



HAL
open science

Ingénierie de l'observation basée sur la prescription en EIAH

Sébastien Iksal

► **To cite this version:**

Sébastien Iksal. Ingénierie de l'observation basée sur la prescription en EIAH. Environnements Informatiques pour l'Apprentissage Humain. Université du Maine, 2012. tel-00991970

HAL Id: tel-00991970

<https://theses.hal.science/tel-00991970v1>

Submitted on 16 May 2014

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.



Année 2012

Habilitation à Diriger des Recherches en Informatique

présentée par

Sébastien IKSAL

Laboratoire d'Informatique de l'Université du Maine (LIUM)

52, rue des docteurs Calmette et Guérin – 53020 Laval Cedex 9

sebastien.iksal@univ-lemans.fr

Sujet

Ingénierie de l'observation basée sur la prescription en EIAH

*Soutenue le 18 septembre 2012, à l'Université du Maine
devant le jury composé de :*

Pierre	TCHOUNIKINE	Président	Université Joseph Fourier (Grenoble 1) / LIG
Angelique	DIMITRACOPOULOU	Rapporteur	Université d'Aegean (Grèce) / LTEEL
Jean-Marc	LABAT	Rapporteur	Université Pierre et Marie Curie / LIP6
Alain	MILLE	Rapporteur	Université Claude Bernard Lyon 1 / LIRIS
Christophe	CHOQUET	Examineur	Université du Maine / LIUM
Jean-Charles	MARTY	Examineur	Université de Savoie / LIRIS

Je souhaite remercier ici . . .

Les personnes qui m'ont fait l'honneur de participer au jury de cette thèse. Je suis très honoré de compter Angélique Dimitracopoulou, Jean-Marc Labat et Alain Mille, pour rapporteurs et je les remercie d'avoir accepté de juger mon travail et pour tous les échanges lors de l'école thématique Traces du CNRS, les ateliers, les projets et sur ce mémoire. Je remercie Jean-Charles Marty et Pierre Tchounikine pour avoir accepté d'examiner mon travail, pour les discussions que nous avons eues et les conseils qu'ils m'ont donnés. Je remercie Christophe Choquet qui m'a accepté dans son équipe, avec qui j'ai collaboré sur les aspects observation du projet de recherche, qui m'a encouragé dans la voie de l'habilitation, pour les relectures et pour avoir accepté de faire partie de mon jury.

Les collègues enseignants et administratifs de l'IUT de Laval qui ont toujours été coopératifs et conciliants au niveau organisationnel de la vie en IUT.

Mes collègues permanents de l'équipe IEIAH du LIUM, les doctorants que j'ai encadré ou non : Hassina, Aïna, Amine, Aymen, Boubekour, Firas, Jean-Pierre, et notamment Diem, Noa et Mohand que j'ai co-encadré ou que je co-encadre encore.

Toutes les personnes avec qui j'ai collaboré au cours de toutes ces années de recherche, que ce soit au LIUM, à l'ENST-Bretagne ou au travers de séminaires ou ateliers. La liste serait trop longue à faire, mais elles ont toutes contribué d'une manière ou d'une autre à m'amener au point où j'en suis aujourd'hui.

Une personne qui compte beaucoup pour moi, qui m'a montré la voie de la recherche, qui m'a formé, conseillé et qui est toujours à mes côtés. Le maître de stage, puis enseignant, encadrant de thèse et maintenant ami : Serge Garlatti.

Je tiens à remercier tout particulièrement ma famille, mes parents et mes beaux-parents, pour leur écoute, leur présence et leur soutien.

Enfin, je tiens à remercier Florence, mon épouse, qui a toujours fait preuve de patience, d'écoute et d'encouragements dans mes avancées et tout particulièrement dans la rédaction de ce mémoire. J'ai une pensée aussi pour ma fille Albane qui a supporté mon indisponibilité ces derniers mois et qui voulait comme Papa "écrire des lettres sur l'ordinateur".

Table des matières

I	TRAVAUX DE RECHERCHE	12
1	Introduction	14
1.1	Questions de recherche et problématique	15
1.2	Préambule au cheminement scientifique	17
1.3	Cheminement scientifique	19
1.4	Plan du mémoire	24
2	L’observation : une analyse de la situation	26
2.1	Terminologie	27
2.1.1	La Trace	27
2.1.2	L’indicateur	28
2.1.3	Les autres termes	29
2.2	Le besoin d’observer	30
2.3	La collecte des informations	33
2.3.1	La création ad hoc de mécanismes de collecte	33
2.3.2	La modélisation des données	33
2.3.3	Position sur la collecte	34
2.4	L’analyse des données	35
2.4.1	Fouiller les données : Faire émerger l’observé	36
2.4.2	Les approches par l’hypothèse	39
2.4.3	Positionnement sur l’analyse	41
2.5	L’interprétation de l’analyse	42
2.5.1	La visualisation au service de la compréhension	42
2.5.2	La modélisation d’apprenants	43
2.5.3	Positionnement sur l’interprétation	45
2.6	La capitalisation	45
2.6.1	DataShop	46
2.6.2	MULCE	46
2.6.3	DPULS	47
2.6.4	Positionnement sur la capitalisation	48
3	Etude préparatoire : La recherche de généricité	50
3.1	Introduction	51
3.2	Etude préparatoire sur les traces	51
3.3	Un modèle sémantique pour l’information contenue dans la trace	52

3.3.1	Les objectifs du modèle	52
3.3.2	Présentation du modèle	53
3.4	Principe de fonctionnement	57
3.5	Mise en œuvre	58
3.6	Synthèse	62
4	L'ingénierie de l'observation par la prescription	64
4.1	Introduction	65
4.2	Présentation de notre approche	66
4.3	La chaîne éditoriale de l'observation prescrite	67
4.3.1	Les rôles concernés	67
4.3.2	Description de la chaîne éditoriale	69
4.4	Un enjeu important : La capitalisation	71
4.5	Les objectifs d'observation	73
4.6	Scénarios d'utilisation	73
4.6.1	Un exemple en réingénierie	74
4.6.2	Un exemple en régulation	77
4.7	Synthèse	78
5	Méthodologie de spécification des données d'observation	80
5.1	Introduction	81
5.2	Représenter les données d'observation : Le modèle DGU	81
5.3	Spécification déclarative des données d'observation	83
5.3.1	Le modèle conceptuel	84
5.3.2	Les modèles d'information	86
5.3.3	De la spécification à l'opérationnalisation : DCL4UTL	93
5.4	Synthèse	96
6	Instrumentation et mise en œuvre	98
6.1	Introduction	99
6.2	Mise en œuvre du processus d'ingénierie	99
6.2.1	Principe général	99
6.2.2	L'éditeur de données	99
6.2.3	L'interpréteur UTL	100
6.2.4	La réutilisation des données	101
6.3	Un exemple concret : Les indicateurs dans Hop3X	102
6.3.1	Présentation de l'environnement	102
6.3.2	La modélisation d'indicateurs	105
6.4	Synthèse	110
7	Conclusion et perspectives	112
7.1	Synthèse des activités menées	113
7.2	Partager et réutiliser au sein de la communauté EIAH	114
7.3	Perspectives à moyen terme	116

7.3.1	L'expression du besoin d'observation dans l'univers métier du prescripteur	117
7.3.2	Le retour d'information ou comment percevoir les indicateurs ? . . .	117
7.3.3	La définition de nouveaux indicateurs en session	117
7.3.4	Vers une composition encore plus sémantique	118

Table des figures

2.1	Représentation de notre terminologie	30
2.2	Cycle de vie des données de traçage en EIAH [May 2009]	32
2.3	Algorithme des k-moyennes avec 5 partitions [Merceron et Yacef 2005] . . .	37
2.4	Fonctionnement général d'un SBT [Settouti 2011]	39
2.5	Exemples de transformations avec SBT [Settouti 2011]	39
2.6	Exemples de visualisation d'indicateurs avec TrAVis [May et al. 2008] . . .	43
2.7	Vue de la classe avec les faces de Chernoff [France et al. 2007]	44
2.8	Représentation avec des zones d'ombre [Heraud et al. 2005]	44
2.9	Parcours types pour IDLS [Bousbia et al. 2010]	44
2.10	Typologie des données du projet DPULS	47
3.1	Le modèle d'information d'un concept traçable	55
3.2	Le modèle d'information d'une trace	56
3.3	Le processus d'utilisation d'UTL 1	58
3.4	Extrait de la modélisation des traces provenant de FSL	59
3.5	Extrait de traces provenant de FSL	60
3.6	Trace FSL d'arrêt de lecture de la vidéo	60
3.7	Interface de consultation des indicateurs	61
4.1	Processus d'ingénierie de l'observation	66
4.2	Chaîne éditoriale de l'observation	69
4.3	Modèle de tâches de la réingénierie	74
4.4	Modèle de scénario et son instance	75
4.5	Réingénierie du modèle de scénario et de son instance	76
4.6	Modèle de tâches de la régulation	77
5.1	Le modèle DGU	82
5.2	Modèle conceptuel du langage UTL	85
5.3	Le modèle d'information d'une donnée brute	87
5.4	Le modèle d'information d'une donnée de production	89
5.5	Le modèle d'information de la donnée additionnelle	90
5.6	Le modèle d'information de la donnée intermédiaire	91
5.7	Le modèle d'information d'un indicateur	92
5.8	Le modèle d'information d'un objectif d'observation	93
5.9	Modélisation du langage DCL4UTL	94

5.10	Exemple de description avec le langage DCL4UTL	95
6.1	Copie d'écran de l'éditeur Edit'UTL sur la partie obtention de l'indicateur	100
6.2	Architecture générale de l'interpréteur UTL [Phâm Thi Ngoc 2011]	101
6.3	Interface Hop3X-Student	103
6.4	Interface Hop3X-Teacher avec les indicateurs venant d'UTL	104

Liste des tableaux

6.1	Section Defining de l'indicateur "Calcul du taux de correction des erreurs à la compilation"	107
6.2	Section Getting de l'indicateur "Calcul du taux de correction des erreurs à la compilation"	108
6.3	Section Using de l'indicateur "Calcul du taux de correction des erreurs à la compilation"	109
6.4	Formule de calcul de l'indicateur "Calcul du taux de correction des erreurs à la compilation"	109

Première partie

TRAVAUX DE RECHERCHE

Chapitre 1

Introduction

Sommaire

1.1	Questions de recherche et problématique	15
1.2	Préambule au cheminement scientifique	17
1.3	Cheminement scientifique	19
1.4	Plan du mémoire	24

Ce document présente les activités de recherche que nous avons menées depuis plusieurs années au sein du LIUM – Laboratoire d’Informatique de l’Université du Maine – dans le domaine des EIAH – Environnements Informatiques pour l’Apprentissage Humain – [Tchounikine 2004] et plus particulièrement dans le domaine de l’ingénierie des EIAH.

Notre intégration au LIUM correspond à la période de restructuration des chercheurs en ingénierie des EIAH. Cette nouvelle structure s’est fédérée autour d’un projet de recherche sous la responsabilité de Christophe Choquet. Le projet REDiM – Réingénierie des EIAH Dirigée par les Modèles – avait pour objectif principal d’intégrer les enseignants et les formateurs de terrain dans le processus de conception d’un EIAH. Aujourd’hui ce projet est terminé et les chercheurs en EIAH du LIUM se sont regroupés au sein de l’équipe IEIAH (Ingénierie des EIAH). Dans cette équipe, nous considérons des situations d’ingénierie et de réingénierie d’un EIAH dans lesquelles le concepteur est un enseignant (milieu académique) ou un formateur (milieu professionnel) et non un chercheur en EIAH. L’approche choisie repose sur le principe de conception continue, supportée par un processus itératif alternant les phases d’analyse des usages et de réingénierie de l’EIAH, dans une approche dirigée par les modèles. Les travaux menés dans ce cadre se sont focalisés sur l’étude des modèles prescriptifs¹, voire prédictifs² qui peuvent être exploités dans la génération des EIAH ainsi que sur l’analyse des usages en EIAH de façon à reconstruire des modèles descriptifs³ de la situation pédagogique. Le résultat de cette analyse devant être, à terme, directement manipulable et interprétable par les concepteurs (enseignants ou formateurs). Au sein de cette équipe, nous avons développé l’axe de recherche sur “la modélisation et l’analyse de traces d’activité dans un EIAH”, les travaux présentés dans ce mémoire présentent nos contributions dans ce domaine.

1.1 Questions de recherche et problématique

Que ce soit dans le domaine de l’adaptation⁴ ou dans le domaine des EIAH, il est nécessaire d’observer la situation et d’en tirer des conclusions sur son déroulement effectif. Les environnements sont capables de laisser des traces, encore faut-il être à même de les exploiter. Le concept de trace est un concept central dans notre problématisation, toutefois il n’existe pas une unique définition de la trace. En effet, certains considèrent la trace comme un parcours cognitif⁵ identifié par une séquence d’évènements observés (cf. le projet TRAILS [TRAILS 2004]), la majorité des autres acteurs de l’observation informatisée considèrent la trace dans une granularité plus fine (un clic de souris, un fichier, un message). Notre point de vue sur la trace se situe plutôt dans la deuxième

1. Le modèle *prescriptif* décrit en amont de la session d’apprentissage la façon dont celle-ci *doit* se dérouler.

2. Dans certains cas, il est préférable d’utiliser le qualificatif *prédictif* plutôt que prescriptif car lorsque le scénario ou l’environnement offre de la souplesse aux usagers, leur comportement peut dévier du scénario prévu, la modélisation représente alors la façon dont la session d’apprentissage *devrait* se dérouler.

3. Le modèle *descriptif* représente le déroulement effectif, c’est-à-dire la façon dont chaque usager a *réellement* réalisé sa situation d’apprentissage.

4. Ce domaine de recherche correspond aux travaux que nous avons menés en thèse.

5. La notion de parcours cognitif utilisée dans ce projet fait référence au cheminement de l’apprenant dans le cadre de son apprentissage.

catégorie. Plusieurs projets européens se sont intéressés ces dernières années au processus d'observation des sessions d'apprentissage [ICALTS 2004][IA 2005][DPULS 2005a]. Ils ont permis de clarifier les définitions d'indicateurs, de patrons d'analyse etc.. L'indicateur dont une définition plus précise est donnée dans le chapitre suivant, est une donnée calculée qui est compréhensible par l'utilisateur et qui répond à un objectif particulier d'observation. Le patron d'analyse est une description capitalisable d'un procédé d'analyse comportant son contexte, son objectif ainsi que la manière dont il est mis en œuvre. Toutefois, dans ces projets, le problème de l'opérationnalisation de l'analyse est rarement abordé, et la question de la réutilisation automatisée est absente. Le raison principale vient du fait que chaque EIAH possède sa propre solution de traçage avec des modèles internes, par conséquent les solutions d'analyse sont développées ad hoc et très difficilement exportables.

Notre objet de recherche est l'ingénierie de l'observation, c'est-à-dire un champ plus large que l'analyse de traces au sens où nous considérons le processus d'observation qui va de l'agent humain exprimant son besoin d'observation jusqu'à l'agent humain (le plus souvent, il s'agit de la même personne) qui exploite le résultat de l'observation. Notre problématique correspond alors à la définition et à l'opérationnalisation du processus d'ingénierie de l'observation en EIAH afin d'aider à l'adaptation et à la réingénierie d'une activité d'apprentissage. Nous partageons le même paradigme que le projet REDiM, à savoir l'utilisation d'EIAH scénarisés. Nous considérons comme une nécessité la prise en compte des acteurs dès le processus de conception de l'observation, c'est pour cela que nous proposons une approche de l'observation par la prescription. Ce qui nous amène à envisager l'expression des besoins d'observation comme un scénario prescriptif d'observation par analogie au scénario prescriptif d'apprentissage.

La question générale sur laquelle repose notre problématique est la suivante : **“Comment aider l'enseignant/concepteur à améliorer sa situation d'apprentissage grâce à l'observation de cette dernière ?”**. Nous adoptons comme premier postulat que l'observation est nécessaire pour l'amélioration d'une situation d'apprentissage. En effet, l'enseignant/concepteur a besoin d'informations sur le déroulement de sa situation afin de savoir où, quand, et comment il doit réagir pour l'améliorer. Le second postulat que nous considérons est que l'observation menée sur une situation d'apprentissage servira soit pour faire évoluer cette même situation dans une nouvelle itération, soit pour être prise en compte dans d'autres situations d'apprentissage, en particulier avec le même enseignant/concepteur. Cela nous amène plusieurs questions auxquelles nous nous intéressons dans le cadre de nos travaux.

La première question est **“Comment aider l'utilisateur à mieux percevoir et comprendre le déroulement de sa session d'apprentissage ?” (Q1)**. L'utilisateur qu'il soit enseignant/concepteur, enseignant, tuteur ou encore apprenant, doit disposer d'informations précises sur son activité ou sur la session d'apprentissage s'il souhaite faire évoluer ces dernières. Afin d'être au plus près des attentes de l'utilisateur, nous avons pris comme hypothèse que l'utilisateur est capable d'exprimer ses besoins d'observation. Il peut vouloir, par exemple, détecter les ressources inutilisées, ou mal utilisées et chercher à en identifier les raisons. Le besoin d'observation correspond donc à une description de ce que l'utilisateur souhaite observer, l'objectif correspondant ainsi que la façon dont il souhaite que l'information lui soit restituée. Cette prescription de l'observation implique alors l'utilisateur

directement dans l'ingénierie de l'observation.

C'est pour cela que nous nous sommes intéressés à la question : **“Comment formaliser le besoin d'observation et le calcul de l'indicateur correspondant afin de le rendre automatisable, réutilisable et adaptable ?” (Q2)**. En effet, le plus souvent les indicateurs sont conçus pour un usage unique (une plate-forme, un type de scénario, etc.). Il est souhaitable aujourd'hui de favoriser la réutilisation en proposant des indicateurs qui peuvent s'instancier dans un maximum de situations. Il est de moins en moins concevable de devoir “redévelopper” des outils à chaque fois que l'on souhaite bénéficier d'une analyse.

