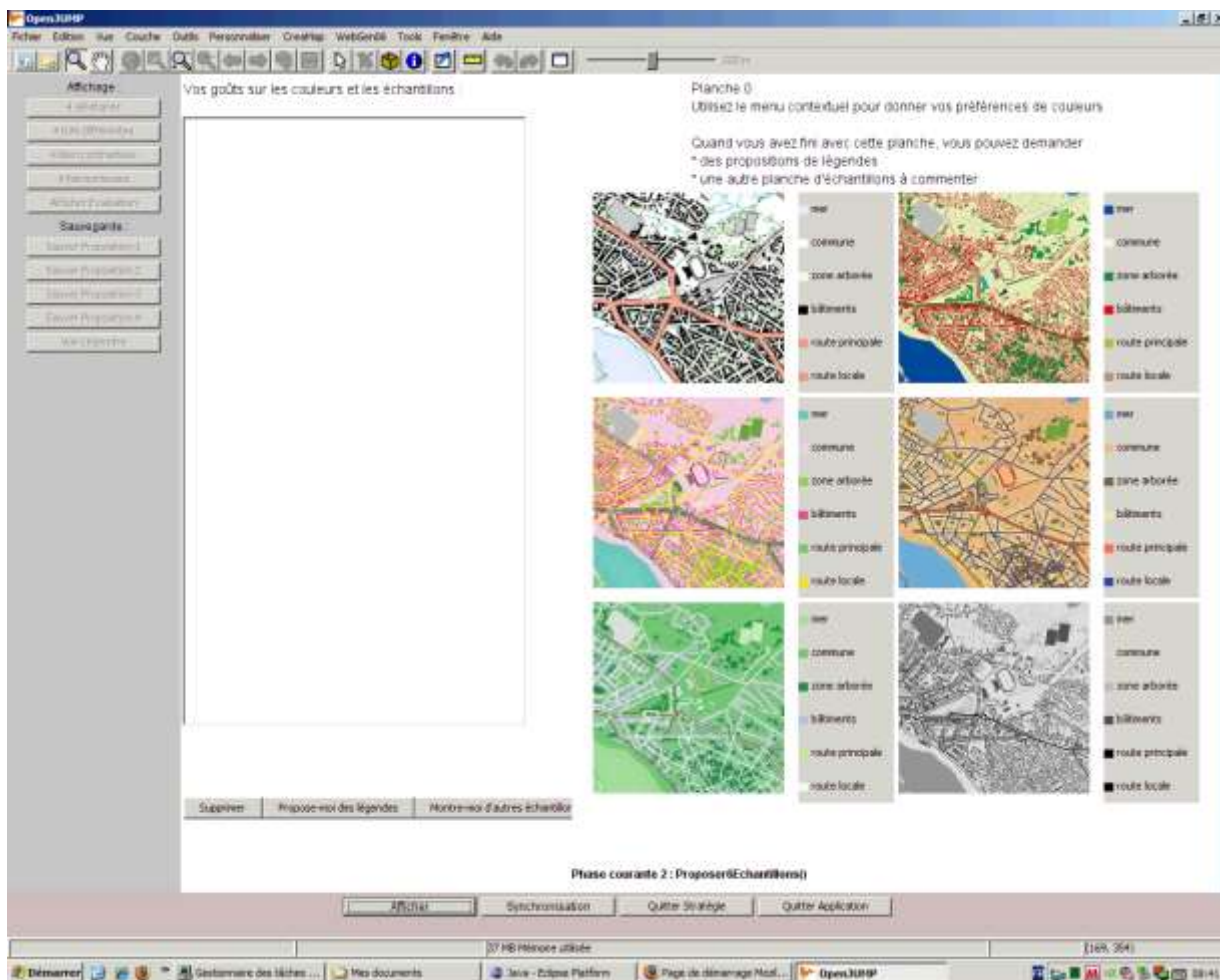


Figure 76. Interface du prototype permettant à l'utilisateur d'exprimer ses préférences à propos d'échantillons.

Lorsque les préférences sont exprimées, l'extrapolation se fait en prenant simplement les couleurs désignées dans les échantillons pour faire des propositions de carte. La difficulté consiste à gérer les préférences pour traiter ensemble celles qui sont cohérentes. De plus, le système peut également intervenir pour modifier la couleur du fond de carte pour améliorer la lisibilité.

Actuellement l'expression des préférences à partir des toiles consiste à en sélectionner une. Le système interprète cette sélection comme une expression d'un goût pour les couleurs utilisées dans la toile et les règles de composition utilisées par le peintre dans cette toile. La toile étant plus loin de la carte que l'échantillon, l'étape d'extrapolation de l'exemple à la carte de l'utilisateur a été particulièrement étudiée. L'expertise du système intervient notamment pour les opérations suivantes :

- affecter le plus correctement possible les couleurs aux thèmes de la carte

- modifier certaines couleurs de façon à ce que les valeurs de contrastes de couleurs soient acceptables et conformes aux relations entre thèmes de la carte.

Des exemples de cartes conçues selon cette méthode sont repris sur la Figure 77.

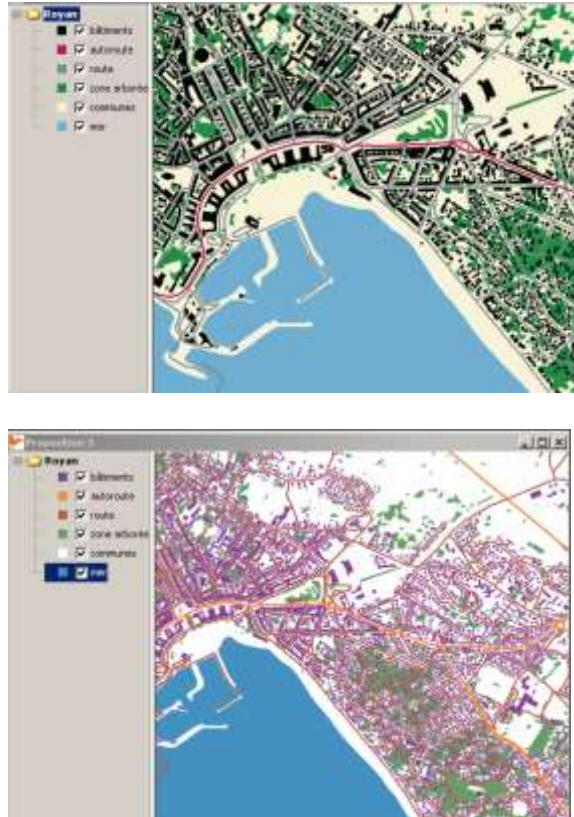


Figure 77. Cartes obtenues à partir des toiles *La Tristesse du Roi* de Henri Matisse (en haut) et *Les Vierges* de Gustav Klimt (en bas).