Ce qui nous amène à la prise en compte de la connaissance métier : **“Comment prendre en compte la connaissance métier dans des indicateurs basés sur des traces spécialisées ?” (Q3)**. Si l'on regarde la façon dont la plupart des indicateurs sont élaborés, ils partent des traces générées par les environnements et transforment directement ces traces afin de construire une information qui a du sens pour l'utilisateur. Cela signifie qu'ils partent du métier de la plate-forme pour arriver au métier de l'utilisateur. Lorsque l'environnement est remplacé, il est nécessaire de retranscrire ce savoir-faire sur le nouveau format de traces. Il serait souhaitable de pouvoir exprimer le savoir-faire indépendamment des formats de trace car la connaissance métier sur la manière de produire un indicateur n'est pas liée à ce format, elle est liée à la sémantique de ce que l'on observe et donc à l'utilisateur.

La question suivante se focalise plus sur l'utilisation de l'observation, en effet, **“Comment permettre à l'utilisateur de concevoir, modifier, adapter au mieux l'EIAH, éventuellement en cours de session (processus de conception continue) ?” (Q4)**. L'idée ici est de permettre aux utilisateurs d'agir sur le déroulement de la situation durant la session d'apprentissage. L'utilisation des indicateurs peut les amener à avoir une réflexion sur leur propre activité et si l'environnement le permet, ils peuvent modifier le scénario prédictif.

Enfin, **“Comment exploiter ce savoir-faire capitalisé pour améliorer la session ?” (Q5)**. L'observation est utile à différents niveaux, que ce soit pour évaluer, pour adapter ou encore pour l'évolution des outils. Toute cette connaissance autour de l'observation, ces indicateurs construits devraient être mis à disposition et exploitables par la communauté de pratique. En effet, pour un enseignant/concepteur, pouvoir bénéficier de l'expérience de ses confrères serait un avantage non négligeable dans la construction d'une situation d'apprentissage ou dans son évolution, par exemple, en prenant en compte des problèmes non rencontrés et auxquels il n'aurait pas pensé.

1.2 Préambule au cheminement scientifique

Ce préambule a pour objectif de présenter les activités que nous avons menées dans le cadre de la thèse car elles permettent de comprendre certains choix que nous avons faits, et particulièrement notre intérêt pour la modélisation, la déclarativité des modèles, la composition de données et la réutilisation. Ensuite, nous présentons notre cheminement scientifique permettant de répondre aux différentes questions énoncées précédemment. Il s'agit d'une méthodologie de type recherche-action, pendant laquelle nous avons cherché

à progresser par étapes afin de bâtir et d'améliorer le processus global de l'ingénierie de l'observation par la prescription.

L'intérêt que nous portons à l'observation trouve son origine dans nos premiers travaux de recherche. Notre activité s'est focalisée au début sur la conception de systèmes d'information adaptatifs et sémantiques. En effet, Internet est devenu une banque d'information gigantesque et aujourd'hui nous avons besoin de systèmes d'information capables de traiter des ressources hétérogènes et disséminées sur de nombreux serveurs. La capitalisation et la réutilisation des ressources sont devenues des enjeux, de même que l'adaptation/personnalisation des systèmes en fonction des caractéristiques (connaissances, besoins, compétences, objectifs, etc.) des utilisateurs. La collaboration avec Serge Garlatti (Télécom Bretagne) autour du projet SWAN (Serveur Web Adaptatif et Naviguant) [Garlatti et Iksal 2000] nous a permis de mettre en place et d'éprouver des méthodes et techniques d'adaptation au sens de Brusilovsky [Brusilovsky 1996]. Dans ce projet, nous avons modélisé des informations maritimes puis nous les avons organisées selon les connaissances de l'utilisateur, selon les caractéristiques techniques de son navire et en fonction de sa localisation géographique. Nous nous sommes donc intéressés à la représentation de l'utilisateur sous une forme exploitable informatiquement. En effet, la définition de règles (ou stratégies) d'adaptation/personnalisation passe nécessairement par une description détaillée des caractéristiques de l'utilisateur ainsi que par une identification des possibilités d'adaptation de l'environnement.

Ensuite, nous avons poursuivi ces travaux dans le cadre des projets ICCARS (Integrated and Cooperative Computer Assisted Reporting System) [Iksal et Garlatti 2002] et CANDLE (Collaborative and Network Distributed Learning Environment). L'idée commune de ces projets reposait sur le fait qu'il fallait indexer sémantiquement les ressources (articles de presse ou éléments pédagogiques) de façon à composer dynamiquement un document en fonction des attentes et des connaissances de l'utilisateur. Cette fois-ci, le système devait non seulement posséder une description de l'utilisateur mais aussi être capable d'interpréter ces connaissances afin d'adapter le contenu et la présentation du document. Nous avons proposé une méthodologie de conception fondée sur un moteur de composition flexible appelé SCARCE – SemantiC & Adaptive Retrieval and Composition Environment [Garlatti et al. 2004]. Dans ce contexte, la cohérence du document produit est assurée par le savoir-faire des auteurs qui spécifient le contenu, l'adaptation et l'organisation du document à produire à l'aide de leurs connaissances communes. Ce moteur de composition flexible est fondé sur une spécification déclarative et sémantique du contenu basée sur des ontologies, une spécification de l'organisation, de l'adaptation et de la présentation ainsi qu'une gestion séparée de ces dernières.

Le constat principal que nous avons pu faire au travers de ces différents projets, réside dans le fait que le plus souvent la modélisation de l'utilisateur repose sur une description qui est faite a priori et par l'utilisateur lui-même. Ces techniques de modélisation permettent de disposer rapidement d'un maximum d'informations sur l'utilisateur, mais reposent sur les propos de l'intéressé, or nous n'avons aucun moyen de vérifier s'il indique la vérité ou s'il la détourne (volontairement ou non). En ce qui concerne les projets CANDLE et ICCARS, nous n'avons pas pu aborder la question de l'évaluation de l'usage des outils (cours ou dossiers thématiques) ou des stratégies d'adaptation mises en œuvre. Comment

savoir ce qui a été exploité dans le cours ? Comment évaluer en temps réel l'effectivité de l'adaptation ?

1.3 Cheminement scientifique

Notre participation au projet CANDLE ainsi que notre recrutement au LIUM ont constitué notre entrée dans le champ des EIAH, en particulier, cela nous a permis dans un premier temps de participer aux actions spécifiques du département STIC⁶ du CNRS : “Web Sémantique et E-Learning (WebLearn)” et “Conception d’une plate-forme technologique pour la recherche en EIAH (Plate-forme)” [Adam et al. 2005]. Ces participations ont confirmé notre volonté de développer l’axe portant sur “la modélisation et l’analyse de traces d’activité dans un EIAH”. Nous avons pu constater ces dernières années que l’analyse de traces est un domaine en pleine effervescence dans le champ des EIAH, en témoignent l’école thématique du CNRS en 2007, les journées d’atelier du GdR I3⁷, les différents projets du réseau d’excellence européen Kaleidoscope, ou encore les thèses et habilitations soutenues.

Dans le cadre de notre intégration au projet REDiM, nous avons développé les travaux liés à l’axe observation de la ré-ingénierie des EIAH, dans le but d’aider à analyser l’activité des usagers selon des besoins d’observation définis par le concepteur en fonction de la modélisation prescriptive de sa situation d’apprentissage (i.e. le scénario d’apprentissage). Alors que les activités de Vincent Barré au LIUM se situaient au niveau du métier des concepteurs par l’analyse automatique du scénario de façon à mettre en évidence les éléments “observables”, nous nous sommes focalisés dans un premier temps sur la séparation entre les indicateurs et le format de représentation des traces informatiques [Iksal et al. 2004]. C’est par cette étape que nous avons abordé l’ingénierie de l’observation pour ensuite proposer un processus plus complet, de l’expression du besoin à l’exploitation par les usagers.

Notre démarche scientifique dans le domaine de l’ingénierie de l’observation a été envisagée afin de répondre aux questions de recherche énoncées précédemment. Elle se décompose en six actions :

1. Une recherche de généralité pour s’abstraire des formats de représentation spécifiques. Il s’agit de mener une réflexion autour de l’objet “trace”, en particulier sur la description sémantique de son contenu, afin que cette description puisse servir de fondement au processus d’observation.
2. L’élaboration d’un cadre méthodologique de l’ingénierie de l’observation par la prescription. Au travers de ce cadre, nous prenons en compte l’usager dès la phase de spécification de l’observation (le besoin d’observation). Nous cherchons à répondre aux questions : qu’est ce que l’usager a besoin d’observer ? Comment ? A quel moment ? Et pourquoi ?
3. La formalisation de l’observation à des fins de partage et de réutilisation. L’observation ne doit pas être une “boîte noire” si l’on veut en faciliter le partage et la

6. STIC : Sciences des Technologies de l’Information et de la Communication.

7. Groupement de Recherche Information-Interaction-Intelligence

réutilisation. De plus, les technologies informatiques évoluant sans cesse, un langage formel peut être traduit pour être utilisé dans différents environnements (Bases de données relationnelles, XML, bases de connaissances etc.), ce qui permet d'améliorer les calculs, et les temps de réponses ainsi qu'une adaptation rapide aux nouveaux environnements.

4. La prise en compte de la perception de l'observation par l'utilisateur. Afin de faciliter l'appropriation de l'observation, il est nécessaire que l'utilisateur y trouve son intérêt. La prescription lui permet d'être impliqué dans la conception. La perception concerne l'usage qu'il fera de cette observation. Cette action a donc pour objet d'aider les usagers à mieux percevoir le résultat de l'observation au travers de la spécification et de l'instrumentation d'une stratégie prescrite.
5. L'exploitation de l'observation au travers de l'adaptation, de l'ingénierie et de la ré-ingénierie. L'observation concerne le plus souvent la compréhension d'une situation, nous souhaitons poursuivre l'étude jusqu'à la prise en compte de ces résultats dans la conception d'une nouvelle situation d'apprentissage, son amélioration ou sur l'adaptation de la situation en cours de session.
6. Faciliter l'émergence d'une communauté d'utilisateurs autour de l'observation. Dès lors que nous disposons d'une représentation formelle de l'observation calculable automatiquement et d'une description sémantique des traces de différentes plate-forme, il est intéressant d'ouvrir ce dépôt d'indicateurs et de permettre aux usagers de s'en servir, de les critiquer et de les améliorer. Le but de cette action est de fédérer les usagers autour de cette objet qu'est l'observation afin d'améliorer leur activité.

La première action que nous avons menée porte sur la troisième question de recherche "*Comment prendre en compte la connaissance métier dans des indicateurs basés sur des traces spécialisées ?*". En effet, avant de nous focaliser sur la formalisation du processus d'analyse, nous avons souhaité proposer une solution de calcul d'indicateurs qui ne dépendait pas d'un format de traces particulier ou d'un langage de scénarisation. L'idée est alors de faire reposer le calcul des indicateurs sur la connaissance métier qualifiant les données d'observation plutôt que sur le format de représentation des traces interne aux EIAH. Nous avons proposé un langage intermédiaire [Iksal et Choquet 2005b] d'extraction et de qualification du contenu des traces avec les concepts observables du scénario pédagogique (au sens de Vincent Barré [Barré et al. 2004]). Un prototype a été développé afin de montrer la faisabilité technique du calcul d'indicateur basé sur ce langage intermédiaire [Iksal et Choquet 2005a]. Ce résultat a été intégré comme tel dans le projet régional METEDI (Méthodes Et Technologies pour l'Enseignement à Distance avec Internet).

Cette action a permis de montrer qu'il était possible de s'abstraire du modèle de traces de l'EIAH pour le calcul des indicateurs au sens où la représentation interne de l'EIAH n'entre pas dans la construction des indicateurs. Cela permet une forme de généricité des indicateurs. En effet, à partir du moment où l'EIAH permet de collecter les informations nécessaires à la valuation de l'indicateur alors cet indicateur peut être réutilisé.

Les limites de cette proposition se situent en particulier dans la "boîte noire" que constitue l'indicateur à ce stade. Pour l'utilisateur, un indicateur est le plus souvent un graphique généré par l'EIAH et qui lui propose une information sur la situation d'apprentissage. La première question de recherche (Q1) que nous souhaitons résoudre correspond à l'aide

que l'on peut lui apporter pour mieux percevoir et comprendre sa session d'apprentissage. Pour cela il est nécessaire d'intégrer l'utilisateur dès la conception de l'indicateur, pour cela nous nous sommes focalisés dans un deuxième temps sur le processus d'ingénierie de l'observation basé sur la prescription.

Suite à ce premier résultat, nous avons collaboré au projet européen DPULS (Design Patterns for recording and analysing Usage of Learning Systems) [DPULS 2005a] sur la capitalisation des patrons d'analyse en EIAH, et co-encadré la thèse de Noa Randriamalaka (soutenue en 2008) portant sur l'«Assistance à la modélisation et à l'analyse des traces d'usage dans un EIAH : une approche par la réingénierie». Les travaux de cette thèse se sont focalisés sur le processus de capitalisation et sur le processus de combinaison de données d'observation existantes afin de faire émerger des usages non prévus [Randriamalaka et Iksal 2006][Randriamalaka et al. 2007]. Ces collaborations nous ont amené à revenir du côté de l'enseignant/concepteur et à nous intéresser plus particulièrement au processus global d'ingénierie de l'observation. Notre objectif pour ce processus est d'étudier où, quand et comment nous pouvons intégrer les acteurs de l'EIAH pour que l'observation les aide à comprendre et améliorer leur situation d'apprentissage (Questions de recherche 1 et 4). Il n'était donc plus question de se limiter au processus d'analyse de traces, mais de travailler sur un processus plus global dans lequel l'analyse de traces ne constitue qu'une étape. Nous proposons d'intégrer les usagers dans le processus d'observation dès la phase de réflexion sur l'observation, plutôt que de les considérer uniquement comme consommateurs du résultat. Il s'agit d'une proposition dans la logique du projet REDiM qui place l'utilisateur au cœur du processus d'ingénierie/ré-ingénierie de l'EIAH. De notre côté, nous avons élaboré une chaîne éditoriale de l'observation [Iksal 2011b] au sens où cette dernière est prise en compte dès la phase de conception de l'activité pédagogique, jusqu'à son exploitation dans un processus de ré-ingénierie ou d'adaptation. Nous avons utilisé le terme de chaîne éditoriale car nous avons un prescripteur qui scénarise l'observation et propose des indicateurs avec leurs moyens de perception en fonction des destinataires de l'observation. Cela se rapproche de nos travaux sur ICCARS dans lequel le journaliste prescrivait un dossier de presse dont la vue dépendait des destinataires de l'information. La notion de perception utilisée dans cette définition correspond aux moyens techniques mis en œuvre pour représenter le résultat d'un indicateur, cela peut être un moyen visuel, auditif, sensitif (vibration) etc.. Dans sa thèse, Noa Randriamalaka s'est intéressé à deux éléments, tout d'abord l'étude d'une approche par patrons pour la spécification de l'observation en lien avec le projet européen DPULS, et ensuite la possibilité pour un enseignant/concepteur de combiner des indicateurs déjà établis et évalués afin de faire émerger de nouveaux usages traçables [Randriamalaka 2008].

L'ingénierie de l'observation par la prescription est un processus qui montre qu'il est possible d'impliquer les acteurs de l'EIAH afin d'intégrer leurs besoins au plus près de la conception. Toutefois, pour que l'ingénierie de l'observation assiste effectivement les usagers, il est nécessaire de l'opérationnaliser et par conséquent de définir et formaliser les différents composants. Il s'agit néanmoins du premier cadre méthodologique de l'ingénierie de l'observation défini de manière aussi précise, focalisé sur les usagers des EIAH et non à destination des chercheurs en analyse de données. La capitalisation et la réutilisation intégrées dans ce processus sont envisagées en considérant comme destinataires l'ensei-

gnant, le tuteur ou encore l'apprenant contrairement à des approches comme DataShop [Stamper et al. 2011], WEKA [Hall et al. 2009] ou encore MULCE [Reffay et al. 2008].

L'intégration des usagers passe par l'expression de leurs besoins d'observation puis par leur déclinaison sous la forme d'un indicateur. C'est pourquoi nous avons souhaité résoudre les questions de recherche 2 et 3 correspondant à la prise en compte de la connaissance métier et à la formalisation des indicateurs capitalisables et réutilisables. Pour cela, deux actions ont été menées, nous avons tout d'abord élaboré en collaboration avec Christophe Choquet [Choquet et Iksal 2007a], un langage de spécification déclarative de données d'observation appelé Usage Tracking Language (UTL). Ce langage a pour objectif de décrire chaque donnée en indiquant ce qu'elle est, la façon de l'obtenir et à quoi elle sert. Pour cela, nous nous sommes appuyés sur notre premier résultat, c'est-à-dire le langage intermédiaire d'extraction et de qualification de l'information dans les traces. En parallèle, il était important de se rapprocher de l'univers métier de l'enseignant/concepteur, et à ce titre Boubekur Zendagui a réalisé une thèse [Zendagui 2010] soutenue en 2010 sur le "Support à la spécification des besoins d'observation dans un contexte de ré-ingénierie des scénarios pédagogiques : une approche dirigée par les modèles". Nous avons pu avancer à la fois du côté de l'opérationnalisation avec le langage et du côté des usagers, en l'occurrence de l'enseignant avec la spécification des besoins d'observation. En ce qui concerne les besoins d'observation, l'originalité de l'approche de Boubekur Zendagui réside dans le fait que la spécification des besoins est guidée par la description de la situation d'apprentissage formalisée par le scénario pédagogique. Il propose un modèle conceptuel des besoins d'observation et un processus d'aide à la spécification de ces besoins qui a pour but principal de structurer les différentes informations nécessaires pour préparer une activité d'observation. Nous n'étions pas co-encadrant de cette thèse mais ses résultats sont importants au niveau de l'expression du besoin d'observation par l'enseignant par l'intermédiaire de langages graphiques de modélisation.

De notre côté, nous avons proposé le langage déclaratif et opérationnalisé UTL pour décrire les données d'observation sous une forme capitalisable et réutilisable. Par données d'observation, nous considérons toutes les données entrant dans le processus de calcul des indicateurs. L'un des points forts de ce langage est son indépendance vis-à-vis des formats de représentation de traces dans les plate-formes, ainsi que d'un langage de scénarisation pédagogique. De plus, il est basé sur XML ce qui permet une manipulation à la fois par un acteur humain et un programme informatique. Afin de le rendre automatisable, il était nécessaire de disposer d'un moyen d'expression du calcul des indicateurs. A ce titre, nous avons co-encadré la thèse de Diem Pham Thi Ngoc, soutenue en 2011, sur la "Spécification et la conception de services d'analyse de l'utilisation d'un Environnement Informatique pour l'Apprentissage Humain" [Phâm Thi Ngoc 2011]. Dans le cadre de cette thèse, un langage déclaratif de composition de données d'observation basé sur UTL a été conçu avec l'objectif d'être indépendant de l'architecture de stockage des données et d'être basé sur la description sémantique et déclarative des données. Il est générique et instrumenté. Cette indépendance nous permet de pouvoir exploiter et tester l'efficacité de différents systèmes de stockage et de différents langages de programmation sans pour autant remettre en cause la description des données. Ce langage est utilisé actuellement au sein de l'Université du Maine avec des calculs d'indicateurs en cours de session. Cette utilisation est présentée

dans le cadre des travaux d'Aïna Lekira sur l'EIAH Hop3X [Lekira et al. 2011a], dont la thèse a été soutenue le 2 juillet 2012.

La limite principale de ce langage est liée à l'utilisateur, en effet, il est difficile pour un enseignant/concepteur et par conséquent un apprenant ou un tuteur, de spécifier la règle de calcul d'un indicateur. Pour cela, le passage par un analyste de données est nécessaire, ce qui rend la spécification en cours de session de nouveaux indicateurs difficile actuellement. De plus, le résultat du calcul n'est pas directement exploitable par les usagers, un couplage avec des mécanismes de perception est nécessaire.

C'est pourquoi, nous avons commencé à travailler sur la cinquième question : "*Comment exploiter ce savoir-faire capitalisé pour améliorer la session ?*", au travers de la thèse de Mohand Akli Ouali qui a démarré en octobre 2011. Cette thèse, que nous co-encadrons, concerne la "Modélisation outillée d'une stratégie contextualisée d'observation dans le cadre d'un environnement d'apprentissage". Selon notre point de vue sur l'observation, l'expression du besoin n'est qu'une étape dans l'activité de l'enseignant, et nous souhaitons plus particulièrement nous intéresser aux stratégies d'observation. Une stratégie d'observation comporte non seulement le besoin d'observation mais aussi ce qu'il est possible de faire du résultat de cette observation, la manière dont l'enseignant souhaite percevoir le résultat ou comment il souhaite que les autres usagers le perçoivent, ainsi que la façon dont il compte lui-même s'en servir. Dans le cadre de ses travaux, Mohand Akli Ouali mène une étude sur le sujet de l'observation et de son utilisation auprès d'enseignants/concepteurs afin d'identifier ce qui compose la stratégie d'observation. Il s'agit d'un travail qui démarre et pour lequel nous n'avons pas encore de résultats. Néanmoins, même s'il existe des tableaux de bords dans certains EIAH [Guéraud et al. 2004], l'enseignant/concepteur n'a jamais été associé à la spécification de sa stratégie d'observation.

Ces travaux constituent d'ores et déjà une avancée intéressante dans notre projet de recherche. Nous avons encore un certain nombre de thèmes à aborder afin de répondre aux différentes questions de recherche, tout d'abord en ce qui concerne le partage et la réutilisation, cela nécessite de développer une communauté de pratique et par conséquent de proposer une méthodologie instrumentée d'exploitation de ces informations et des techniques d'observation associées. Ensuite, nous souhaitons faire évoluer la spécification des données d'observation de manière à rendre plus autonome le prescripteur, grâce par exemple à des langages graphiques, ce qui permettrait la spécification en cours de session de nouveaux indicateurs. La définition d'une communauté passe aussi par la prise en compte des avis des usagers sur les indicateurs ainsi que par la collecte des solutions mises en œuvre, que ce soit lors de la réingénierie, du tutorat ou encore de l'adaptation. A ce sujet, nous avons déposé le projet ANR OPALE⁸ dont l'objectif principal est de proposer des solutions de capitalisation, de partage et de réutilisation des patrons d'observation dans un but d'adapter le cours en session ou de faire de la ré-ingénierie sur une session de formation. Ce projet en partenariat avec le LIRIS et l'IRIT n'a pas été retenu cette année, les différents partenaires envisagent de le soumettre dans un autre type d'appel à projets.

8. Observational Patterns for Adaptation in Learning Environments

1.4 Plan du mémoire

Notre cheminement scientifique est composé de plusieurs actions de recherche permettant une avancée par étape de nos travaux. Ce mémoire suit une organisation similaire. Il s'organise autour de sept chapitres en comptant celui-ci. Tout d'abord, **ce chapitre introductif** présente notre problématique et notre cheminement scientifique. Ce qui permet de situer nos travaux par rapport aux questions de recherche que nous voulons résoudre ainsi que par rapport à notre problématique. La présentation de ces éléments a pour objectif de préparer la lecture des chapitres suivants.

Avant de présenter nos actions de recherche et nos résultats en détail, nous avons organisé **le deuxième chapitre** autour de l'observation. Notre objectif, dans ce chapitre, est de présenter la terminologie du domaine de l'observation et d'introduire les étapes les plus courantes du processus d'observation grâce aux différents travaux du domaine. Nous avons souhaité marquer la différence entre le processus d'observation qui décrit la mise en œuvre de cette dernière et l'ingénierie de l'observation qui relève d'une méthodologie plus globale. L'ingénierie de l'observation correspond à notre proposition, elle intègre le processus d'observation mais aussi toute la réflexion autour de l'expression du besoin jusqu'à la capitalisation et la ré-ingénierie de l'observation. Ce chapitre nous permet aussi de positionner nos travaux par rapport à d'autres dans le domaine notamment des EIAH.

La question de recherche principale qui nous guide porte sur l'aide que l'on peut apporter à l'enseignant/concepteur afin d'améliorer sa situation d'apprentissage grâce à l'observation de cette dernière. Nous avons choisi de focaliser nos travaux sur l'ingénierie de l'observation par la prescription comme cadre méthodologique permettant de répondre à cette problématique. Avant de proposer ce cadre, nous avons souhaité comprendre le processus d'observation au travers d'expérimentations menées au sein de différents projets (Thèse d'Alain Corbière [Corbière 2006], travaux du laboratoire SYSCOM [Iksal et al. 2007] et le projet VAPS [Klinger 2006] [Phâm Thi Ngoc et al. 2009]). **Le troisième chapitre** présente notre action autour de la recherche de généralité de manière à proposer des moyens d'observation indépendants des EIAH. Nous avons étudié les différents formats de traces et proposé un langage intermédiaire de représentation de l'information contenue dans les traces [Iksal et al. 2004] [Iksal et Choquet 2005a] [Iksal et Choquet 2005b]. Ce langage intermédiaire servant de base sur laquelle s'appuie la conception des indicateurs. La différence principale avec les autres approches de définition de langage intermédiaire repose sur l'utilisation de la sémantique métier pour décrire le contenu de l'information disponible dans les traces.

Une fois cette étude pragmatique menée, nous nous sommes intéressés à l'ingénierie de l'observation par la prescription. **Le quatrième chapitre** présente notre position sur l'ingénierie et notamment la chaîne éditoriale [Iksal 2011b] qui permet de situer les différents acteurs de l'observation du point de vue de l'ingénierie. Dans notre approche, nous plaçons le prescripteur (l'enseignant/concepteur, le tuteur, le chercheur, etc.) au cœur du processus. Nous avons choisi de baser l'observation sur les besoins du prescripteur et non sur une approche de type "fouille de données". Notre approche est complémentaire

car le principe de la fouille de données et de se baser sur les traces collectées pour faire émerger de la connaissance non prévue, or dans notre approche, nous savons exactement ce que nous cherchons à obtenir ou à prouver.

Notre approche de l'ingénierie de l'observation par la prescription repose sur une représentation des données d'observation. **Le cinquième chapitre** a pour objectif de présenter une modélisation capitalisable des données basée sur la sémantique de l'information contenue dans les traces. Nous proposons un modèle de description en trois parties : ce qu'est la donnée, comment l'obtenir et son utilisation. Ce modèle est utilisé dans un langage de représentation de données qui est opérationnalisable, le langage UTL – Usage Tracking Language [Choquet et Iksal 2007a] [Choquet et Iksal 2007b] [Phâm Thi Ngoc et al. 2010b] [Phâm Thi Ngoc et al. 2010a].

Afin d'éprouver nos propositions, nous avons instrumenté notre processus et nous l'avons exploité dans un EIAH. **Le sixième chapitre** présente les outils développés autour du langage de représentation de données UTL ainsi qu'un exemple d'utilisation sur Hop3X [Lekira et al. 2011b], un EIAH pour l'apprentissage de la programmation. Ce chapitre a pour objectif de valider notre proposition et de présenter une marche à suivre pour son utilisation.

Le dernier chapitre dresse une synthèse de nos travaux. Nous présentons aussi des perspectives en deux temps, tout d'abord les directions de recherche à destination de la communauté, ainsi que quelques pistes de recherche que nous souhaitons développer par rapport à nos propositions.

Chapitre 2

L'observation : une analyse de la situation

Sommaire

2.1 Terminologie	27
2.1.1 La Trace	27
2.1.2 L'indicateur	28
2.1.3 Les autres termes	29
2.2 Le besoin d'observer	30
2.3 La collecte des informations	33
2.3.1 La création ad hoc de mécanismes de collecte	33
2.3.2 La modélisation des données	33
2.3.3 Position sur la collecte	34
2.4 L'analyse des données	35
2.4.1 Fouiller les données : Faire émerger l'observé	36
2.4.2 Les approches par l'hypothèse	39
2.4.3 Positionnement sur l'analyse	41
2.5 L'interprétation de l'analyse	42
2.5.1 La visualisation au service de la compréhension	42
2.5.2 La modélisation d'apprenants	43
2.5.3 Positionnement sur l'interprétation	45
2.6 La capitalisation	45
2.6.1 DataShop	46
2.6.2 MULCE	46
2.6.3 DPULS	47
2.6.4 Positionnement sur la capitalisation	48

L'observation dans le domaine informatique est une problématique souvent abordée, que ce soit au travers de l'analyse de traces (par exemple en analysant des web logs) ou encore de la fouille de données. L'idée est souvent d'observer un comportement, un apprentissage, l'appropriation de l'outil. L'observation peut servir aussi à l'évaluation ou à l'adaptation et à la construction de modèles utilisateurs. Nous allons présenter dans ce chapitre une analyse qui n'est pas exhaustive mais qui présente les différentes facettes de l'observation au travers des travaux de recherche menés depuis plusieurs années. Cette analyse a pour but principalement de situer nos activités et aussi d'en définir la portée. Nous avons choisi d'organiser ce chapitre autour d'un cycle classique d'observation qui est expliqué dans la section 2.2. Ce cycle est composé de quatre étapes : (i) la collecte des traces ; (ii) l'analyse de traces ; (iii) l'interprétation des résultats d'analyse ; (iv) la capitalisation et la réutilisation de l'observation. Avant de commencer l'analyse, il nous semble nécessaire de clarifier le vocabulaire utilisé dans nos travaux, puis de présenter les motivations liées à l'observation.

2.1 Terminologie

L'observation et l'analyse de traces sont étudiées dans de nombreux domaines (IHM, EIAH, etc.), il est important pour nous de clarifier dès à présent la terminologie que nous utilisons. Nous commençons par les termes "*Trace*" et "*Indicateur*" qui sont souvent employés dans les articles scientifiques, puis nous terminerons sur les autres termes de notre vocabulaire.

2.1.1 La Trace

Il s'agit d'un terme qui est souvent utilisé mais pour lequel il n'existe pas vraiment de définition commune actuellement, même s'il existe un noyau commun. Ce problème de définition vient souvent du fait que chacun définit la trace dans un contexte particulier (trace d'interaction, trace d'apprentissage, etc.). Le contexte correspond essentiellement à l'environnement informatisé utilisé ainsi qu'aux objectifs du traçage. Nous allons aborder ici quelques unes de ces définitions avant de proposer la définition que nous adoptons pour le terme "*trace*".

Mostow et Beck [Mostow et Beck 2006a] définissent la "trace brute" comme une séquence d'observés directement enregistrée par un vecteur de collecte donné. Concrètement, dans le cadre de leur projet LISTEN, il s'agit d'un flux d'interactions qu'ils segmentent ensuite en une cinquantaine de variables.

La trace d'interaction de Jermann [Jermann et al. 2001] est définie comme l'observation ou l'enregistrement des actions réalisées par l'acteur sur le système au moment des interactions, ces informations étant collectées et stockées afin d'être traitées ensuite.

Dans le cadre de leurs travaux, Lund et Mille [Lund et Mille 2009] ainsi qu'un peu plus tard Settouti [Settouti 2011] se sont positionnés sur la définition d'une "trace numérique d'interactions" : "Il s'agit d'une suite temporellement située d'éléments, qui relève soit d'une interaction entre humains, médiatisée et médiée de diverses façons par une machine soit d'une suite d'actions et réactions entre un humain et une machine". On observe dans

cette définition l'importance du temps dans la description de la trace, et aussi le fait qu'il s'agisse d'une trace numérique d'interactions.

Settouti cite aussi une définition de Pernin indiquant que la trace est un indice de l'activité des acteurs d'une situation d'apprentissage qu'elle soit ou non instrumentée.

Sébastien George [George 2010] parle "d'une empreinte laissée par un acteur à un endroit et à un moment donnés". Dans cette définition, le concept d'empreinte reste flou sur la granularité de l'information.

Dans le cadre du projet TRAILS [TRAILS 2004], la trace est considérée comme un parcours cognitif identifié par une séquence d'évènements observés. A ce niveau, la trace est déjà un objet évolué, construit, organisé et qui fait sens aux usagers. Cette définition ne se situe pas au même niveau que les précédentes.

En ce qui nous concerne, nous considérons la trace comme relevant de l'observation directe ou indirecte des actions ou des productions de l'utilisateur (les empreintes de Sébastien George), et donc de l'activité de l'utilisateur. Il ne fait nul doute que le facteur temps est un facteur important dans la description de la trace. Elle est toujours collectée à un moment bien précis, que cet instant soit exprimé dans un format calendaire ou dans une simple relation d'ordre. Nous considérons par contre que ce facteur temps n'est pas indispensable à la construction de tous les indicateurs. C'est pour cela que de notre point de vue, le temps est une variable comme une autre, présente ou non dans la représentation de la trace. Prenons par exemple, un indicateur que nous avons modélisé dans le cadre du supermarché virtuel pour le réapprentissage. Cet indicateur correspond à la vérification que le patient a collecté tous les articles de la liste qui lui a été confiée, le plus important ici est de savoir si l'article a été pris ou non, le facteur temps n'entre pas en ligne de compte. Nous rejoignons aussi la définition de Pernin sur les aspects numériques, au sens où la collecte peut se faire par vidéo ou même au travers d'un agent humain d'observation. Le traitement de l'information sera par conséquent différent et impliquera alors un agent humain.

Notre définition de la trace :

De notre point de vue, *“la trace est le résultat d'une observation ou d'une production enregistré dans un vecteur de collecte particulier à un instant précis, elle est composée d'un ensemble de descripteurs reflétant l'activité des individus observés”*. Il s'agit d'une définition volontairement large, car elle ne se limite pas à la prise en compte de l'interaction, mais elle intègre aussi la notion de production.

2.1.2 L'indicateur

Sur la définition d'indicateur en EIAH, le consensus est plus évident. Les travaux menés dans le cadre des projets ICALTS [ICALTS 2004] et IA [IA 2005] servent de référence pour les autres définitions. Dans ICALTS, Dimitracopoulou et al. [Dimitracopoulou et al. 2004] proposent la définition suivante “learning indicators are variables that indicate [...] the mode, the process or the 'quality' of the considered 'cognitive system' activity, the features or the quality of the interaction product, the mode or the quality of the collaboration”. Il

s'agit ici de variables donc de données calculées qui servent à comprendre et/ou évaluer le système, l'interaction ou la collaboration. Le projet DPULS [David 2005] s'est basé sur cette définition pour proposer la suivante : "An indicator is a feature of a datum (usually of a derived datum). It highlights a relation between the datum and an envisaged event. In as far as they have value-oriented meaning, indicators can be considered variables. Indicators have to be interpreted, taking into account the learning activity, the profile and roles of the actors, as well as the interaction context.". Cette définition est plus précise, elle parle d'interprétation par rapport au contexte et de lien entre la donnée calculée et l'évènement concerné. Dans ses travaux, Gendron a préféré restreindre sa définition aux indicateurs collaboratifs. Elle indique qu'un indicateur de collaboration est un élément qui décrit ou mesure les caractéristiques d'une activité courante ou passée dans le but d'améliorer la collaboration [Gendron et al. 2012]. Dans cette définition apparaît le fait qu'un indicateur puisse être évalué en cours de session, ce qui n'est pas précisé dans les autres définitions, et qu'un indicateur peut être une description ou un calcul. Elle fait donc la distinction entre un agent humain et un outil informatique pour valuer les indicateurs.