Enfin, l'utilisateur a la possibilité de demander la relaxe de certaines règles cartographiques (les règles sur les conventions de couleurs). L'application lui propose un curseur entre les règles de grammaire du peintre et les règles de compositions dictées par la cartographie. L'effet de ce curseur est illustré sur la Figure 78.



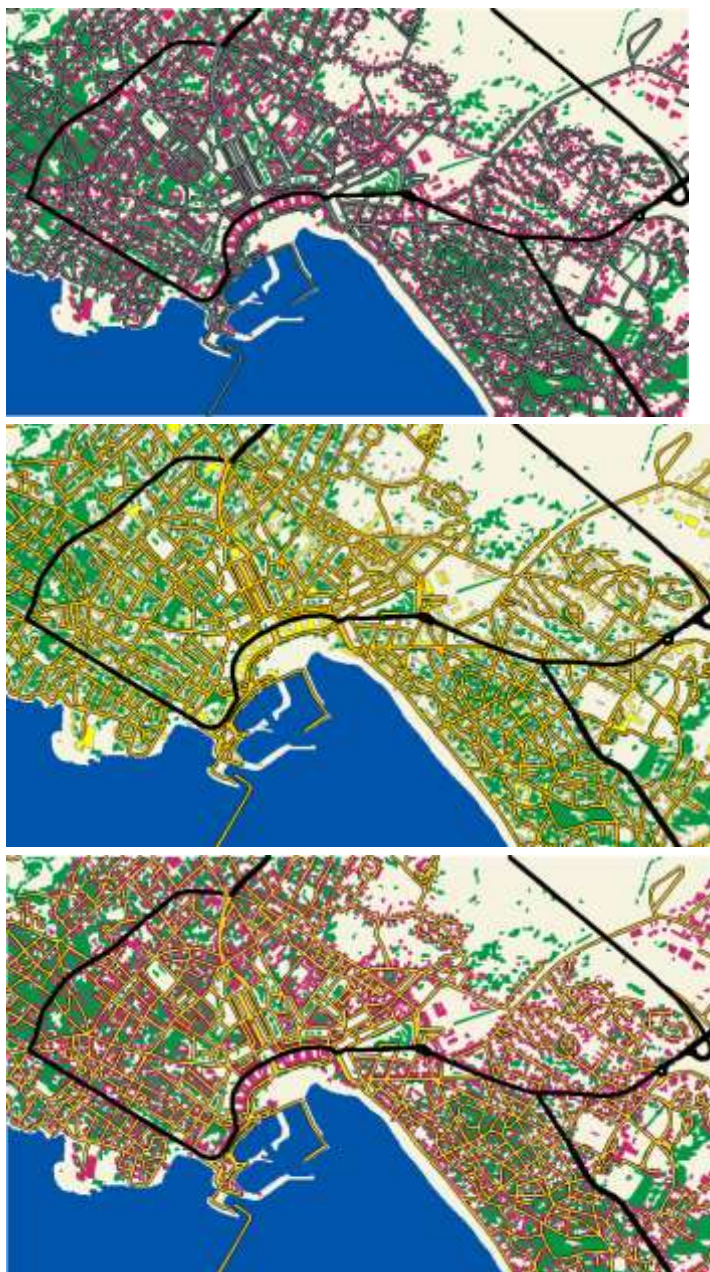


Figure 78. Cartes obtenues à partir d'une toile de Matisse en reprenant les couleurs de la toile et les règles de composition de la toile puis en introduisant peu à peu les règles cartographiques pour améliorer la lisibilité de la carte.

2.3 Bilan

Cette section propose un bilan de cette deuxième partie du mémoire.

Un premier apport est notre modèle de description de traitements. L'originalité de ce modèle est de faire le lien entre des concepts du domaine des services (fonction fournie et contexte) et des concepts familiers aux développeurs (ressources logicielles) via des objets *Activités*. Un autre résultat précieux est la capitalisation d'expériences liées à la description de traitements qui nous ont permis de comprendre des difficultés de ce domaine. L'approche actuelle consiste à utiliser un modèle conceptuel unifié mais à construire des prototypes dédiés : un prototype dédié à l'élaboration d'un vocabulaire commun dans ce domaine et un prototype dédié au partage effectif de méthodes développées dans la plate-forme Geoxygene. Pour aller plus loin, il est important d'intégrer des travaux existants sur la prise en compte du contexte pour inciter des développeurs de méthodes par exemple à intégrer les critères proposés dans la plate-forme MOBIDIC. Il est également important de mettre en place des plate-formes de services dédiées à la recherche en géomatique, à la comparaison des méthodes et également à la construction de nouvelles applications à partir de méthodes provenant de différentes équipes. Cette approche est amorcée dans le groupe de travail 'Mutualisation' du GDR MAGIS que j'anime avec Florence Leber. Nous étudions en particulier les travaux WebGen pour extrapoler leur approche aux équipes du GDR ainsi que les travaux Mobidic pour permettre aux équipes de développer des services utilisables en contexte mobile.

Un autre apport est notre proposition globale pour un processus Carte à la Carte qui s'appuie sur une explicitation des intentions associées à la carte en cours de conception. De plus, nous avons introduit un nouvel élément par rapport aux propositions similaires : l'impact émotionnel attendu de cette carte. Un résultat important est un modèle de spécification de carte. Un autre résultat est le catalogue de style et les méthodes associées pour permettre au rédacteur de la carte de réutiliser des styles existants. Ce résultat me semble opérationnalisable et il est important de bien le présenter aux services de l'IGN concernés. Un autre apport est d'avoir formalisé une part de l'expertise nécessaire à la création d'une carte : la créativité. Cette formalisation s'inspire des modèles existants en aide à la conception : utiliser des sources d'inspiration et extrapoler une solution en respectant des contraintes. Informatiquement, notre proposition opérationnelle d'application de conception coopérative et de dialogue

homme-machine sont également intéressants car ce sont des applications de techniques de modélisation à des problèmes réels.

Bilan et Perspectives

Ce mémoire a présenté des travaux que j'ai conduits dans le contexte de ma thèse, de ma participation à des projets ou de l'encadrement de l'action de recherche sur l'accès en ligne aux données et aux processus de dérivation géographiques. Voici un bilan global de ce mémoire, qui fait suite aux bilans intermédiaires de chacune de ces deux parties.

Principaux apports

Les apports de nos travaux ont été présentés dans les bilans intermédiaires. Je rappelle les principaux:

La thèse de Sandrine Balley a proposé un **modèle de formalisation de la structure complète d'un jeu de données géographiques** qui a été exploité pour la manipulation de schéma. Ce modèle ouvre des perspectives par exemple pour l'aide à l'édition collaborative de données géographiques conformes à des spécifications (thèse de Carmen Brando). Elle a aussi proposé un **modèle de transformation dynamique de schémas**. Ce modèle ouvre des perspectives pour la fusion et la propagation des mises à jour.