De notre point de vue, ces définitions présentent ce qu'est un indicateur, mais il manque un élément important que Gendron ébauche dans le cas de l'indicateur de collaboration. En effet, un indicateur est défini pour répondre à un besoin, il y a nécessairement un objectif derrière l'indicateur. Nous considérons deux types d'objectifs, l'objectif d'observation qui relève du processus d'observation et de l'usage qui en sera fait, et l'objectif pédagogique qui concerne plus particulièrement l'enseignant/concepteur et l'intérêt d'un point de vue pédagogique de l'indicateur.

Notre équipe a donc adopté la définition suivante de l'indicateur :

Définition d'un indicateur :

"Un indicateur est un observable signifiant sur le plan pédagogique, calculé ou établi à l'aide d'observés, et témoignant de la qualité de l'interaction, de l'activité et de l'apprentissage dans un EIAH. Il est défini en fonction d'un objectif d'observation et motivé par un objectif pédagogique."

2.1.3 Les autres termes

La terminologie que nous utilisons dans nos travaux est présentée dans le schéma 2.1. Cette figure permet de situer les termes les uns par rapport aux autres. En plus de la définition de la trace et de l'indicateur que nous venons de donner, voici les définitions que nous avons adoptées pour les principaux termes liés à notre activité :

- Observable : un observable est la spécification d'une observation. Il est question de définir ici ce qu'il est possible d'observer dans un système. Ce concept est à différencier du besoin d'observation qui correspond à la spécification de ce que souhaite observer l'utilisateur.
- Observé : un observé est le résultat d'une observation.
- Trace brute : il s'agit d'une trace directement générée par l'EIAH sans transformation.

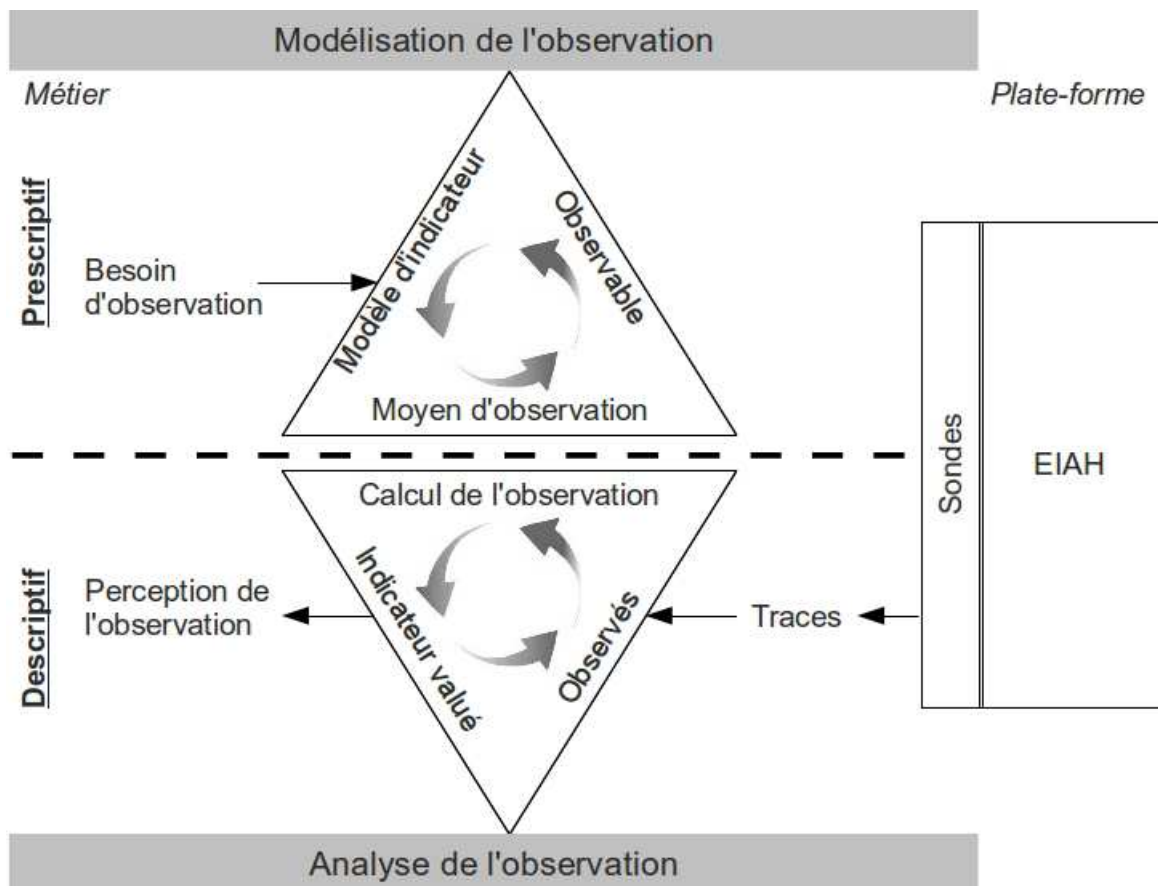


FIGURE 2.1 – Représentation de notre terminologie

- Collecte : la collecte est l'acte, situé dans le temps, d'enregistrement des traces brutes.
- Donnée brute : la donnée brute correspond à l'information extraite de la trace qui sert au calcul d'indicateurs.
- Sonde : la sonde est la fonction d'observation qui génère des traces au sein d'une plate-forme.
- Moyen d'observation : fonction de transformation appliquée sur un ensemble d'observés pour obtenir un autre observé.
- Modélisation de l'observation : c'est la spécification des observables et des moyens d'observation tendue par l'identification d'indicateurs.
- Analyse de l'observation : elle consiste en l'application des moyens d'observation sur les observés de manière à établir des indicateurs à partir des traces brutes.

2.2 Le besoin d'observer

Dans un premier temps, il est nécessaire de se poser la question de ce qu'est l'observation.

Définition tirée du Petit Robert [Le Robert 2011] :

D'un point de vue sémantique, l'observation est un examen attentif qui peut s'entendre de trois manières : (i) L'action de considérer avec une attention suivie [...] afin de mieux connaître. (ii) Le procédé scientifique d'investigation, constatation attentive des phénomènes tels qu'ils se produisent, sans volonté de les modifier. (iii) Surveillance attentive (d'un suspect, d'un patient).

Si l'on se place du côté informatique, ces trois déclinaisons de l'observation se retrouvent. En effet, nous allons observer pour mieux connaître un individu (construction d'un modèle utilisateur [Frias-Martinez et al. 2006] [Settouti et al. 2011]), et en tirer des conclusions par exemple sur ses connaissances, ses préférences etc.. Nous observons aussi pour constater des phénomènes comme l'utilisation ou non d'une ressource dans un environnement de formation ou encore l'usage de tout ou partie des fonctionnalités d'un environnement. Nous sommes aussi amenés à faire de la surveillance lorsque par exemple, le tuteur prend en charge un groupe d'apprenants et veille au bon déroulement de la situation d'apprentissage.

Dans notre domaine, l'observation est souvent associée à l'analyse de traces. En effet, il est pratique d'installer des sondes dans les environnements informatiques et d'enregistrer les traces de l'activité de l'utilisateur dans un format exploitable automatiquement. C'est pourquoi de nombreux travaux se sont intéressés à la découverte d'informations à partir de ces traces. De notre point de vue, l'analyse de traces n'est qu'une étape dans le processus d'observation, c'est pourquoi nous décrivons dans le chapitre 4, une ingénierie de l'observation.

Selon que l'on se place du point de vue de la trace ou de l'analyse, nous allons parler d'un cycle de vie des données ou d'un processus d'analyse de traces. Dans la communauté EIAH, May [May 2009] propose un cadre général représentant le cycle de vie des données de traçage (Figure 2.2).

Ce cycle se décompose en trois parties : (1) les environnements utilisés qui nécessitent d'être étudiés afin d'identifier ce que nous traçons et pourquoi nous le traçons. Madeth May considère que cela permet d'identifier les besoins des usagers ; (2) les systèmes de traçage spécifiques aux plate-formes permettant la collecte des informations d'interaction notamment ; (3) il s'agit du processus général d'exploitation des données tracées.

Ce point de vue sur les données de traçage nous semble restrictif. En effet, il préconise l'étude des environnements de formation pour identifier les besoins des usagers, nous considérons que l'utilisateur doit être la source de l'expression de ses propres besoins. L'analyse d'une plate-forme décrit le métier que le concepteur de cette dernière a choisi de représenter, l'utilisateur se retrouve alors avec une représentation imposée et souvent réduite de son métier. De plus, rien de nous empêche d'ajouter de nouveaux moyens de collecte d'information non pris en charge par la plate-forme mais fonctionnant en parallèle, afin de répondre à des objectifs précis des usagers. Par contre, nous sommes d'accord avec le fait que le plus souvent les systèmes de traçage sont spécifiques à une plate-forme.

D'une façon plus générale, nous constatons que le processus d'analyse de traces "classique" se décompose en trois étapes :

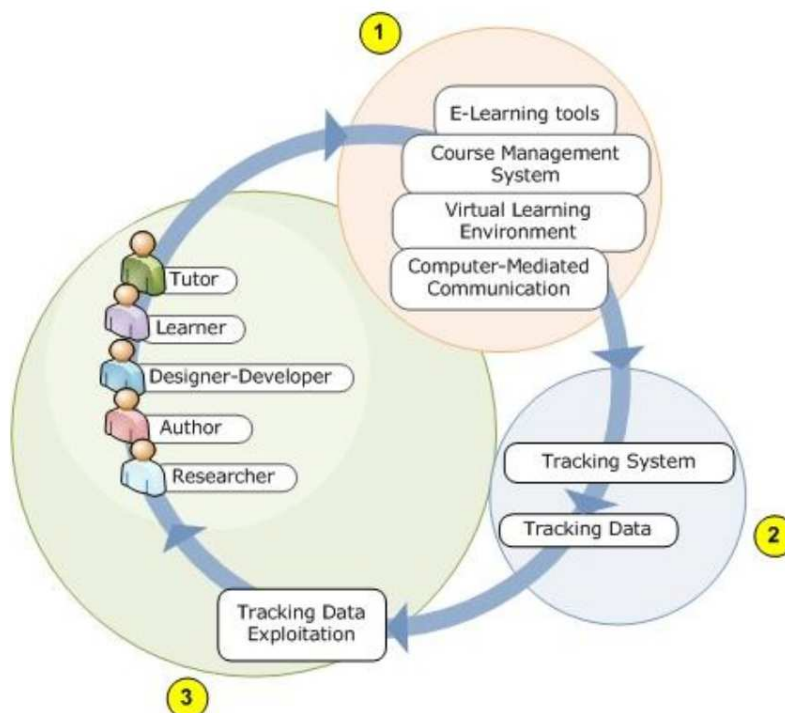


FIGURE 2.2 – Cycle de vie des données de traçage en EIAH [May 2009]

- **La collecte des informations** : Il est question ici de récupérer les informations à analyser. Souvent il s'agit de données comportementales stockées dans la base de données de l'outil ou dans des fichiers de logs. Parfois, lorsque l'environnement ne s'y prête pas, la collecte peut s'avérer beaucoup plus complexe.
- **L'analyse des données** : Il s'agit de l'étape la plus importante car elle fait passer la trace du statut d'information collectée au statut de donnée signifiante pour l'utilisateur. Elle est composée d'opérations de transformation, de combinaison, d'enrichissement ou encore de calcul sur les informations collectées.
- **L'interprétation de l'analyse** : L'analyse de données fournit une connaissance sur la situation d'apprentissage qu'il faut ensuite exploiter afin d'en tirer des conclusions. Il s'agit donc d'interpréter les données analysées et de faciliter cette tâche grâce à des mécanismes de perception plus ou moins adaptés aux usagers destinataires de ces informations.

Lorsque l'observation est à destination des chercheurs, une quatrième étape est prise en compte, il s'agit de la **capitalisation et du partage des corpus de traces et des algorithmes d'analyse**. C'est le cas dans les projets DataShop [Stamper et al. 2011] et MULCE [Reffay et al. 2008]. Nous considérons cette étape intéressante et importante car elle permet de fédérer des individus autour de l'observation, malheureusement ces projets ne sont qu'à destination des chercheurs. Nos objectifs sont similaires mais nos usagers sont les enseignants/concepteurs. Nous souhaitons les aider à initier un processus d'observation ou à améliorer le processus existant.

Dans la suite de ce chapitre, nous allons revenir sur chacune de ces quatre étapes pour positionner nos travaux par rapport aux autres approches du domaine.

2.3 La collecte des informations

Tout processus d'observation commence par une étape de collecte d'information. Cette dernière peut être automatisée ou non, systématique et/ou organisée selon un protocole particulier. Cette information peut prendre différentes formes, la plus classique étant la trace informatique, mais elle peut être sous la forme d'un enregistrement audio ou vidéo. Lorsqu'il n'y a pas de système d'enregistrement, la collecte d'information peut se présenter sous la forme d'une enquête ou d'un questionnaire qui sera renseigné par l'utilisateur directement ou par un observateur. Ces techniques non informatisées nécessitent par conséquent d'extraire les informations pour les rendre exploitables avec des procédés automatiques. Nous allons présenter quelques solutions de collecte créées pour des besoins spécifiques, puis nous introduirons la notion de modèle de données nécessaire à la manipulation des informations.

2.3.1 La création ad hoc de mécanismes de collecte

Lorsqu'il est question de tracer l'activité ou le comportement des usagers, la solution la plus courante est de développer directement dans l'application les différents capteurs. C'est le cas par exemple dans les projets LISTEN [Rasmussen et al. 2011], Hop3X [Peña de Carrillo et al. 2011], LOGIC-ITA [Merceron et Yacef 2003], pour n'en citer que quelques uns. Souvent cela repose sur l'intégration des événements dans une base de données, comme le pratiquent certaines plate-formes de formation à distance (MOODLE [MOODLE 2012], etc.), sur l'écriture dans des fichiers de logs comme FreeStyleLearning [vom Brocke 2001], ou encore sur l'envoi de données structurées ou formatées (en XML par exemple) directement sur un serveur dédié [Lekira et al. 2011a].

Parfois, lorsqu'il n'existe pas de plate-forme spécifique ou lorsqu'il n'est pas possible d'intervenir dans le développement de la plate-forme, une autre solution consiste à développer des systèmes fonctionnant en parallèle et permettant de capter avec ou sans l'aide des usagers ce qui se passe sur la machine, c'est le principe des enregistreurs de frappes au clavier (ou KeyLogger) par exemple. L'équipe SYSCOM [France et al. 2007] a développé un petit applicatif dans lequel l'apprenant indique son activité et ses changements d'activité afin qu'un outil de suivi, basé sur les faces de Chernoff¹, soit mis à jour en temps réel.

2.3.2 La modélisation des données

Étape obligatoire pour un traitement automatisé de l'information, la modélisation (d'autres parlent de "représentation") des données se heurte à différents problèmes. Tout

1. Système de visualisation dans lequel l'apprenant est représenté par une figure dont les caractéristiques évoluent dynamiquement dans le temps, selon les traces laissées par l'apprenant pendant la session. Cette solution est présentée à nouveau dans la section 2.5.1

d'abord, en raison de la multiplicité des environnements informatiques, l'analyste se retrouve face à une diversité importante de formats de stockage des traces de l'utilisateur, lorsque, dans le meilleur des cas, il dispose d'un environnement produisant des traces. Le premier modèle auquel nous sommes confrontés est donc celui de la collecte des données. Effectivement, même s'il n'est pas nécessairement explicite, la représentation choisie par les concepteurs de l'environnement ou des questionnaires d'observation constitue la base des travaux d'analyse. Lorsque les travaux sont menés spécifiquement pour un environnement, ces modèles sont directement exploités. Le projet LISTEN [Rasmussen et al. 2011] existe depuis 1993, les auteurs ont été amenés régulièrement à remettre en cause leurs modèles en raison de nouvelles fonctionnalités et des enrichissements nécessaires à leurs observations. Dès lors qu'ils maîtrisent toute la chaîne de leur EIAH, il est logique de l'adapter en permanence. Toutefois, même s'il s'agit de modèles extrêmement précis quant au contenu, ils sont la plupart du temps non documentés. Ce qui en diminue les possibilités de réutilisation.

D'autres approches consistent à passer par un modèle intermédiaire qualifié souvent de générique au sens où il n'est pas lié à un environnement particulier. Ces modèles sont particulièrement intéressants car ils proposent des approches différentes au niveau de l'exploitation des traces, c'est le cas notamment des Systèmes de Gestion de Bases de Traces [Djouad et al. 2009], ou encore des entrepôts DataShop [Stamper et al. 2011] et MULCE [Reffay et al. 2008] que nous abordons plus loin dans ce chapitre. Cette technique a été adoptée aussi dans le projet TrAVIS [May et al. 2008]. En effet, les auteurs ont adopté un format intermédiaire indépendant des outils et facilitant la conception d'outils d'analyse réutilisables. Le format intermédiaire qu'ils ont proposé a été conçu exclusivement pour les outils de communications médiatisées. Cette solution est similaire à celle que nous avons adoptée lors de notre première action de recherche, décrite plus précisément dans le chapitre 3. Il s'agissait dans notre cadre de proposer un langage intermédiaire pour le calcul d'indicateur permettant d'être indépendant des formats de représentation de traces propres à chaque plate-forme. De notre côté, nous n'avons pas limité notre approche aux communications médiatisées, ce qui est naturel étant donné que nous envisageons des EIAH scénarisés sans spécificités.