D'autres travaux se sont concentrés sur **le partage et la réutilisation de traitements, fondés sur une description orientée services des ressources logicielles existantes**, des processus connus et des algorithmes abstraits.

Nous proposons deux approches **d'aide à la conception d'une carte** qui s'appuie sur la représentation explicite des objectifs de la carte : une approche la plus automatisée possible qui s'appuie sur un catalogue de styles prédéfinis et une approche d'aide à la création de styles fondée sur une formalisation de la créativité.

Notre **modèle de dialogue**, appliqué à l'affectation de couleurs dans une légende, peut être réutilisé dans d'autres contextes qui nécessitent un dialogue. Ces contextes me semblent nombreux. Il s'agit de situations dans lesquelles des utilisateurs sont amenés à visualiser des données, à les éditer et à les transformer et que des experts en géomatique peuvent leur apporter une expertise mais que cette expertise doit être confrontée avec leur propre expertise et leurs intuitions.

L'ensemble de ces travaux ont permis **la constitution au COGIT d'une expertise nouvelle liée au domaine de la diffusion**. Cette expertise est alimentée par une expérience sur des projets abordant des domaines techniques adjacents de nos problématiques classiques et par une veille scientifique continue.

Revisiter, faire évoluer et diffuser la théorie cartographique et la géomatique

L'évolution des technologies et des pratiques est une caractéristique forte de la diffusion d'information géographique en ligne. Lors de la séance d'ouverture de la 23^{ème} conférence de l'Association Internationale de Cartographie en 2007, Menno Jan Kraak et Ferjan Ormeling ont interpellé l'audience de la façon suivante (en substance): « La théorie cartographique n'évolue pas assez vite et est dépassée par les nouvelles technologies. Nous, chercheurs, devons trouver un moyen de faire évoluer notre expertise pour nous adapter à ces technologies alors même que leur étude satisfaisante mobiliserait entièrement notre attention. » Il faut souligner que cette préoccupation (faire évoluer la cartographie) n'est pas nouvelle. Dès la fin des années 80, avec l'évolution des technologies de l'information et de la communication, la cartographie a dû évoluer (Visvalingam 1989). En 1988, l'Association Internationale de Cartographie, sensible à cette nécessité, a demandé à ses membres de redéfinir le domaine de la cartographie. Pour faire évoluer la cartographie il fallait avant tout s'appuyer sur une définition de ce domaine. Il s'est avéré que nous ne possédions pas une telle définition faisant consensus (par exemple, certains pays y intègrent la photogrammétrie, d'autres non). (Freitag 1993) souligne par ailleurs que toutes les définitions avancées excluaient ou minoraient les aspects 'usage de l'information géographique'. A l'approche de 2010, la remarque de (Freitag 1993) doit être révisée car les usages sont largement intégrés dans les thèmes des conférences géomaticiennes.

Mutualiser l'expertise et les ressources entre chercheurs en géomatique

Aujourd'hui, un enjeu fort est la diffusion de l'expertise entre communautés scientifiques étudiant la géomatique pour faire progresser ce domaine scientifique, comme c'est le cas au sein du Groupement de Recherche Pluridisciplinaire MAGIS. En pratique, des projets comme Geotopia visant le partage de ressources documentaires annotées sont des approches pragmatiques à la communication et à la collaboration entre chercheurs sur un même domaine. Au sein du projet Mutualisation du GDR MAGIS, Florence Le Ber et moi-même avons observé le besoin de mutualisation de composants logiciels entre des équipes du GDR et la pertinence de cette mutualisation.

Cette tâche est compliquée du fait que des problématiques à caractère 'information spatiale' par l'apparition de nouveaux domaines dans le champ 'est d'autant plus Les frontières entre les objets d'étude de la géomatique et ceux d'autres domaines évoluent, au fur et à mesure que de nouvelles problématiques à caractère 'information spatiale'

surgissent. Par exemple, citons trois communautés qui étudient le problème de la recherche d'information géographique :

- la communauté purement géomaticienne type AGILE.
- la communauté recherche d'information (qui travaille sur les moteurs de recherche plein texte).
- la communauté du Web sémantique avec par exemple la conférence Where2.0.

Il faut donc établir une communication et une collaboration entre ces communautés. Le workshop Geographical Information Retrieval (GIR) réunit ainsi une jeune communauté composée par la collaboration de géomaticiens et de chercheurs en recherche d'information. Son équivalent français, Georama, a vu le jour récemment. La collaboration des deux communautés (recherche d'information et géomatique) est fructueuse comme nous l'avons expérimenté au sein de projet européen SPIRIT. Il est nécessaire d'identifier les communautés scientifiques avec lesquelles nous devons collaborer et de réussir à collaborer avec elles (les deux tâches peuvent être ardues).

Des 'communautés scientifiques' liées au Web et aux logiciels géographiques opensource doivent aussi être considérées. Il faut alors tenir compte de leur mode de travail qui se démarque un peu des approches académiques. Ce sont des communautés qui augmentent leur savoir via les forums et via google. Ces communautés ne poursuivent pas certains buts des communautés scientifiques (faire progresser la science, capitaliser le savoir, ...) et ont des méthodes différentes. Le principe roi est d'être 'community-driven'. Mais leurs sujets d'étude sont parfois très proches. Ainsi, en 2007 a eu lieu la première édition d'une 'nonconférence' pour les geohackeur, sponsorisée par Google : conférence gratuite. Voici les thèmes annoncés : augmented reality, usability and design science, remote sensing, geoinformatics, forestry and agriculture, food chain transparency, civil engineering, emergency disaster relief, urban planning, local search, context awareness, place hacking, social cartography, citizen journalism, locative gaming, psychogeography, locative art, iphones androids and the way the web is falling into mobile.

En définitive, nous devons apprendre d'autres domaines et leur apporter notre expertise. Contrairement à ce que nous avons pu penser, il existe des modèles simples pour créer dynamiquement des cartes sur internet, il existe des modèles économiques réussis nouveaux. Et par ailleurs ces outils ne suffisent pas pour créer de bonnes cartes. La collaboration récente de Google et ESRI montre l'importance de la géomatique et de ses modèles pour le Web géographique.

Intégrer l'expertise dans les processus de prise de décision

Par ailleurs, les nouvelles infrastructures d'information et les nouvelles interfaces ont réduit certaines chaînes d'interventions d'experts humains dans des prises de décisions liées à l'espace géographique. Si on facilite l'usage de données géographiques, il faut aussi rendre disponible (sous forme de services d'évaluation et de messages d'alerte

éventuellement) l'expertise requise pour prendre des décisions. Si le décideur politique a accès à une interface qui ne lui présente pas uniquement une carte de synthèse mais lui permet d'interagir pour simuler facilement de nouveaux aménagements, des problèmes peuvent surgir (l'aménageur est court-circuité). Par ailleurs, il est important de prendre en compte la production croissante de contenu localisé et de cartes par de nouveaux acteurs. Il nous semble essentiel d'identifier les éléments à ajouter dans une architecture de diffusion pour s'assurer que toutes les expertises nécessaires à un bon processus seront utilisées : quelle expertise, à quel moment, par qui ? Plusieurs réponses sont généralement possibles. Il est parfois possible d'intégrer une expertise directement dans un processus de façon transparente pour l'utilisateur. Cette approche est très utile pour faciliter l'accès à des fonctionnalités qui demandent une expertise et dont on veut qu'elles soient effectuées dans les règles. Par exemple cela revient à proposer automatiquement des couleurs conformes aux règles de sémiologie dans un processus de carte à la carte (approche présentée en section 2.2.3.3). Verrouiller les processus pour qu'ils se déroulent de façon conforme à ce que notre expertise nous apprend n'est pas toujours le bon choix. Parfois, il est nécessaire en plus de communiquer cette expertise voire de la transmettre pour que les utilisateurs apprennent à la mettre en application. Il semble qu'en terme de définition des styles par exemple, cette dernière solution soit prometteuse. Si la plupart des personnes reconnaissent la capacités des agences cartographiques de produire de 'belles cartes', il est parfois difficile pour l'utilisateur novice d'admettre l'importance du respect de règles de sémiologie dans une carte qu'il construit lui-même. Or si une application de cartographie en ligne bride sa liberté créatrice, il se tournera peut-être vers d'autres solutions existantes plus flexibles.

Organiser le travail sur la diffusion dans un univers en constante évolution

Il faut également s'intéresser à l'organisation des ressources de développement pour réaliser les composants de l'architecture de diffusion. Il y a un équilibre à trouver pour satisfaire deux types d'enjeux.

— L'efficacité et l'efficience de l'architecture pour réaliser les fonctionnalités qui lui ont été assignées (par exemple : la carte doit s'afficher rapidement,...). La satisfaction de cette exigence tend à renforcer le couplage entre composants de l'architecture (qu'ils soient bien d'accord à l'avance sur la façon dont ils vont ensemble réaliser la fonctionnalité dans tous les cas qui vont se présenter).

— Le cycle de vie de l'architecture et des éléments qui la composent (les fonctionnalités de maintenance issues de l'analyse fonctionnelle ne couvrent pas toujours tout, d'autant plus que ces enjeux sont souvent implicites pour les ingénieurs système). La satisfaction de ces exigences tend à maximiser l'autonomie de chaque composant c'est-à-dire à minimiser le couplage entre eux.

Nous transposons ces notions de couplage entre les services concernés au sein de l'IGN. Il est important d'analyser les communications et interfaces nécessaires entre services pour coopérer efficacement sur ces problématiques de diffusion. Ce qui nous semble

important : savoir s'abstraire des couches de bas niveaux et des préoccupations spécifiques à un service ou à un autre pour échanger des informations compréhensibles par chacun et pertinentes.

Quelques enjeux techniques

Après ces considérations méthodologiques, voici une liste d'enjeux techniques.

Le premier est la facilitation de l'intégration et de la recherche de contenu localisé dans les infrastructures d'information spatiale. Cela passe par exemple par des techniques de qualification de la sémantique d'un contenu (son lien avec la réalité) et de ses usages possibles. Il est également nécessaire de proposer des méthodes d'interrogation des métadonnées géographiques par des personnes, en utilisant par exemple des techniques de visualisation.

La performance des serveurs de données géographiques est également un enjeu technique de taille. Définir les bons systèmes d'index qui croisent le spatial avec le temporel et le thématique et qui facilitent le changement de niveau de détail est encore un domaine de recherche dans lequel nous devons progresser.

L'amélioration de la diffusion passe aussi par des solutions pour faciliter l'usage des représentations de l'espace géographiques (généralement des cartes) par les destinataires. Un aspect important est par exemple de faciliter l'accès au bon niveau de détail, le passage d'un niveau de détail à un autre voire la juxtaposition de plusieurs niveaux de détails. La plupart des activités de l'homme ont besoin de plusieurs niveaux de détails.

Un autre enjeu technique important est également le partage de connaissance entre les experts impliqués dans la recherche sur le développement durable et la gestion du risque, via des SIG participatifs ou via des modeleurs. Comme mentionné dans (Bucher 2002 thèse), le raisonnement et la prise de décision relatifs à l'espace géographique devraient intégrer d'autres méthodes issues de domaines comme la géographie, l'hydrologie, l'architecture. Une solution de diffusion devrait donc permettre l'intégration de ces méthodes dans un processus de construction de connaissances géographiques (un peu à la façon des assistants dans la plupart des logiciels complexes).

Bibliographie

- Abadie N., 2009, Schema Matching Based on Attribute Values and Background Ontology, in *proceedings of the 12th AGILE International Conference on Geographic Information Science*, Hanovre (Germany)
- Abd El Kader, Y., 2006, *Un serveur de saisie et de consultation de métadonnées sur les traitements géographiques*, Thèse en informatique de l'Université de Caen
- Ahonen-Rainio P., 2001. Description of the content of geographic datasets. in *proceedings of 8th Scandinavian Research Conference on Geographic Information Science*, SCANGIS'2001, As, Norway
- Andrienko, G., Jern, M., Dykes, J., Fabrikant, S., I., Weaver, C., 2007, Geovisualization and synergies from InfoVis and Visual Analytics, *11th International Conference Information Visualization*
Proceedings, IEEE Computer Society Press, 2007
- Aumaitre, G., Balley, S., Levin, S., 2004, *Evaluation methodology*, SPIRIT IST-2001-35047 deliverable D19 7201
- Aumaitre, G., Balley, S., Levin, S., Bucher, B., 2004b, *Evaluation of the interim prototype*, SPIRIT IST-2001-35047 deliverable D20 7202
- Balley, S., Bucher, B., Petrelli, D., Ruas, A., van Kreveld, M., Sanderson, M., Sester, M., 2002, *User Requirements Specification*, SPIRIT IST-2001-35047 deliverable D3 7101
- Balley S., 2007, *Aide à la restructuration de données géographiques sur le Web - Vers la diffusion à la carte d'information géographique*. Thèse en informatique de l'Université de Paris-Est Marne-la-Vallée, Saint Mandé
- Barrault M., Regnaud N., Duchêne C., Haire K., Baeijs C., Hardy P., Mackaness W., Ruas A., Weibel R., 2001, Integrating multi-agent, object oriented and algorithmic techniques for improved automated map generalization, in *the proceedings of the 20th International Cartographic Conference*, Beijing (China)
- Becam A., 2002, *YEHUDI : Un environnement pour l'interopérabilité de modèles urbains distribués et homogènes*. Thèse de Doctorat en Informatique et Information pour la Société de l'INSA de Lyon
- Benamara, F., Gardent, C., Saint-Dizier, P., 2003, Réponses coopératives et logiques de description, *Journées GDR Sémantique et Modélisation*, Paris
- Berners-Lee, T., Hendler, J., Lassila, O., 2001, The semantic Web, *Scientific American*, Vol 284 (5)