2.3.3 Position sur la collecte

Le processus de collecte d'information peut être considéré selon deux points de vue. Tout d'abord, la collecte des interactions, des productions ou des commentaires des usagers au sein même de l'EIAH, elle concerne plus particulièrement le concepteur de l'artefact technique. Le second point de vue est celui des analystes, au sens où il s'agit de la collecte des informations de traçage disponibles dans l'EIAH.

La collecte au sein de l'EIAH est une étape souvent négligée par les différents projets dans l'analyse de traces. Nombreux sont les travaux qui exploitent directement les traces produites par une plate-forme sans chercher à s'impliquer dans l'évolution de son système de traçage. Il serait intéressant, et ce n'est pas fait actuellement, de pouvoir agir sur les capacités de traçage des plate-formes en général. Au mieux, nous trouvons des projets dans lesquels la plate-forme est développée ad hoc, ce qui permet une évolutivité des mécanismes

de traçage car le code source est disponible et donc modifiable. Nous considérons qu'il est nécessaire de sensibiliser les développeurs d'EIAH aux possibilités de traçage et favoriser la conception de sondes activables². Cela reste cependant un artefact technique, or il est aussi souvent intéressant de disposer d'informations supplémentaires qui nécessitent un observateur externe. Ce dernier peut interroger après coup l'utilisateur afin qu'il justifie certaines actions, certains choix.

Dans le processus de collecte du point de vue de l'analyste, il est intéressant de noter la variété des modèles de données. En effet, de nombreux chercheurs dans le domaine d'analyse des traces ont élaboré un modèle spécifique pour leurs traitements des données (TrAVIS, IDLS, DataShop, etc.). Le plus souvent ces modèles sont dédiés à une activité comme l'analyse de l'interaction, l'analyse d'outils de communication médiatisée, etc.. De notre point de vue, il est intéressant de disposer d'un modèle de données qui sert de fondation à différents services d'analyse, dans le cas contraire, les modèles sont rapidement oubliés et inutilisés. En ce sens, le système IDLS est spécifique à la détection des styles d'apprentissage, sa généralisation à d'autres indicateurs est limitée. La modélisation des traces de SBT est intéressante et capitalisable. Les traces de plus haut niveau permettent d'obtenir une grande variété d'indicateurs. La différence entre nos deux approches concerne essentiellement l'approche d'analyse que nous présentons dans la section suivante.

2.4 L'analyse des données

Différentes approches co-existent [Gaudioso et Talavera 2006], elles sont catégorisables de deux manières. Tout d'abord, les approches d'analyse de données basées sur les traces et qui consistent à manipuler, fouiller ou transformer ces traces afin d'arriver à de la connaissance qui devrait faire sens pour l'utilisateur (souvent l'enseignant en EIAH). Nous situons dans cette catégorie les approches de type "fouille de données (ou data-mining)" [Hand et al. 2001], ainsi que les systèmes de gestion de bases de traces [Djouad et al. 2010].

Une autre façon d'aborder l'analyse de traces consiste à rechercher une information ou à valider une hypothèse. Nous les qualifions d'approches par l'hypothèse. Il est question de prouver qu'une information est présente ou de montrer qu'un comportement a bien eu lieu. Cela revient à concevoir l'analyse à partir de ce que l'utilisateur veut chercher et tenter de le prouver ou de le retrouver. Nous situons dans cette catégorie les travaux autour du calcul d'indicateurs, de la construction du cheminement de l'apprenant, de la création de modèles d'apprenants ou d'utilisateurs en général. Notre approche par la prescription qui est basée sur les besoins d'observation exprimés par l'utilisateur est pleinement dans cette mouvance.

Ces deux catégories sont complémentaires, les utilisateurs ont besoin aussi bien de systèmes leur permettant de prouver ce qu'ils ont imaginé, mais aussi de systèmes capables de faire émerger de nouveaux faits qui pourront impacter le comportement des utilisateurs sur la

2. Une sonde capte l'observable dans la plate-forme pour produire une trace. La notion de sonde activable signifie que la plate-forme a été conçue avec d'importantes capacités de traçage et que les utilisateurs activent ou désactivent au besoin la production de traces.

situation d'apprentissage. Nous avons organisé cette section selon ces deux catégories afin de présenter leurs particularités.

2.4.1 Fouiller les données : Faire émerger l'observé

Les analystes en fouille de données effectuent leurs analyses sur des volumes importants de données en utilisant des algorithmes pré-établis [Romero et al. 2010]. Dans leur ouvrage "Principles of Data Mining", Hand, Mannila et Smyth proposent la définition suivante "*Data mining is the analysis of (often large) observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner.*" [Hand et al. 2001]. On y retrouve la notion de grands volumes de données ainsi que le fait de découvrir "des relations non prévues". Il s'agit donc de faire éventuellement émerger des connaissances sur les usagers et leurs pratiques ou de présenter de façon synthétique les informations observées. La fouille de données repose donc sur trois étapes : (1) la conversion des traces en données sous forme de variables ; (2) le lancement d'un algorithme de fouille de données ; et (3) la synthèse du résultat d'une manière qui devrait faire sens à l'utilisateur (le plus souvent sous une forme graphique à interpréter). Dans les faits, la restitution du résultat nécessite des compétences spécifiques d'interprétation des graphiques, comme c'est le cas par exemple avec les nuages de points du projet TADA-Ed (explication reprise dans la section 2.4.1.2).

Nous avons choisi de nous arrêter sur quelques projets représentatifs de la fouille de données. Le projet WEKA fait office de référence dans le sens où il s'agit d'un environnement complet, capable d'intégrer des bases de traces et d'y appliquer des méthodes de fouilles de données connues ou en phase de test. WEKA est un environnement généraliste, en complément, nous présentons les projets TADA-Ed et LISTEN qui correspondent à la fouille de données en EIAH. TADA-Ed est aussi un système d'intégration de bases de traces et d'algorithmes dédiés aux enseignants. De son côté, LISTEN est intéressant car il s'agit d'un EIAH qui nécessite la fouille de données pour aider l'apprenant. Le dernier projet que nous présentons est le projet SBT (Systèmes à Bases de Traces) du LIRIS³ qui se différencie des autres par une approche plus constructiviste de la fouille de données. Dans cette approche, les traces sont modélisées puis elles subissent des transformations successives afin d'établir une connaissance significative pour l'utilisateur.

2.4.1.1 WEKA : Un environnement complet de fouille de données

Le système WEKA (Waikato Environment for Knowledge Analysis) [Hall et al. 2009] est parmi les plus représentatifs du domaine. Il a pour mission de proposer un ensemble d'algorithmes de data-mining et de machine learning (apprentissage automatique) permettant de manipuler différentes données et d'en faire émerger une certaine connaissance pertinente pour les utilisateurs et en particulier les chercheurs. Il permet l'importation de jeux de données ainsi que de nouveaux algorithmes. Les chercheurs peuvent alors mettre en concurrence différents algorithmes afin de tester la qualité de leurs résultats. Ces méthodes (clustering, classification, règles d'association, etc.) peuvent être associées à des

3. Laboratoire d'InfoRmatique en Image et Systèmes d'information, Lyon, France

mécanismes de visualisation afin de rendre le résultat plus convivial. Il ne fait aucun doute qu'en la matière WEKA est un outil très complet, toutefois, si l'on considère l'observation comme une aide à l'activité d'un enseignant, il n'est pas imaginable de laisser l'enseignant manipuler cet outil. La compréhension de son usage et de ses résultats reste l'apanage des analystes. Toutefois, WEKA peut s'avérer très utile pour diagnostiquer le ou les algorithmes pouvant entrer dans le calcul d'un indicateur qui pourra être aisément réutilisé par l'enseignant.

2.4.1.2 TADA-Ed : La fouille de données au service de l'éducation

Merceron et Yacef partagent notre avis sur la non pertinence de ce type d'outils dans le domaine de l'éducation [Merceron et Yacef 2005]. C'est pour cela qu'elles ont proposé l'outil TADA-Ed (Tool for Advanced Data Analysis in Education). L'idée principale de leur projet est de mettre les outils de fouille de données au service des enseignants. Ces derniers peuvent alors fouiller les données liées à l'activité de leurs apprenants afin de mettre en évidence des " patrons pertinents pédagogiquement ". TADA-Ed ne se limite pas aux données d'interaction ou aux fichiers de logs, mais prend aussi en compte des données plus sémantiques comme la pertinence d'une action ou encore la validité de la réponse. La logique de l'application est similaire à WEKA au sens où il existe des filtres pour importer des jeux de données, et le système inclut des algorithmes de data mining ainsi que des mécanismes de visualisation. L'interface se veut plus abordable pour un enseignant et notamment elle est conçue autour de son métier (comptage des erreurs faites par les apprenants, notion d'exercice à évaluer, etc.). Il s'agit d'un pas intéressant du côté des enseignants, toutefois, TADA-Ed repose sur le fait que l'enseignant maîtrise un minimum le vocabulaire de la fouille de données et qu'il sache aborder les outils de visualisation. En effet, dans l'exemple suivant (Figure 2.3) l'enseignant a choisi l'algorithme des K-moyennes (K-Means) et obtient ce graphique qu'il doit donc exploiter. Nous considérons que ces éléments ne sont pas évidents à appréhender, de plus ils sont difficilement exploitables en cours de session.

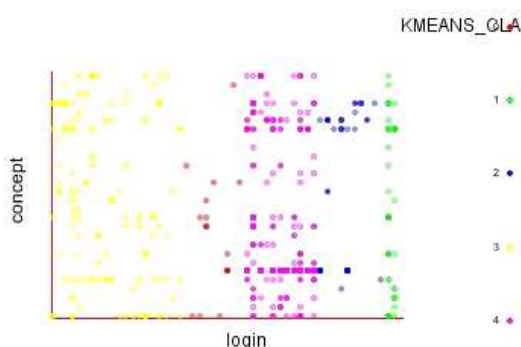


FIGURE 2.3 – Algorithme des k-moyennes avec 5 partitions [Merceron et Yacef 2005]

2.4.1.3 LISTEN : La fouille de données pour aider à l'apprentissage de la lecture

Dans le cadre du projet LISTEN (Literacy Innovation that Speech Technology ENables), les chercheurs de CMU (Carnegie Mellon University) ont développé un tuteur d'apprentissage à la lecture. Cet outil génère des traces qu'ils utilisent afin d'améliorer ce tuteur et de valider leurs propositions scientifiques [Rasmussen et al. 2011]. Le principe consiste à suivre la lecture de l'apprenant, d'identifier de façon automatique le mot qu'il est en train de lire puis de vérifier si la lecture est correcte. Il s'agit d'un tuteur automatique capable d'écouter l'apprenant et d'interagir avec lui afin d'améliorer sa lecture et aussi sa compréhension. S'il le souhaite, le tuteur peut aider à l'écriture de textes. Il s'agit d'un projet de recherche démarré en 1993 et qui continue d'évoluer aujourd'hui. Les auteurs ont proposé un modèle de données [Mostow et Beck 2006b] qui a évolué de nombreuses fois en fonction des fonctionnalités qui ont enrichi la plate-forme. Toutefois, cela reste un modèle spécifique et difficilement réutilisable, c'est le principal reproche que l'on peut lui faire.

2.4.1.4 Les systèmes à base de traces

Les Systèmes à Base de Traces (SBT) [Djouad et al. 2009] [Djouad et al. 2010], ou systèmes de gestion de bases de traces ont pour objectif de proposer un cadre conceptuel (cf. Figure 2.4) permettant de manipuler les traces dans un principe de généricité et d'interopérabilité. Ce système de gestion considère les traces comme une séquence temporelle d'observés toujours accompagnée d'un modèle capitalisable [Settouti et al. 2007]. Il s'agit d'une approche par modèles qui est très pertinente, entre autres parce qu'elle permet de manipuler et de fusionner des traces aux formats hétérogènes, provenant de plusieurs sources. Le modèle SBT considère toute trace, y compris celles obtenues par transformation de la trace brute, comme temporellement située (un indicateur étant une trace de très haut niveau). Il est à noter l'utilisation constante du terme trace. Dans leur terminologie, l'objet manipulé est et reste une trace.

L'approche retient trois types de transformations : la sélection, la réécriture de motifs et la fusion temporelle. Le système à base de trace est composé de trois services : la collecte, l'interrogation et la transformation. Une trace est composée d'un ensemble d'observés et d'un modèle de la trace. On passe ensuite de la M-trace première (M-trace pour trace modélisée) à l'indicateur grâce à des opérations de transformations successives (cf. Figure 2.5). L'indicateur est une M-trace dont le niveau d'abstraction est tel qu'elle devient significative pour l'utilisateur. On dispose alors d'un langage permettant d'exprimer la M-trace et d'un langage permettant sa manipulation.

Les aspects génériques sont présents dans le langage qui sert de base à la construction d'un SBT. En effet, ce langage permet de construire tous les SBT, et permet la réutilisation des modèles ou des transformations [Djouad et al. 2009].

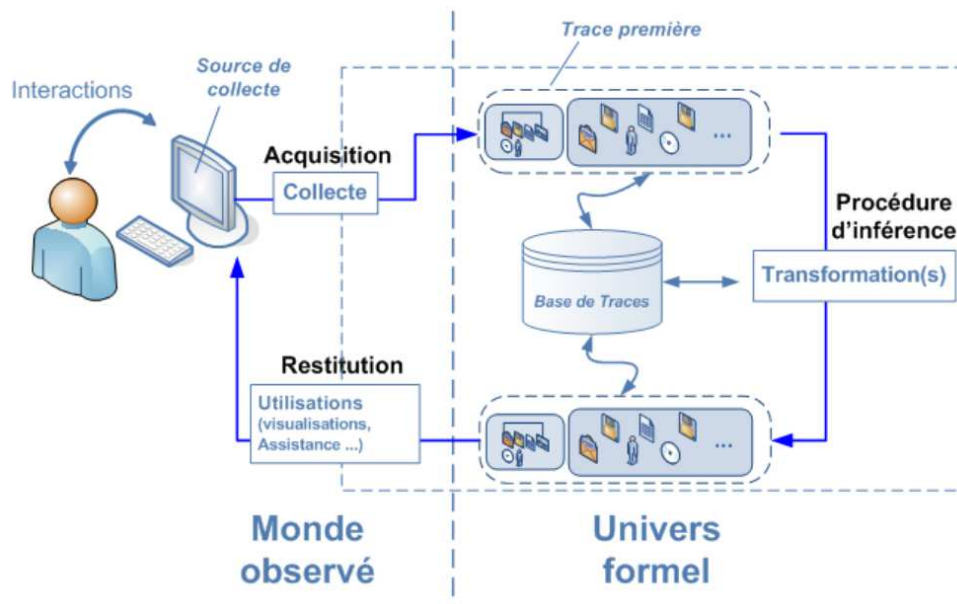


FIGURE 2.4 – Fonctionnement général d'un SBT [Settouti 2011]

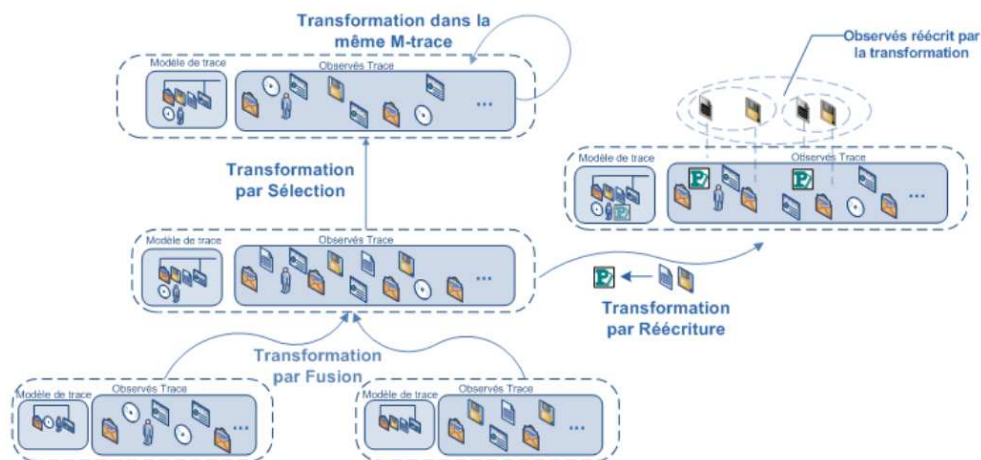


FIGURE 2.5 – Exemples de transformations avec SBT [Settouti 2011]

2.4.2 Les approches par l'hypothèse

Les approches par l'hypothèse sont basées sur un principe différent de la fouille de données au sens où l'analyse est guidée par l'information qu'il souhaite obtenir. L'idée n'est pas de faire émerger une nouvelle connaissance mais de rechercher un schéma, de démontrer des hypothèses. Cette approche est complémentaire par rapport à la fouille de données. En effet, dans le cas où la fouille de données a fait émerger un fait ou un comportement, il est possible de le spécifier ensuite sous la forme d'un indicateur sans pour autant réappliquer les algorithmes de fouille de données. Le plus souvent, la spécification de ce qui est à retrouver se fait indépendamment d'un algorithme de fouille. Nous avons choisi quelques projets représentatifs de l'approche par l'hypothèse. Tout d'abord

le projet TRAILS dont le but est d'identifier le parcours cognitif de l'apprenant à partir d'une reconstruction de sa navigation dans les ressources. Ensuite le projet ICALTS et le système IDLS se sont focalisés sur la notion d'indicateurs à modéliser pour identifier des comportements, ou des styles d'apprentissage.