- Bertin, J., 1983. *Semiology of Graphics: Diagrams, Networks, Maps*. University of Wisconsin Press, Madison
- Bianchi-Berthouze, N., Hayashi, T., 2002, Interpretation of complex data: requirements to support kansei mining, in *Proceeding of International workshop on multimedia data mining*, in conjunction with the Seventh ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Edmonton, Canada
- Boucheritte, A., 2009, *Exploitation des enrichissements sémantiques d'un monde de synthèse: aide à l'exploration*, rapport de stage de Mastère SI, Université de Marne la Vallée Paris Est, Ecole Nationale des Sciences Géographiques
- Brewer, C., Battenfield, B., 2007, Framing Guidelines for multi-scale mapmaking using databases at multiple resolutions, in *Cartography and Geographic Information Science*, Vol 34(1)
- Buard E., Ruas A., 2007, Evaluation of colour contrasts by means of expert knowledge for on-demand mapping, in *proceedings of the XXIIIe International Cartographic Conference*, Moscow
- Bucher, B., Clough, P., Joho, H., Purves, R.S., & Syed, A. K., 2005, Geographic IR Systems: Requirements and Evaluation. in *Proceedings of the 22nd International Cartographic Conference*, A Coruña, Spain
- Bucher, B., Balley, S., Levin, S., Syed, A., K., Petrelli, D., Weibel, R., Abdelmoty, A., Beaulieu, M., Heinzle, F., 2004, *User requirements specification reassessment*, SPIRIT IST-2001-35047 deliverable D3 7101B
- Bucher, B., Clough, P., Finch, D., Joho, H., Purves, R., Syed, A., K., *Evaluation of SPIRIT prototype following integration and testing*, SPIRIT IST-2001-35047 deliverable D31 7301
- Bucher, B., 2002, *Metadata and markup languages*, SPIRIT IST-2001-35047 deliverable D2 6101
- Bucher B., 2003, L'accès à l'information géographique: l'apport de la modélisation de tâches géographiques, *Revue Internationale de Géomatique*, Vol 13 N°4/2003
- Bucher, B., 2007, La Carte à la carte sur le Web, dans *Le Monde des Cartes, bulletin du Comité Français de Cartographie*, No193
- Bucher B., Balley S., Richard D., Cébelieu G., Hangouët J-F., 2005, Shareable descriptions of data production processes, in proceedings of the 8th AGILE conference, Estoril, Portugal
- Bucher B., Jolivet L., Buard E., Ruas A., 2007, The need for Web legend services, in *7th International Symposium on Web and Wireless GIS (W2GIS)*, J.M.Ware and G.E.Taylor (Eds), Springer Lecture Note in Computer Science, Cardiff, UK
- Bucher, B., Jolivet, L., 2008, Acquiring service oriented descriptions of GI processing software from experts. in *proceedings of the 11th AGILE Conference*, Girona, Espagne
- Bucher, B., Vatan, B., 2008, *Spécification des services d'interface de contenu*, Livrable D3.1.3.2-2, affaire TerraData-1
- Bucher B., Balley S., 2007, A generic preprocessing service for more usable data processing services, 10th AGILE Conference, 8-11 May, Aalborg (Denmark)

- Burghardt, D., Schmid, S. and Stoter, J.E., 2007, Investigations on cartographic constraint formalisation, in *proceedings of the workshop of the ICA commission on generalization and multiple representation*, at the XXIIe International Cartographic Conference, Moscow, Russia
- Cañamero, L.D., 2001. Emotions and Adaptation in Autonomous Agents: A Design Perspective, *Cybernetics and Systems: An International Journal*, Vol 32 (5)
- Carneiro, T., Camara, G., *A gentle introduction to TerraME*, v1.0, INPE report, 2006
- Cartwright, W., 2008, Delivering geospatial information with Web2.0, in *International perspectives on maps and the internet*, Peterson, M. P., (Ed), Berlin: Springer
- Chandrasekaran, B., Josephson, J.R., Richard Benjamins, V., 1998, The Ontology of Tasks and Methods, in *proceedings of the Eleventh Workshop on Knowledge Acquisition, Modeling and Management*, KAW'98, Alberta, Canada
- Chesneau E., 2006, *Modèle d'amélioration automatique des contrastes de couleurs en cartographie - Application aux cartes de risque*, Thèse de doctorat, Université de Marne La Vallée
- Chesneau E., Ruas A., Bonin O., 2005, Colour Contrasts Analysis for a better Legibility of Graphic Signs on Risk Maps. in *proceedings of the XXIIeme International Cartographic Conference ICC' 2005*, La Coruna, Spain
- Christophe S., Bucher B., Ruas A., 2007, A dialogue application for creative portrayal, *International Cartographic Conference*, Moscow (Russia)
- Claramunt C., Mainguenaud M., 1996, A Spatial Data Model for Navigation Knowledge, in *proceedings of SDH '96*
- Clouard, R., Elmoataz, A., Revenu, M., 2002, Une méthodologie de développement d'applications de traitement d'images, *actes des Rencontres Francophones en Intelligence Artificielle*, Angers
- Clouard, R., Renouf, A., Revenu, M., 2008, Un atelier d'ingénierie des connaissances en traitement d'images, *actes des 8^{èmes} journées francophones Extraction et Gestion des Connaissances, atelier ECOI*, Sophia Antipolis
- Coote, A., Rackham, L., 2008, Neogeographic data quality – is it an issue ?, AGI Conference, UK
- Craglia, M., Goodchild, M. F., Annoni, A., Camara, G., Gould, M., Kuhn, W., Mark, D., Masser, I., Maguire, D., Liang, S., Parsons, E., 2008, Next-Generation Digital Earth: A position paper from the Vespucci Initiative for the Advancement of Geographic Information Science Export, *International Journal of Spatial Data Infrastructures Research*, Vol. 3
- Devillers, R., 2004. Conception d'un système multidimensionnel d'information sur la qualité des données géospatiales. Thèse de doctorat, Département des sciences géomatiques, Université Laval (Canada) & Institut Francilien des Géosciences, Université de Marne-la-Vallée
- Devogele, T., Parent, C., Spaccapietra, S., 1998, On Spatial Databases Integration, *International Journal of Geographical Information Science*, vol 12, pp.335-352

- Domingues C., Bucher B., 2006, Legend design based on map samples, in proceedings of the 4th GIS Science conference, Extended abstracts, Raubal, Miller, Frank and Goodchild (Eds), IFGI Prints, Münster
- Donaubauer, A, Straub, F, Schiilcher, M., 2007, md-WFS : A concept of web-enabling semantic transformation, in *Proceedings of the 10th AGILE Conference on Geographic Information Science*
- Doucet, A., Gañarski, S., 2001, Entrepôts de données et BD multidimensionnelles, *Bases de données et internet*, Doucet, A., Jomier, G., (Eds), Informatique et Systèmes d'Information, Hermes Lavoisier, Paris
- Doucet, A., 2006, *PADOUE (Partage de Données Utiles en Environnement), rapport de fin de projet*, projet PADOUE ACI GRID 2002, Paris
- Edsall, R., 2007, Globalization and cartographic design : implications of the growing diversity of map users, in *proceedings of the 23rd ICC conference*, Moscow
- Edwardes, A., Burghardt, D., Neun, M., 2005, Interoperability in Map Generalisation Research, *proceedings of the International Symposium on Generalization of Information*
- Finch, D., technical group, 2003, Specification of System Functionality, SPIRIT IST-2001-35047 deliverable D4 1101
- Foerster, T., Stoter, J., Kobben, B., 2007, Towards a formal classification of generalisation operators, *proceedings of the XXIIIe International Cartographic Conference, Moscow*
- Foerster, T., Schäffer, B., 2007, A client for distributed processing on the Web, in *proceedings of the 7th Symposium of Web and Wireless GIS*, Ware, J.M., Taylor, G., E. (Eds), Cardiff
- Foerster, T., Burghardt, D., Neun, M., Regnauld, N., Swan, J., Weibel, R., 2008, Towards and interoperable Web generalisation services framework – current work in progress, in *Proceedings 11th ICA Workshop on Generalisation and Multiple Representation*, Montpellier, France
- Forrest, D., 1999, Developing rules for Map Design: a Functional Specification for a cartographic design Expert Sytem, *Cartographica*, vol 36 (3)
- Freitag, U., 1993, Map functions, in *The Selected Main Theoretical Issues Facing Cartography: report of the ICA-Working Group to Define the Main Theoretical Issues on Cartography*, Kanakubo, T. (Ed)
- Friedmannová, L., Konečný, M., Staněk, K, 2006, An adaptive cartographic visualisation for support of the crisis management. In *proceedings of Auto-Carto 2006*, Vancouver.
- Gardarin, G., 2000, *Internet/intranet et bases de données*, Eyrolles, Paris
- Gartner, G., Bennett, D., A., Morita, T., 2007, *Towards Ubiquitous Cartography, Cartography and Geographic Information Science* (special content issue), Vol 34 (4)
- Gartner G., Peterson, M., P., 2007, Experiences from hands-on webmapping workshops of the ICA commission on map and the internet, in *proceedings of the 23^r ICA conference*, Moscow

- Gesbert N., 2005, Etude de la formalisation des spécifications de bases de données géographiques en vue de leur intégration, Thèse de doctorat en informatique de l'Université de Marne-La-Vallée
- Gibson J., 1979, *Ecological Approach to Visual Perception*, published by Lawrence Erlbaum Associates
- Giff, G., A., Crompvoets, J., 2008, Performance indicators a tool to support Spatial Data Infrastructure assessment, *CEUS* Vol 32 (5)
- Göder, G., 2003, *Représentation comparée de schémas et spécifications de contenu*, Rapport de stage du DESS cartographie et SIG, ENSG - Paris I
- Goodchild, M., 2007, Citizens as sensors: the world of volunteered geography , *GeoJournal*, Vol. 69, No. 4
- Goodwin, J., Dolbear, C., Hart, G., 2009, Geographical Linked Data: the Administrative Geography of Great Britain on the Semantic Web, *Transactions in GIS*, Vol 12 (1)
- Harrie, L., Sarjakoski, L. T. and L. Lehto, 2002, A variable-scale map for small-display cartography. In *proceedings of the Joint International Symposium on "GeoSpatial Theory, Processing and Applications"* (ISPRS/Commission IV, SDH2002), Ottawa, Canada
- Harrower, M. and C. Brewer, 2003, ColorBrewer: An online tool for selecting color schemes for maps. *The Cartographic Journal* 40(1) pp27-37.
- Harvey, F., Tulloch, D., 2006, Local-government data sharing: evaluating the foundations of spatial data infrastructures, in *International Journal of Geographical Information Science*, Vol 20 (7)
- Heinzle, F., Kopczynski, M., Sester, M., 2003, Spatial Data Interpretation for the Intelligent Access to Spatial Information in the Internet, in *'Proceedings of 21st International Cartographic Conference*, Durban, South Africa
- Hirose, M., Hiramoto, R., Sumiya, K, 2007, GeminiMap - Geographical Enhanced Map Interface for Navigation on the Internet, in *proceedings of the 7th Web and Wireless GIS Symposium*, Ware, J.M., Taylor, G., E. (Eds), Cardiff, pp279-292
- Hoarau, C., 2009, *Cartographie sur Dispositif Mobile*, rapport de stage de mastère CarthaGéo de l'Ecole Nationale des Sciences Géographiques et de l'Université Paris I, Saint Mandé
- Horrocks, I., 2008, Ontologies and the Semantic Web, in *proceedings of the ACM conference*, Vol 51 (12)
- Hugentobler, M., Iosifescu-Enescu, I., 2009, Cartographic web services with open standards and open software, in *proceedings of the 1st OpenSource Geospatial Research Symposium (OGRS)*, Springer Lecture Notes in Geoinformation and Cartography, Bocher, E., Ertz, O. (Eds), (to be published)
- Jakobsson, A., 2002, *User requirements for mobile topographic maps*, *GiMoDig deliverable IST-2000-30090 D2.1.1*
- Jankowski, P., Nyerges, T., 2001, *Geographic Information Systems for Group Decision Making, towards a participatory, geographic information science*, Taylor & Francis Publishers, UK