2.4.2.1 De la trace au parcours : Le projet TRAILS

Dans le cadre du Réseau d'Excellence Européen Kaleidoscope [KALEIDOSCOPE 2004], le projet TRAILS (Personalized and Collaborative Trails of Digital and Non-Digital Learning Objects) [TRAILS 2004] s'est intéressé aux parcours cognitifs (trails) que les apprenants suivent et construisent quand ils naviguent dans un ensemble de ressources pédagogiques. Les apprenants interagissent avec les ressources pédagogiques en suivant des parcours identifiés par une collection d'événements observés temporellement situés. TRAILS s'est focalisé sur l'observation des parcours individuels au sein d'un ensemble de ressources pédagogiques, en cherchant à identifier le parcours cognitif de l'apprenant. Dans cette approche, la trace est assimilée à une connaissance et le parcours correspond à un apprentissage. Dans nos travaux, nous considérons les traces collectées lors des sessions d'apprentissage, nous ne restreignons pas le sens du terme *trace* à une séquence d'actions ou aux traces d'interactions (comme dans TATIANA [Dyke 2009] par exemple), mais nous considérons comme trace toute donnée fournissant de l'information sur une session d'apprentissage. Cependant nous pensons, comme le propose le projet TRAILS, que le processus d'analyse des traces amène à abstraire ces dernières de manière à les utiliser pour mieux comprendre le parcours cognitif et l'activité d'un apprenant par exemple. Pour faciliter cette compréhension et de façon à ce que les résultats soient plus pertinents pour le concepteur, TRAILS envisage la comparaison entre le parcours planifié – planned trail – et le parcours émergent – emergent trail. Dans notre contexte d'EIAH scénarisé, nous proposons que les résultats de cette analyse soient éventuellement représentés comme un scénario descriptif qui peut être comparé avec le scénario prédictif. Ce qui implique d'exprimer dans la mesure du possible, et de la volonté du prescripteur, les deux scénarios dans le même langage.

2.4.2.2 Le calcul d'indicateurs

Le projet ICALTS (Interaction and Collaboration Analysis' supporting Teachers and Students' Self-regulation) [ICALTS 2004] et le projet IA (Interaction Analysis – Supporting participants in technology based learning activities) [IA 2005] qui l'a prolongé, issus tous deux du réseau Kaleidoscope, “proposent que la conception des EIAH ne porte pas seulement sur les besoins initiaux d'action et de communication, mais soit étendue à la spécification des besoins d'analyse des interactions très complexes entre les acteurs d'une activité pédagogique, qu'ils travaillent individuellement ou en collaboration”. Ainsi, la définition des besoins d'analyse de l'utilisation d'un EIAH est une activité de conception pédagogique, et les traces elles mêmes sont des objets pédagogiques. Ces projets ont introduit et défini le concept d'Indicateur d'Analyse de l'Interaction (Interaction Analysis Indicator) comme étant “une variable généralement calculée ou établie à l'aide de données observées, témoignant du mode, du processus ou de la qualité de l'interaction”. Nous

partageons entièrement cette position.

L'indicateur d'un point de vue technique, est la donnée établie sur la base d'observés qui est signifiante pour l'utilisateur [Dimitracopoulou et al. 2004]. Nombreux sont les projets qui ont traité de la notion d'indicateur et surtout qui ont proposé des descriptions automatisées ou non d'indicateurs. Dans le domaine éducatif, le projet ICALTS [ICALTS 2004] a listé un certain nombre d'indicateurs liés aux traces d'interactions. Ces indicateurs ont servi de référence lors de la présentation et du test de certains outils [Djouad et al. 2009] [Phâm Thi Ngoc et al. 2010a].

La définition d'indicateurs a toujours un objectif particulier : évaluer, diagnostiquer, adapter, assister etc.. Prenons par exemple le système IDLS (Indicators for the Detection of Learning Styles) [Bousbia et Labat 2007] [Bousbia et al. 2010] [Bousbia 2011] ; il s'agit d'un outil spécialement conçu pour le diagnostic des styles d'apprentissage des apprenants. Il est nécessaire d'avoir une connaissance des modèles de traces afin de les convertir par transformation dans le modèle interne à IDLS. Cette transformation se fait grâce au langage XSLT ce qui présuppose des formats balisés pour les traces. Ensuite le système reconstruit le graphe de parcours de l'apprenant dans la session pour calculer des indicateurs qui permettront de déduire les styles d'apprentissage. Les auteurs introduisent la notion d'indicateur de niveau haut, intermédiaire et bas nécessaire au calcul d'autres indicateurs. Ce découpage se rapproche de nos concepts de données brutes, données intermédiaires et indicateurs. L'approche par notifications dans l'arbre des indicateurs qui permet de savoir si un indicateur est calculable ou non est très intéressante. Par contre, l'auteur ne mentionne pas l'intérêt de ses indicateurs de bas niveau. Par exemple, l'indicateur "durée d'une activité" est un indicateur de bas niveau parce qu'il est basé directement sur les traces. Nous pensons que le terme d'indicateur à ce niveau n'est pas pertinent, la durée d'une activité seule n'est probablement pas significative pour l'utilisateur, par contre le fait de savoir si l'activité a pris plus ou moins de temps que prévu se rapproche plus de la définition d'un indicateur.

2.4.3 Positionnement sur l'analyse

Nous avons présenté ici plusieurs solutions considérant le processus d'analyse de données. Si l'on se place du point de vue de l'enseignant/concepteur qui souhaite analyser sa situation d'apprentissage, il est évident que des solutions purement fouille de données ne sont pas pertinentes et cela d'autant plus si l'analyse doit avoir lieu en cours de session. L'approche par l'hypothèse et donc avec la construction d'indicateurs semble la plus pertinente sauf que ces indicateurs ont souvent été conçus pour une plate-forme donnée ou pour un objectif particulier et dans ce cas le modèle servant de base est très spécifique. Ce qui amène l'enseignant à devoir convertir ses données dans plusieurs formats pour bénéficier de plusieurs indicateurs. L'approche des systèmes de gestion de bases de traces est indépendante d'un objectif particulier, les modèles de traces premières sont élaborés selon les différentes plate-formes, ce qui permet un usage plus étendu. Cette solution consiste à manipuler les traces au travers d'un atelier de transformation jusqu'à obtenir une trace modélisée qui soit signifiante pour l'enseignant/concepteur. Cette étape est difficilement réalisable par l'enseignant, un spécialiste de la fouille de données est de notre point de vue

nécessaire afin d'identifier les transformations pertinentes. De plus, dans cette approche, le besoin d'observation de l'enseignant n'est pas pris en compte, alors qu'étant souvent le principal intéressé, il est aussi le premier prescripteur de l'observation.

2.5 L'interprétation de l'analyse

L'analyse une fois effectuée, fournit un ensemble de résultats qui souvent ne sont pas exploitables directement par les usagers, c'est pourquoi il est nécessaire de proposer des mécanismes de perception comme des représentations visuelles. Guéraud et al. ont souligné l'importance de la représentation visuelle pour les tuteurs et les apprenants dans le projet FORMID [Guéraud et al. 2004]. Ces représentations ont la particularité d'être aisément appréhendables par les usagers, nous présentons ici quelques exemples pour illustrer ces mécanismes. Dans certains cas, la perception directe n'est pas nécessaire car le résultat de l'analyse a un autre objectif, par exemple rassembler de la connaissance pour élaborer un modèle d'apprenant. Nous avons choisi de présenter le projet TrAVis [May et al. 2008] qui est assez complet sur l'analyse des communications médiatisées en utilisant, comme dans notre approche, un modèle intermédiaire. Toujours dans le domaine de la visualisation, le laboratoire SYSCOM a proposé plusieurs approches intéressantes [France et al. 2007] et faciles à appréhender pour un enseignant. Enfin, le système IDLS que nous avons présenté un peu plus tôt [Bousbia et al. 2010], propose aussi une visualisation des styles d'apprentissage. Afin d'illustrer une autre interprétation de l'analyse, nous présentons les modèles d'apprenants de Settouti [Settouti et al. 2011] qui montrent la reconstruction d'information à partir des traces.

2.5.1 La visualisation au service de la compréhension

Le projet TrAVis (Tracking data Analysis and Visualization platform) [May et al. 2008] a été développé pour assister les acteurs d'une formation à distance (enseignants-tuteurs et apprenants) lors de l'exploitation des traces des outils de communication médiatisées. Cet environnement propose d'analyser et de visualiser les traces sous différentes formes ainsi qu'à différentes échelles (cf. Figure 2.6). Avec TrAVis, les enseignants analysent des traces en construisant les indicateurs sur les interactions des apprenants. Quant aux apprenants, ils peuvent utiliser TrAVis comme un outil réflexif sur leurs activités de communication grâce à des indicateurs leur permettant de prendre conscience de leur propre comportement et de celui des autres. Chaque sphère dans la figure représente une activité, le diamètre de la sphère est proportionnel à la durée passée sur l'activité. La distance entre deux sphères représente le temps entre deux lectures de message et la couleur de la sphère représente les actions de l'utilisateur (défilement, affichage du message, lecture partielle ou masquage du message).

Les chercheurs du laboratoire SYSCOM se sont eux aussi intéressés aux mécanismes de visualisation de l'information analysée. Avec les faces de Chernoff [France et al. 2007] (cf. Figure 2.7), l'idée est d'identifier les apprenants en difficulté d'un simple coup d'œil : plus la face devient foncée (rouge dans l'application) plus l'apprenant est en retard dans sa tâche. L'interprétation de la figure 2.7 est la suivante [Iksal et al. 2007] :

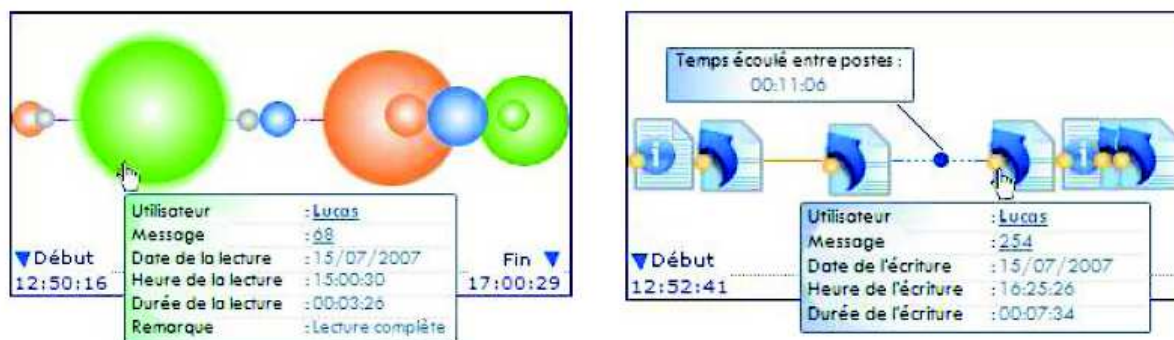


FIGURE 2.6 – Exemples de visualisation d'indicateurs avec TrAVIS [May et al. 2008]

- Mélanie est en train de passer d'une activité à l'autre. Elle est déjà passée dans toutes les activités et on constate (G) qu'elle a déjà passé beaucoup trop de temps dans les activités (A) et (B) qui sont en rouge dans son diagramme.
- La couleur de fond des activités (en bleu) fonce (B) lorsqu'en moyenne les apprenants dépassent le temps imparti pour la réalisation des activités concernées.
- (C) est l'activité compte-rendu et personne n'est en train de la faire actuellement.
- (A) est actuellement réalisée par 3 apprenants dont un depuis trop longtemps (visage foncé) et un quatrième est en train de passer à cette activité (lien).

Ces différentes visualisations évoluent dynamiquement dans le temps, en fonction des traces laissées par l'apprenant pendant la session. Par exemple, la forme des yeux du "visage" (D) d'un apprenant donné (fermée, ouverte) correspond à la quantité d'événements produits par l'apprenant pour une certaine activité. Si un apprenant "dort" dans une activité particulière, le tuteur peut intervenir en lui suggérant de participer plus activement. Une flèche (F) entre deux activités indique qu'un apprenant est passé d'une activité à une autre. Enfin, plus l'épaisseur d'une flèche est importante (E), plus il y a d'apprenants ayant choisi de passer de cette activité à la suivante. Cette représentation permet d'identifier des parcours types en suivant les traits les plus épais.

Une autre façon de visualiser l'information est la proposition des barres d'ombre [Heraud et al. 2005]. Elles permettent la représentation des anomalies sur l'activité des apprenants, dans la figure 2.8 par exemple, nous avons trois zones significatives, la zone (a) qui correspond à la durée préconisée pour l'activité par le concepteur, la zone (b) qui correspond à un dépassement de cette durée, et la zone (c) qui précise qu'aucune activité n'a été observée sur ce laps de temps.

Dans le cadre des styles d'apprentissage, le système IDLS [Bousbia et al. 2010] propose des patrons représentatifs et graphiques (Scolaire, Etoile et Dispersion cf. Figure 2.9) pour décrire visuellement le parcours de l'apprenant.

2.5.2 La modélisation d'apprenants

L'analyse des traces d'apprentissage est utilisée, par exemple, pour la création de modèles d'apprenants, cela permet d'adapter au mieux l'environnement de formation au public. Les travaux de Lemya Settouti ([Settouti et al. 2010][Settouti et al. 2011]) ont pour

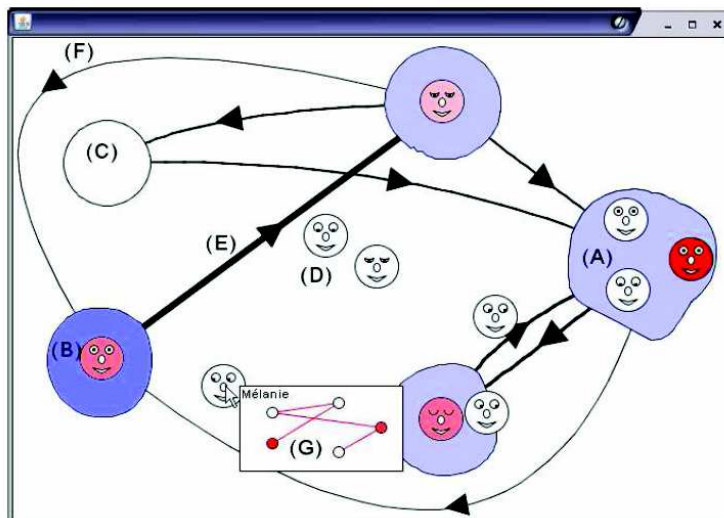


FIGURE 2.7 – Vue de la classe avec les faces de Chernoff [France et al. 2007]



FIGURE 2.8 – Représentation avec des zones d'ombre [Heraud et al. 2005]

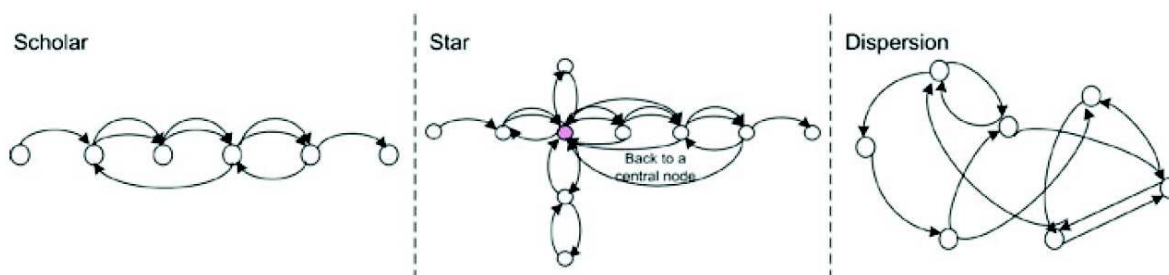


FIGURE 2.9 – Parcours types pour IDLS [Bousbia et al. 2010]

objectif de proposer un framework générique dédié à la modélisation d'apprenants basée sur les traces. Il est question ici d'utiliser les Systèmes à Bases de Traces afin d'extraire de la connaissance sur l'apprenant, cette connaissance induit un profil d'apprenant. Le système repose sur une modélisation des traces représentant le comportement des apprenants, une modélisation de la structure du modèle d'apprenant, ainsi qu'une description des règles de diagnostic permettant l'instantiation du profil. Le fonctionnement de ce framework est basé sur des requêtes faites avec le langage SPARQL, ces dernières pouvant être abstraites afin d'obtenir des patrons de requêtes stéréotypiques et réutilisables

(par exemple un patron de requête détectant le comportement “gaming the system” de l'apprenant).

2.5.3 Positionnement sur l'interprétation

L'interprétation de l'analyse est un travail aussi conséquent que l'analyse elle-même. Cette interprétation passe d'abord par la compréhension du résultat de l'observation. En effet, il est toujours préférable de disposer de moyens de perception qui font sens dans l'activité de l'utilisateur pour qu'il puisse appréhender un indicateur. Il existe de nombreuses propositions, nous en avons présenté quelques unes. Le reproche principal que nous pouvons faire correspond aux aspects spécifiques de ces mécanismes, ils ont été conçus le plus souvent pour répondre à un besoin lié à un EIAH ou à un type d'EIAH particulier [Guéraud et al. 2004] [May et al. 2008]. Or, nous pensons qu'il serait intéressant de disposer d'une banque de mécanismes de perception que les usagers pourraient utiliser à condition que ces derniers soient cohérents avec l'indicateur envisagé. C'est pour cela que nous proposons dans le cadre de la thèse de Mohand Akli Ouali [Ouali 2012] d'étudier le concept de stratégie d'observation. Cette stratégie est une description, externe à un EIAH, de la façon dont s'organise l'observation du côté du destinataire. La prise en compte du contexte, des capacités de perception du demandeur ou encore de la période de mise à disposition des indicateurs sont les éléments clés de cette description.

La perception est le moyen le plus courant d'exploitation des résultats de l'observation, toutefois, l'interprétation de l'analyse concerne aussi l'usage qui est fait de l'observation. Nous pouvons être amenés à exploiter de façon automatique ces résultats pour construire des modèles de connaissances ou de compétences, comme c'est le cas avec les modèles d'apprenants. De notre côté, nous envisageons une collaboration sur l'exploitation de l'observation dans le cadre de l'adaptation. Il s'agit du projet OPALÉ déposé auprès de l'ANR qui n'a pas été retenu mais que nous comptons déposer à nouveau dans un autre appel avec les mêmes partenaires.