- Jenny, B., Jenny, H., Råber, S., 2008, Map design for the Internet, *International perspectives on maps and the internet*, Peterson, M., P., (Ed), Lecture Notes in Geoinformation and Cartography, Springer
- Joho, H., Sanderson, M., 2004, The SPIRIT Collection: an overview of a large web collection. *SIGIR Forum*, Vol 38 (2)
- Jolivet, L., Bucher, B., 2008, *Implémentation de services Web d'enrichissement de données pour TerraNumerica*, livrable WP3.1.3.2 – D2, affaire Terradata-1
- Jones, C.B., Abdelmoty, A.I., Finch, D., Fu, G. & Vaid, S. 2004. The SPIRIT Spatial Search Engine: Architecture, Ontologies and Spatial Indexing. In *Proceedings of the 3rd International Conference on Geographic Information Science (GIScience 2004)*, Maryland, USA. LNCS
- Kavouras, M., Kokla, M., 2008, *Theories of geographic concepts – Ontological Approaches to Semantic Integration*, CRC Press
- Kent, A., J., 2009, Cartographic landscapes and the new noise: finding the good view in a topographical mashup, *The Bulletin of the Society of Cartographers*, Vol 42
- Kraak, M.J., 2006, Where do maps lead us? keynote presentation at Visual Analytics & Spatial Decision Support Workshop of the GIScience conference, Muenster
- Kurokawa, C., Ota, M., 2007, Portrayal schema design and mechanism for the map personalization, in *proceedings of the 23rd ICC conference*, Moscow
- Illert, A., Afflerbach, S., 2003, *Report on Global Schema*, GiMoDig IST-2000-30090, Deliverable D5.2.1
- Jolivet L., Buard E., Bucher B., Ruas A., 2007, Amélioration de légende sur le Web, *actes de la conférence SAGEO'07*, Clermont-Ferrand, France
- Krygiel, J., 1995, Cartography as an art and a science? *The Cartographic Journal* Vol 32 N°6
- Larson, R.R. 1996. Geographic Information Retrieval and Spatial Browsing. In *GIS and Libraries: Patrons, Maps and Spatial Information*, Linda Smith and Myke Gluck, Eds., University of Illinois
- Laucius, S., Bouju, A., Bertrand, F., 2006, Chargement et visualisation dynamiques de données géoréférencées pour un utilisateur mobile, *actes de la conférence SAGEO*, Strasbourg
- Lehto, L., 2003, *Architecture Specification*, GiMoDig IST-2000-30090 deliverable D4.4.1
- Lemmens, R., 2006, *Semantic interoperability in distributed geo-service*, PhD thesis, ITC, Enschede
- Lilburne, L. R., Webb, T.H., Benwell G.L., 2004, The Scale Matcher : a procedure for assessing spatial data and models, in *International Journal of Geographical Information Science*, vol.18, 3
- Lin, Y., Libourel, T., Mougenot, I., 2009, Aide à la décision pour les domaines environnementaux, mise en place de chaînes de traitements pour la tracabilité et le diagnostic, *actes des journées DOCTISS 2009*
- Lopez-Velasco, C., Gensel, J., Villanova-Oliver, M., Martin, H., 2009, Vers une plateforme de génération de SIG mobiles adaptés au contexte d'utilisation, revue internationale de géomatique, *SIG mobiles*, Martin, H., Badard, T., (Eds), Paris

- Lüscher, P., Weibel, R., Mackaness, W., A., 2008, Where is the Terraced House? On the Use of Ontologies for Recognition of Urban Concepts in Cartographic Databases, in *Headway in Spatial Data Handling, proceedings of the SDH Conference*, Ruas, A., Gold, C., (Eds), Lecture Notes in Geoinformation and Cartography
- Mackaness, W., A., Beard, M., K., Buttenfield, B., P., 1994, *Selected Annotated Bibliography on Visualization of the Quality of Spatial Information*, Research Initiative 7, NCGIA report 94-10
- Mackaness, W., A., 2007, Understanding geographic space, in *Generalisation of Geographic Information: Cartographic Modelling and Applications*, Mackaness and Ruas (eds), Elsevier
- McMaster, R. B., Shea, S., 1992, Generalization in digital cartography, *Association of American Geographers publications*, Washington D.C
- Masser, I., All shapes and sizes: the first generation of national spatial data infrastructures. *International Journal of Geographical Information Science*, Vol 13, pp. 67-84, 1999
- Mennis, J. L., Peuquet, D. J., Qian, L., 2000, A conceptual framework for incorporating cognitive principles into geographical database representation, in *the International Journal of Geographical Information Science*, Vol 14 (6)
- Michaux, P., 2003, *Création d'une interface de consultation pour les traitements cartographiques*, rapport de stage de DESS imagerie électronique, Saint Mandé
- Moellering, H., 2007, Expanding the ICA conceptual definition of a map, in *proceedings of the 23rd International Cartographic Conference*, Moscow
- Moisuc, B., 2007, *Conception et Mise en Œuvre de Systèmes d'Information Spatio-Temporelle Adaptatifs : le framework ASTIS*, thèse en informatique de l'Université Joseph Fourier
- Moisuc, B., Gensel, J., Davoine, P.-A., Martin, H., 2006, Designing Adaptive Spatio-Temporal Information Systems for Natural Hazard Risks with ASTIS, in *proceedings of the 6th International Symposium on Web & Wireless Geographic Information Systems (W2GIS'2006)*, Hong Kong, China
- Morita, T., 2005, A working conceptual framework of ubiquitous mapping, in *proceedings of the XXIInd International Cartographic Conference*, A Coruna, Espagne
- Morita, T., 2007, Concept of a real scale map and the allocation of reference points in ubiquitous mapping, in *proceedings of the 23rd ICC conference*, Moscow
- Neutens, T., Schwanen, T., Witlox, F., De Maeyer, P., 2008, My Space or your space? Towards a measure of joint accessibility, *CEUS* Vol 32 (5)
- Nivala, A.-M., Brewster, S.A., Sarjakoski, L.T., 2007, Usability Problems of Web Map Sites. in *proceedings of the XXIII International Cartographic Conference, Cartography for everyone and for you*, Moscow, Russia
- Noucher, M., 2009, Coproduire sans appauvrir, enjeux et dispositions pour cartographier à plusieurs les territoires, *actes des 1^{ères} rencontres SIG La Lettre*, Marne La Vallée
- Open Geospatial Consortium (OGC), 2007, Styled Layer Descriptor profile of the Web Map Service Implementation Specification, v1.1, OGC® Implementation Specification, Lupp, M. (Ed), 2007