2.6 La capitalisation

Notre objectif avec la capitalisation est de faciliter l'activité de l'enseignant ou du tuteur en permettant la réutilisation d'indicateurs existants. De notre point de vue, la réutilisation passe par une appropriation de l'observation et par les possibilités que peut offrir un système dans son intégration avec un EIAH au sens environnement et non plateforme. La capitalisation et la réutilisation sont deux fonctions souvent abordées ensemble. D'un côté, il est nécessaire d'étudier comment mettre en œuvre et améliorer la capitalisation afin de proposer la réutilisation. De l'autre, il faut faciliter l'appropriation par les usagers. Les premières questions à se poser portent sur l'objet concerné. Les projets actuels concernent la capitalisation des corpus d'apprentissage et dans une moindre mesure la capitalisation des moyens d'analyse. Afin d'illustrer les travaux actuels, nous présentons des projets dédiés à la capitalisation et au partage dans le domaine des EIAH. Le projet DataShop [Stamper et al. 2011] est intéressant car il propose la capitalisation et la réutilisation des deux types d'objets que sont les corpus d'apprentissage et les outils orientés

fouille de données permettant de les analyser. Le projet MULCE [Reffay et al. 2008] a des objectifs similaires, il est intéressant sur deux points, tout d'abord, il ne se limite pas à l'interaction comme DataShop et il prend en compte aussi les aspects plus organisationnels de la capitalisation comme la gestion des droits et de l'anonymat. Enfin, nous présentons le projet DPULS [DPULS 2005b] car il se préoccupe des usagers, en particulier les enseignants/concepteurs, plutôt que les chercheurs. De plus, ce projet s'est intéressé à la capitalisation des moyens d'observation.

2.6.1 DataShop

L'expérimentation est un point clé des activités de recherche, le projet DataShop [Koedinger et al. 2010] [Stamper et al. 2011] a pour objectif de permettre aux chercheurs dans le domaine de l'éducation et plus particulièrement dans le domaine de l'observation en EIAH, de tester et de vérifier leurs méthodes d'analyse. Pour cela, DataShop est un dépôt de corpus de données d'interaction, il est composé d'environ 300 corpus formatés selon leur propre modèle interne. En effet, le modèle est composé de trois types d'éléments : (1) le contexte du message (L'apprenant, le problème et la session avec le tuteur); (2) l'outil, et plus particulièrement l'action réalisée dans l'outil par le tuteur ou l'apprenant ; (3) la réponse du tuteur à l'apprenant. En complément, des outils d'analyse et leurs outils de visualisation de données sont mis à la disposition des usagers. Il est donc possible à l'aide de ce système, de proposer et de réutiliser des corpus, mais aussi d'appliquer nos propres outils d'analyse sur ces corpus ou de réutiliser les outils proposés. Si l'approche en elle-même est très convaincante du fait que la plate-forme fonctionne et est exploitée actuellement, elle possède deux inconvénients. Tout d'abord, elle a été conçue afin d'éprouver des mécanismes de fouille de données sur des données d'interaction uniquement, il est donc nécessaire de transformer les données d'observation à analyser dans leur format pour ensuite les utiliser. Cette transformation implique nécessairement une perte des informations ne correspondant pas à l'interaction. Ensuite, cette plate-forme a été imaginée pour des chercheurs en EIAH, elle n'est donc pas conçue pour être exploitée par une partie importante des destinataires de l'observation que sont les enseignants, les tuteurs et les apprenants. Il n'est pas envisageable de l'utiliser, par exemple, en cours de session.

2.6.2 MULCE

Le projet MULCE [Reffay et al. 2008] [MULCE 2009] s'est particulièrement intéressé au partage de corpus d'apprentissage et d'enseignement ainsi qu'à leur considération comme objet d'étude scientifique. Tout comme DataShop, ce projet insiste sur l'importance et les bénéfices du partage des données entre chercheurs et sur la nécessité de la contextualisation de ces données. La plate-forme associée permet le partage en ligne de corpus et offre la possibilité aux chercheurs d'effectuer des recherches selon des critères de classification des données partagées. Les corpus sont classés selon deux catégories : (1) les corpus globaux qui contiennent toutes les données collectées lors d'une expérimentation ; (2) les corpus distinguables qui sont à nouveau catégorisés selon l'activité des chercheurs : association d'une publication scientifique aux données, les données prêtes à l'emploi pour

des outils libres, partager des analyses avec leurs outils associés. Pour cette plate-forme, un format commun de partage et de représentation des données a été utilisé. Ce format intègre les traces, mais aussi le contexte, le scénario pédagogique, le protocole de recherche, les informations concernant les droits de licence, les fichiers de suivi d'apprenants, les articles de recherche etc..

2.6.3 DPULS

Le projet DPULS (Design Patterns for recording and analyzing Usage of Learning Systems) [DPULS 2005a] du réseau d'excellence Kaleidoscope [KALEIDOSCOPE 2004] s'est attaché à l'étude de l'analyse de l'usage d'un EIAH dans des contextes de formation en entreprise ou académique, avec l'objectif d'aider les concepteurs enseignants ou formateurs à définir des situations pédagogiques adaptées aux usages observés. Le projet a défini un ensemble structuré de patrons [DPULS 2005b]. Ces derniers formalisent des solutions éprouvées à des problèmes d'analyse des usages d'un EIAH sous une forme réutilisable. Ce projet a également proposé une définition plus large pour le concept d'indicateur qui est vu comme “une variable signifiante sur le plan pédagogique, calculée ou établie à l'aide de données observées, et témoignant de la qualité de l'interaction, de l'activité et de l'apprentissage dans un EIAH”. Comme le montre la figure 2.10, ce projet différencie une “donnée dérivée” (derived-datum), calculée ou établie à l'aide d'autres données, d'une “donnée primaire” (primary-datum). Une donnée primaire est “subjective” (subjective-datum), c'est-à-dire établie ex nihilo par l'analyste de la session, “brute” (raw-datum), c'est-à-dire directement collectée avant, pendant ou après une session d'apprentissage, ou “additionnelle” (additional-datum), utilisée pour établir une donnée dérivée. Une donnée additionnelle peut être d'ordre contextuel (contextualised-datum), disponible avant la session, comme un scénario prédictif, la métadonnée d'une ressource, une taxonomie décrivant le domaine d'apprentissage, ou prédictive (predictive-datum), à produire par les acteurs de la situation d'apprentissage pendant la session, comme la production d'un étudiant à évaluer, le compte-rendu d'activité d'un tuteur.

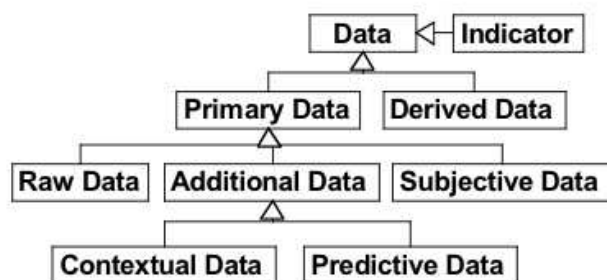


FIGURE 2.10 – Typologie des données du projet DPULS

2.6.4 Positionnement sur la capitalisation

Nous avons vu ici plusieurs approches concernant la capitalisation, ces approches mettent en évidence un besoin clair de partager la connaissance dont nous disposons. Cette connaissance correspond par exemple aux différentes données collectées, c'est-à-dire les corpus de MULCE ou de DataShop qui sont alors modélisés dans un but précis, ou pour des outils particuliers (dans DataShop). L'autre intérêt de la capitalisation correspond au savoir-faire en analyse. Dans une approche orientée fouille de données, le projet DataShop permet d'intégrer et de tester des algorithmes à condition qu'ils soient préparés au format interne de la plate-forme. DPULS de son côté s'est orienté vers des patrons permettant de modéliser les problèmes et solutions liés à l'analyse d'usage d'une façon non automatisable mais à destination d'utilisateurs humains. Ces solutions ne sont pas suffisantes. En effet, afin d'aider les enseignants/concepteurs dans leur activité, il est nécessaire de les prendre en compte en tant qu'acteurs impliqués, or MULCE et DataShop sont clairement à destination des chercheurs. Ce qui signifie que les modélisations adoptées dans ces outils concernent des corpus d'apprentissage dans un spectre beaucoup trop large par rapport aux besoins que nous avons choisis de traiter. Pour DataShop, la transformation entre la base de traces et leur modèle nécessite au préalable une interprétation des traces. De plus, les auteurs expliquent qu'un traitement est nécessaire la nuit, ce qui rend les données analysables uniquement le lendemain. Le défaut principal de DPULS est que les patrons ne sont pas directement opérationnalisables, l'objectif du projet étant le partage de connaissances et la fédération de communautés autour des problèmes d'analyse. Nous considérons qu'une solution pertinente serait une plate-forme du type de DataShop mais à destination des utilisateurs des EIAH. En effet, il serait utile pour les praticiens de disposer d'une banque d'indicateurs exploitables sur leurs données, certes cela nécessite une transformation de leurs traces dans le format de la plate-forme, comme c'est le cas pour DataShop, mais si le résultat est l'exploitation directe de dizaines d'indicateurs, c'est un savoir-faire conséquent et intéressant à exploiter. En ce qui concerne notre proposition de modèle capitalisable de données d'observation, nous nous situons dans une approche orientée utilisateurs et nous souhaitons permettre à l'utilisateur d'exploiter les indicateurs existants pendant ou après la session d'apprentissage. L'utilisation de modèles déclaratifs favorise l'intégration de l'analyse avec l'environnement de formation en lui associant l'interpréteur que nous avons développé. C'est en ce sens que notre approche est intéressante pour la capitalisation.

Chapitre 3

Etude préparatoire : La recherche de généricité

Sommaire

3.1	Introduction	51
3.2	Etude préparatoire sur les traces	51
3.3	Un modèle sémantique pour l'information contenue dans la trace	52
3.3.1	Les objectifs du modèle	52
3.3.2	Présentation du modèle	53
3.4	Principe de fonctionnement	57
3.5	Mise en œuvre	58
3.6	Synthèse	62

3.1 Introduction

De nombreux outils proposent de “tracer” l’activité de leurs utilisateurs, le plus souvent l’activité est mémorisée dans des fichiers de “logs” qui ne possèdent pas de grammaire standardisée. Devant la diversité de ces formats, les applications d’analyse de logs sont le plus souvent dédiées. En plus de la diversité des formats de traçage, et comme nous centrons nos activités sur les EIAH scénarisés, il existe de nombreux modèles de scénarisation pédagogique (par exemple IMS-LD [IMS-LD 2003], LDL [Martel et al. 2006]). Afin de proposer un processus d’ingénierie, nous avons commencé nos travaux par une étude sur la généricité des indicateurs. Ce qui revient à répondre à la troisième question de recherche “*Comment prendre en compte la connaissance métier dans des indicateurs basés sur des traces spécialisées.*”. En effet, souvent les indicateurs étaient développés dans la plate-forme de formation et reposaient sur les choix de conception de cette plate-forme ([Mazza et Milani 2004]). Etant donnée la diversité des plate-formes, il est préférable que les indicateurs soient basés sur les connaissances métier de l’enseignant/concepteur plutôt que sur le métier de la plate-forme.

Dans cette première étude, l’objectif principal est de montrer qu’un indicateur peut être créé de manière générique au sens où il ne dépend pas de la représentation des données au sein d’une plate-forme spécifique. Nous nous sommes intéressés à la sémantique de l’information disponible dans les traces pour construire ces indicateurs. Pour cela, nous avons proposé un langage¹ déclaratif intermédiaire qui permet l’extraction des informations contenues dans les traces de l’EIAH ainsi que leur qualification en fonction des concepts présents dans le scénario pédagogique. Il s’agit d’un langage déclaratif donc aisément interprétable et modifiable par un acteur humain, et d’un langage intermédiaire entre le format de stockage des traces dans la plate-forme et le langage de programmation des indicateurs.

Nous commençons dans ce chapitre par présenter l’étude préparatoire qui nous a amené à proposer ce langage intermédiaire que nous appelons le modèle sémantique pour l’information contenue dans la trace, pour ensuite présenter son principe de fonctionnement et sa mise en œuvre dans un prototype.

3.2 Etude préparatoire sur les traces

Afin de modéliser la sémantique embarquée dans les traces, nous avons commencé par analyser les traces de différentes expérimentations. Tout d’abord, nous avons utilisé une expérimentation réalisée au sein du LIUM [Corbière 2006] qui reposait sur le composant FreeStyleLearning² de la plate-forme “Open University Support System” (Open-USS³). Cette plate-forme génère des traces sous la forme d’un fichier texte structuré. Le concepteur a modélisé son scénario avec le langage de modélisation IMS Learning Design [IMS-LD 2003]. Dans le cadre de travaux pratiques en réseaux menés au département Services et Réseaux de Communication de l’IUT de Laval en 2002-2003, les étudiants

1. Nous utilisons pour représenter le même concept les mots “langage” et “modèle sémantique”

2. <http://www.campussource.de/org/software/openuss/>

3. <http://openuss.sourceforge.net/>

ont utilisé le composant `FreeStyleLearning` afin de comprendre le fonctionnement et de développer un serveur HTTP. Nous avons eu accès à ces traces ainsi qu'à des exemples de besoins d'observation de la part de l'enseignant. Ensuite, nous avons eu accès à une deuxième base de traces provenant d'un supermarché virtuel utilisé par des patients atteints de maladies dégénératives de type Parkinson ou Alzheimer dans le cadre de l'aide au diagnostic et du réapprentissage [Klinger 2006]. En complément de la base de traces, nous avons aussi obtenu les besoins d'observation de la conceptrice de la plate-forme. Ces données couplées aux autres modèles existants (logs, bases de données, etc.) nous ont permis de proposer un modèle sémantique de l'information contenue dans la trace indépendant d'une plate-forme et d'un format de stockage de traces.

3.3 Un modèle sémantique pour l'information contenue dans la trace

L'objectif lié à cette étude est de proposer une modélisation permettant une certaine forme de généralité dans la création d'indicateurs. La généralité se trouve dans le fait que le calcul de l'indicateur ne doit plus reposer sur une plate-forme particulière et devient par conséquent indépendant du format de représentation des traces spécifique à l'EIAH. Nous avons choisi de baser notre modélisation sur la sémantique de l'information. Sans pour autant proposer une ontologie représentant les connaissances métier de l'enseignant/concepteur car il s'agit d'un travail beaucoup trop important pour cette étude préparatoire, nous intégrons dans notre modèle des concepts du métier de l'enseignant/concepteur. Nous aurions pu nous baser sur un langage de scénarisation pédagogique, sauf que le plus souvent ce type de langage est peu ou pas outillé, c'est-à-dire difficile à intégrer dans une plate-forme. Par conséquent, les enseignants n'utilisent pas de langage spécifique de scénarisation, parfois il s'agit tout simplement d'un plan ou d'un organigramme sur une feuille. Nous avons alors choisi de nous baser pour cette étude sur les concepts liés à l'environnement plutôt que le scénario directement, ce qui correspond au vocabulaire des enseignants/concepteurs rencontrés et nécessite par conséquent la modélisation de ces concepts.

3.3.1 Les objectifs du modèle

Pour instrumenter la transformation des traces générées par un dispositif d'apprentissage en données brutes⁴ exprimées dans un format indépendant de ce dispositif d'apprentissage, nous avons décidé de proposer un modèle générique et sémantique permettant de décrire cette transformation [Iksal et Choquet 2005b] [Iksal et al. 2004]. Le modèle est générique au sens où ses descripteurs permettent de prendre en compte divers formats de représentation des traces, et il est sémantique car il apporte une qualification qui fait sens à l'enseignant/concepteur. L'idée fondatrice est que chaque trace d'utilisation prise en considération lors de la modélisation de l'observation concourt à témoigner d'un usage observé. Extraire le contenu de la trace consiste à le mettre en relation avec les concepts

4. Une donnée brute est une donnée significative extraite de la trace, section 2.1

du scénario pédagogique prédictif afin de pouvoir le qualifier et l'utiliser dans l'analyse de la situation d'apprentissage. Le modèle de cette étude préparatoire est centré sur la transformation des traces dans un format indépendant du dispositif d'apprentissage, et sur leur association, en tant que témoignage d'un usage observé, à une représentation du scénario pédagogique indépendante du modèle de scénarisation utilisé par le concepteur. Les méthodes d'analyse permettant d'établir des indicateurs sont, dans le contexte d'utilisation de ce langage, implémentées dans des outils acceptant en entrée des traces exprimées dans ce langage et une représentation d'une partie du scénario pédagogique prédictif indépendante du modèle d'expression pédagogique. Ces outils peuvent alors être combinés entre eux et réutilisés avec d'autres jeux de données, provenant d'autres dispositifs d'apprentissage, et/ou pour aider à l'évaluation d'autres scénarios pédagogiques, éventuellement exprimés selon un modèle d'expression pédagogique différent.

3.3.2 Présentation du modèle

Ce modèle ne considère que des traces collectées de manière automatique, il ne prend donc pas en compte les enregistrements sonores ou audiovisuels, ni les observations manuscrites. Afin d'anticiper sur les autres questions de recherche de notre problématique, nous avons décidé de proposer un modèle sous la forme d'un langage déclaratif. Cela facilite d'une part sa spécification en fonction des besoins de l'enseignant/concepteur et d'autre part sa capitalisation et sa réutilisation. Nous avons fait le choix de traduire le modèle sous la forme d'un schéma XML ce qui lui permet d'être utilisé avec tout méta-modèle d'expression pédagogique disposant d'un "binding" XML et avec tout type de format de traces collectées automatiquement par un dispositif d'apprentissage. De plus, l'utilisation du langage XML nous permet de ne pas nous restreindre à un choix d'architecture, aujourd'hui il est possible d'intégrer et d'importer des données XML dans la plupart des outils de gestion de données. Ce choix technique reposait sur un prototypage rapide avec les outils que nous utilisons couramment, tout en conservant la possibilité de transformer nos modèles et données si nécessaire dans d'autres langages comme RDF/RDFS [Lassila et Swick 1999] ou OWL [Dean et al. 2002].