- Open Geospatial Consortium (OGC), 2009, Geographic Linkage Service (GLS) Specification, v0.12, OGC® Draft Implementation Specification, Schut, P. (Ed)
- Ormeling, F., Kraak, M.-J., 2008, Maps as predictive tools: Mind the Gap, *Cartographica*, vol 43(2)
- Ostreika, A., Fryml, J., Pospíšil, M., Charvat, K., Čerba, O., Loidold, M., Bernhardt, M., Moreno, P., 2007, *Report on Processes Analysis*, Humboldt project deliverable A2.1-D2
- Open Geospatial Consortium, 2007 a, OGC Web Services Initiative - Phase 5 (OWS-5), Annex B OWS-5 Architecture
- Open Geospatial Consortium, 2007 b, OpenGIS® Web Processing Service, OpenGIS® standard
- OWL Services Coalition, 2004, *OWL-S : Semantic Markup for Web Services*, Submission to W3C
- Özsu, T., Valduriez, P., 1999, *Principles of distributed database systems*, Prentice Hall, New Jersey, US
- Page, M., Gensel, J., Capponi, C., Bruley, C., Genoud, P., Ziébelin, D. 2000. Représentation de connaissances au moyen de classes et d'associations : le système AROM, *actes de la conférence LMO*
- Paolino, L., Sebillio, M., Tortora, ., Vitiello, G., 2007, Framy- Visualizing Spatial Query Results on Mobile Interfaces, in *proceedings of the 7th Web and Wireless GIS Symposium*, Ware, J.M., Taylor, G., E. (Eds), Cardiff, pp175-186
- Petit, M., Ray, C., Claramunt, C. 2007, A user context approach for adaptive and distributed GIS. in *Proceedings of the 10th International Conference on Geographic Information Science: AGILE'07*, Wachowicz, M., Fabrikant, S. (eds.), Springer-Verlag LNG&C, Aalborg, Denmark
- Peterson, M., (Ed) 2008, *International Perspectives on Maps and The Internet*, Lecture Notes in Geoinformation and Cartography, Springer
- Philippe, P., 2006, *Définition d'un jeu d'échantillons cartographiques qualifiés*, rapport de stage de mastère Carthagéo Pro de l'Ecole Nationale des Sciences de l'information Géographique, Saint Mandé
- Purves, R.S., Yang, B., 2005, Graphical query and presentation interface, SPIRIT deliverable D26 401
- Purves, R.S., Clough, P. and Joho, H. 2005. Identifying imprecise regions for geographic information retrieval using the web. In *Proceedings of GISRUK*
- Purves R.S., Clough P., Jones C.B., Arampatzis A., Bucher B., Finch D., Fu G., Joho H., Khirini A.S., Vaid S., Yang B., 2007, The design and implementation of SPIRIT: a spatially-aware search engine for information retrieval on the Internet, *International Journal Geographic Information Systems (IJGIS)*, Vol21(7)
- Raper, J., 1996, Unsolved problems of spatial representation, in *proceedings of SDH'96*
- Renard, J., 2008, *Caractérisation et analyse des légendes des cartes topographiques européennes*, Rapport de stage de mastère Carthagéo de l'Université Paris 1, Paris 7 et de l'ENSG, Saint Mandé

- Rinner, C., Keßler, C., Andrulis, S., 2008, The Use of Web 2.0 Concepts to Support Deliberation in Spatial Decision-Making, *Computers, Environment and Urban Systems* (CEUS) volume 32, issue 5, pp. 386-395
- Ruas, A., 2004, *Le changement de niveau de détail dans la représentation de l'information géographique*, Habilitation à diriger des recherches, Université de Marne-la-Vallée
- Ruas, A., Plazanet, C., 1996, Strategies for automated generalization. In *Advances in GIS Research II*, M.J. Kraak and M. Molenaar (eds.), *Proceedings of 7th International Symposium on Spatial Data Handling*, London: Taylor & Francis
- Sanderson, M., Joho, H., 2003, *Test collection formation methods*, SPIRIT deliverable D11 2102
- Sawahata, L., 2001. *Color Harmony Workbook, a Workbook and Guide to Creative Color Combinations*, Gloucester, Massachusetts: Rockport Publishers
- Schreiber, A. Th. , Akkermans, J. M. , Anjewierden, A. A., de Hoog, R., Shadbolt, N. R., Van de Velde, W., Wielinga, B. J., 2000, *Knowledge Engineering and Management, The CommonKADS Methodology*, MIT Press
- Singh, R., 2004, Geo Blogging: collaborative, peer-to-peer geographic information sharing, in *proceedings of URISA Public Participation in GIS 3rd Annual Conference*
- Sykora P., Schnabel O., Iosifescu Enescu I., Hurni L., 2007, Extended Cartographic Interfaces for Open Distributed Processing, in *Cartographica*, Vol 42 (3)
- Stanek, M. Konecny, L. Friedmannova, 2007, An Adaptive Cartographic Visualisation for Support of the Crisis Management, in *proceedings of the 23rd International Cartographic Conference*, Moscow, Russia
- Staub, P., Gnägi, H., R., Morf, A., 2008, Semantic interoperability through the definition of conceptual model transformations, *Transactions in GIS*, Vol 12 (2)
- Takahashi N., 2008, An elastic map system with cognitive map-based operations, *International perspectives on maps and the internet*, Peterson, M., P., (Ed), *Lecture Notes in Geoinformation and Cartography*, Springer
- Timpf S., Volta G., Pollock D., Egenhofer M., 1992, A conceptual model of wayfinding using multiple levels of abstraction, *Theories and Methods of Spatio-temporal Reasoning in Geographic Space*, A. Frank, I. Campari, U. Formentini (eds), Springer-Verlag, Italia
- Tsou, M.-H., Curran, J. M., 2008, Chapter 20. User-Centered Design Approaches for Web Mapping Applications: A Case Study with USGS Hydrological Data in the United States, in *International Perspectives on Maps and the Internet*, Peterson, M. P. (Ed), Berlin: Springer, pp. 301-321
- Vatant, B., Bucher, B., 2007a, *Glossaire TerraNumerica*, Document 3.1.3.2_1bis, projet TerraData-1
- Vatant, B., Bucher, B., 2007b, *Etat de l'art: langages, normes et services*, Document 3.1.3.2_1, projet TerraData-1
- Viana, W., Gensel, J., Andrade, R., Villanova-Oliver, M., Martin, H., 2006. XMobile : un environnement pour la génération d'interfaces adaptées aux dispositifs mobiles, *actes de la conférence INFORSID 2006*, Hammamet, Tunisie

- Visvalingam, M., 1989, Cartography, GIS and Maps in Perspective, *Cartographic Journal*, Vol 26, N°1
- W3C Working Group, 2007, Semantic Annotations for WSDL and XML Schema — Usage Guide, Note 28, Akkijaru, R., Sapkota, B. (Eds)
- Weger, G., 1999, *Cartographie, sémiologie graphique et conception cartographique*, cours de l'Ecole Nationale des Sciences Géographiques, Marne la Vallée
- Williamson, I., P., Rajabifard, A., Feeney, M.-H., 2003, *Developing Spatial Data Infrastructures: From Concept to Reality*, CRC Press
- Wood, J., Dykes, J., Slingsby, A., Clarke, K., 2007, Interactive Visual Exploration of a Large Spatio-temporal Dataset: Reflections on a Geovisualization Mashup, in *IEEE Transactions on Visualization and Computer Graphics*, Vol 13 (6)