Cette sous-section présente les modèles d'information des deux éléments constituant le langage. Tout d'abord, nous nous sommes intéressés à la représentation d'une partie du modèle de scénarisation. En effet, suite aux travaux de Vincent Barré [Barré et al. 2004], nous avons identifié la notion de *concept traçable* dont le modèle d'information est représenté dans le paragraphe 3.3.2.1. Un concept traçable est un concept du scénario qu'il est possible d'observer dans l'EIAH. Cela nécessite d'avoir identifié au préalable dans l'environnement ce qui peut être observé. A chaque concept traçable du scénario pédagogique, il est possible d'associer un ou plusieurs *usages observés*. L'usage observé correspond à la nature de ce qui est observé, cela fait partie de la sémantique que nous intégrons au modèle. En effet, pour un même concept du scénario, il est possible d'observer plusieurs usages. L'usage correspond à ce que les usagers font avec l'instance du concept dans le scénario. Par exemple, pour le concept traçable "Etude de texte" de type "Ressource", nous pouvons avoir comme usage observé le "chargement d'une page" ou encore le "défilement dans une page". Ce qui signifie que l'on peut observer le chargement des différentes

pages de la ressource. Un indicateur possible serait par exemple : “Est-ce que toutes les pages de la ressource ont été chargées par l’apprenant X?”. Le choix de définir un concept traçable vient du fait que la conception du scénario ne passe pas systématiquement par l’utilisation d’un langage spécifique, mais que le concepteur est capable d’identifier les concepts importants de son scénario. De plus, lorsqu’il existe un lien associatif direct avec un langage de scénarisation, l’utilisation du concept propre au vocabulaire de l’enseignant/concepteur permet par exemple d’y associer la terminologie d’autres langages de scénarisation et donc de faciliter la réutilisation. Cette association pourra reposer sur l’utilisation d’une ontologie représentant les langages de scénarisation afin d’améliorer la sémantique de l’information.

Ensuite, nous nous sommes intéressés à la représentation de la transformée des traces générées par le dispositif d’apprentissage. Elle permet de décrire les traces comme un ensemble de données, identifiées chacune par une méthode d’accès (ou d’extraction) dans les traces générées par l’EIAH. Des exemples de ces opérations de transformation sont donnés dans la sous-section 3.5. Comme nous l’avons indiqué, les traces représentées ici sont toutes dans un format textuel (Fichier texte, base de donnée, fichier XML). Le modèle d’information de cette représentation de la trace est représenté par le paragraphe 3.3.2.2.

Nous avons adopté un formalisme analogue à celui du modèle d’information de IMS Learning Design. Ce formalisme sera utilisé aussi dans le chapitre 5.

1. Seuls les éléments sont présents (pas les attributs).
2. Les diagrammes sont des structures de type arbre, à lire de gauche à droite. Un élément situé à gauche contient les éléments situés à droite. L’élément le plus à gauche est la racine de l’arbre, c’est-à-dire la donnée modélisée.
3. Une relation OU (exclusif) est représentée par le symbole “<”.
4. Une relation ET est représentée par le symbole “[”.
5. Un élément participant zéro ou n fois à la relation est précédé par “*”.
6. Un élément participant au moins une fois à la relation est précédé par “+”.
7. Un élément participant au plus une fois à la relation est précédé par “?”.
8. Quand aucun des signes *, + ou ? n’est utilisé, c’est que l’élément participe exactement une fois à la relation.

3.3.2.1 La représentation du concept traçable

Au travers du concept traçable, nous établissons une sorte d’ontologie représentant le modèle métier d’un enseignant/concepteur, celui pour lequel l’étude a été menée. L’idée ici est de s’appuyer sur cette représentation pour qualifier le contenu des traces. Nous différencions dans ce modèle le “*concept*” et le “*concept traçable*” (Traceable-concept dans la figure 3.1), en effet, nous considérons le “*concept*” à un niveau plus abstrait que le “*concept traçable*”. Par exemple, si l’on reprend la terminologie d’IMS-LD [IMS-LD 2003], les termes “*Resource*” ou “*Activity*” seront des concepts, et dans un scénario “*Vidéo introductive*” sera le “*concept traçable*” associé au “*concept*” “*Resource*”. Nous avons fait ce choix de modélisation car toutes les instances d’un même concept ne sont pas observables de la

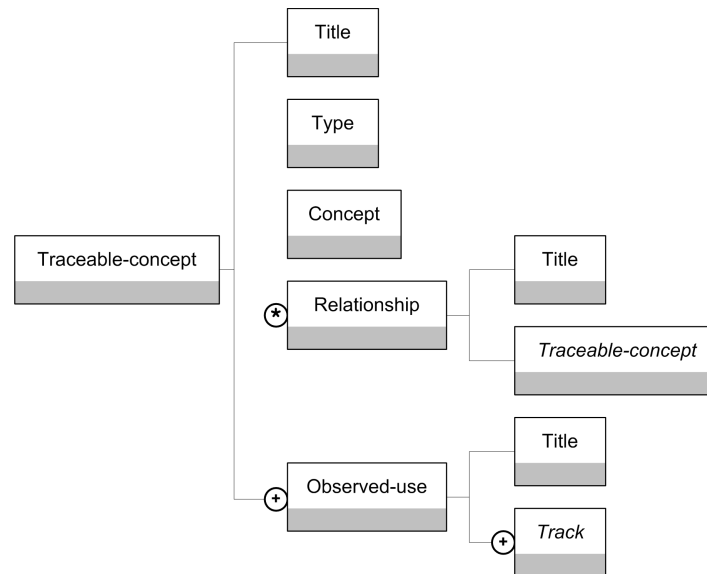


FIGURE 3.1 – Le modèle d’information d’un concept traçable

même façon. Nous n’avons pas les mêmes usages observés sur une ressource de type texte et sur une ressource de type vidéo.

De la même façon, nous avons choisi de relier les concepts traçables entre eux en nommant (Title) cette relation (Relationship), afin d’être au plus proche du vocabulaire utilisé par l’enseignant/concepteur. Par exemple, plutôt que de relier le concept “Activité” avec le concept “Ressource”, nous avons relié le concept traçable “Vidéo introductive” avec le concept traçable “Voir les objectifs”.

Afin de distinguer les concepts venant de la scénarisation et ceux venant de la plate-forme, nous avons ajouté la caractéristique “*Type*”. En effet, ici deux types de connaissances métier peuvent être exploitées par l’enseignant/concepteur. Nous avons souhaité les distinguer pour deux raisons. Tout d’abord, l’enseignant/concepteur peut scénariser à partir d’un modèle de scénarisation existant (un EML - Educational Modeling Language) ou à partir de son propre modèle de représentation qui, le plus souvent, est une représentation mentale qu’il se fait de la plate-forme qu’il utilise. Ensuite, dans le cadre de la création d’un indicateur cette distinction pourrait être utile, notamment pour faire le lien entre un modèle de scénarisation et la plate-forme.

L’élément *titre* (title) permet de nommer le concept traçable.

Enfin, l’objectif de ce modèle est d’apporter de la sémantique à l’information contenue dans les traces, c’est pour cela que nous avons intégré la notion d’“*usage observé*” (Observed-use). L’idée ici est de considérer pour chaque concept traçable qu’il est possible de tracer différents usages. Par exemple les réponses d’un étudiant à un QCM ou, comme dans l’exemple présenté dans la sous-section 3.5, le pilotage interactif d’une ressource vidéo. Pour chaque usage, nous indiquons l’ensemble des traces permettant de le caractériser. Il s’agit donc de l’ensemble des traces utilisées par l’outil informatique permettant de calculer le ou les indicateurs liés à l’usage. Nous rappelons que dans le cadre de cette étude préparatoire, le calcul des indicateurs était directement écrit dans un lan-

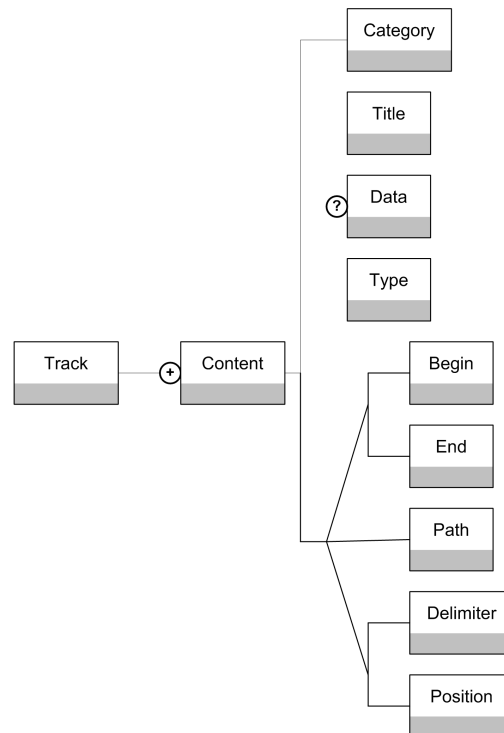


FIGURE 3.2 – Le modèle d’information d’une trace

gage de programmation. Dans ce contexte, les indicateurs nécessitent des compétences informatiques fortes afin d’être créés et/ou modifiés. Le “*titre*” (Title) de l’usage observé permet de le nommer, les éléments de représentation des “*traces*” (Track) sont décrits dans le paragraphe suivant.

3.3.2.2 La représentation de la trace

Une trace est généralement composée de plusieurs informations dont seules certaines sont riches en signification pour le concepteur. Cette partie du modèle a pour objectif de décrire la méthode d’extraction de ces informations. Il s’agit d’une solution générique au sens où elle est indépendante d’une modélisation spécifique liée à une plate-forme éducative ou un serveur web, par contre l’instance quant à elle sera bien spécifique. Nous proposons une implémentation qui est compatible avec la majorité des formats de traces collectées de manière automatique (textuel ou base de données).

Une trace est structurée avec des symboles ou expressions fixes et des expressions variables. Si nous prenons l’exemple suivant tiré de FreeStyleLearning :

```
[04/12/2002:03:18:55 +0962] [FreeVideoPlayer] start() currentTime=0s
[04/12/2002:03:21:58 +0775] [FreeVideoPlayer] stop() currentTime=182.0s
```

Nous constatons que le mot “*FreeVideoPlayer*” et le mot “*currentTime*” apparaissent systématiquement, nous pouvons en déduire que ces expressions vont permettre d’identifier ces traces dans un ensemble plus conséquent. FreeVideoPlayer fait clairement référence à

un lecteur vidéo utilisé dans l'environnement de formation. Ces traces contiennent aussi les mots “*start*” et “*stop*”, ce qui signifie que ces traces représentent l'usage d'une vidéo avec l'instant de démarrage et l'instant d'arrêt de la lecture. Au niveau des expressions variables, nous pouvons identifier la date et l'heure de l'action ainsi que le positionnement dans la ligne de temps de la vidéo.

Nous avons donc des informations fixes permettant d'identifier la trace, et des informations variables qui constituent les données sur lesquelles nous pouvons construire des indicateurs. Nous nous servons de ces constatations pour proposer une représentation de la trace basée sur des mots clés (les expressions fixes) et des valeurs (les expressions variables).

La description d'une *trace* (Track) se décompose alors en objets de deux *catégories* (Category) : les *mots clés* (Keyword) et les *valeurs* (Value). Nous avons souhaité intégrer le fait que les informations puissent être dans des fichiers structurés de type textuels, mais aussi XML ou dans des bases de données, ces outils disposent de langages de requête que nous utiliserons pour spécifier l'accès à l'information. En ce qui concerne les fichiers textuels plus classiques, nous nous basons sur des séparateurs ou la position exacte de l'information. Cette solution permet de répondre à une quantité importante de formats de traces.

Le *titre* (Title) d'une trace désigne le contenu et lui associe un sens (par exemple : “date de début d'activité”, “réponse au QCM”).

Le champ *donnée* (Data) est utilisé pour stocker la valeur ou le mot clé.

Le type correspond au type de stockage de l'information. Il peut prendre pour valeur “Text”, “XML” ou “Database”. Cette liste est ouverte, à condition de disposer d'un langage de requête.

Pour une trace de type “Database” ou “XML”, le *chemin* (Path) contient le chemin d'accès à la donnée, soit une requête XPath ou SQL.

En ce qui concerne les fichiers de “logs” sous forme de fichiers texte structurés, nous utilisons un ensemble de champs permettant d'indiquer la localisation de la donnée dans une chaîne de caractères, soit par la position des caractères, soit en utilisant des délimiteurs.

- *Début* (Begin) donne la position du premier caractère du contenu.
- *Fin* (End) donne la position du dernier caractère (-1 pour la fin de la ligne).
- *Délimiteur* (Delimiter) correspond au délimiteur utilisé pour séparer la chaîne de caractères en différentes sections.
- *Position* donne la position de la section de chaîne de caractères à conserver.

3.4 Principe de fonctionnement

Ce langage a été conçu de façon essentiellement “informatique”, cela signifie que le travail repose essentiellement sur un informaticien. Il s'agit d'un travail préparatoire et en aucun cas une réponse adaptée aux usagers. L'enseignant/concepteur rédige ses besoins d'observation dans des fichiers textes classiques éventuellement complétés par des graphiques. Si le besoin d'observation est relativement simple, le développeur joue le rôle d'un analyste de données, sinon il est nécessaire de faire appel à un analyste de données expérimenté afin de décrire le moyen d'observation, c'est-à-dire la façon dont l'indicateur

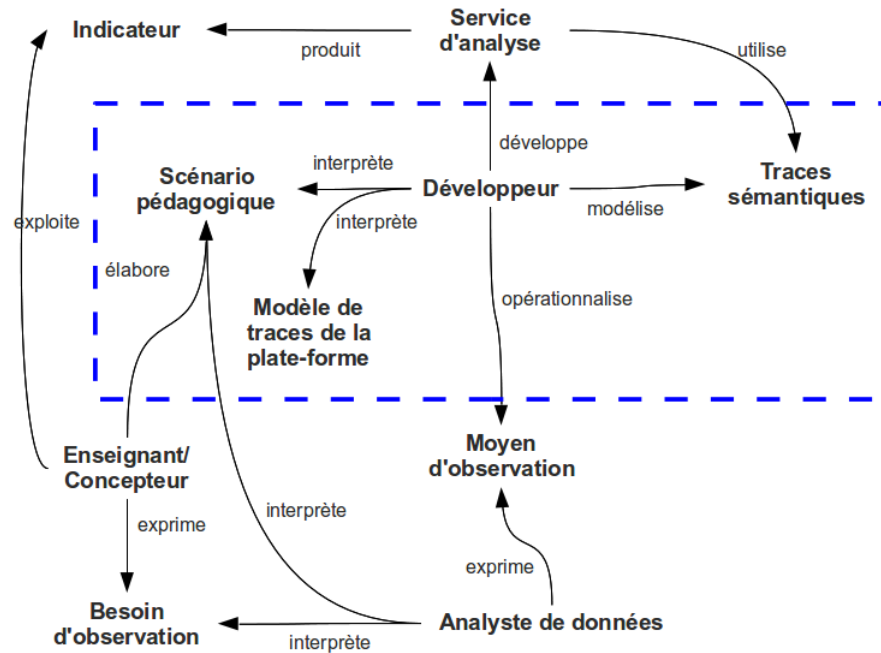


FIGURE 3.3 – Le processus d'utilisation d'UTL 1

doit être calculé ainsi que l'ensemble des informations nécessaires au calcul. A nouveau ce travail est fait sans formalisme particulier. Enfin, le développeur est responsable de la tâche importante de modélisation des traces et des concepts traçables. Ensuite, il implémente les indicateurs sous la forme par exemple de services d'analyse. La figure 3.3 décrit ce fonctionnement.

3.5 Mise en œuvre

Cette section présente un exemple d'utilisation de notre langage sur l'expérimentation consacrée à la conception, au déploiement, à l'analyse et à la réingénierie d'un dispositif d'apprentissage de la programmation d'un serveur HTTP basée sur FreeStyleLearning et la plate-forme Open-USS. Afin de valider notre proposition, nous avons aussi déployé notre modèle sur les traces et les indicateurs de deux environnements externes à notre laboratoire. Tout d'abord, nous avons utilisé les informations du supermarché virtuel pour le diagnostic et le réapprentissage de patients atteints des maladies de Parkinson et d'Alzheimer [Klinger 2006]. Ensuite, nous avons collaboré avec le laboratoire SYSCOM et utilisé l'une de leurs expérimentations sur l'apprentissage de la modélisation UML [Iksal et al. 2007].

Pour cet exemple de mise en œuvre, nous avons :

1. défini la transformation de IMS-LD en identifiant, avec le concepteur, les concepts de ce langage de scénarisation dont au moins une des instances était susceptible de devenir un *concept traçable*,
2. utilisé cette transformation pour décrire les concepts traçables du scénario pédago-

- gique prédictif défini par le concepteur,
3. représenté les traces dans notre modèle en les situant dans les fichiers de logs générés par l’environnement FreeStyleLearning et la plate-forme Open-USS,
 4. défini des outils génériques d’analyse permettant d’établir des informations significatives pour le concepteur, à partir des représentations UTL du scénario pédagogique et des traces.

Le scénario prédictif de l’enseignant/concepteur est composé de six activités et d’un certain nombre de ressources pour réaliser ces activités. Les activités sont les suivantes : Voir les objectifs, approfondir les concepts, synthétiser les concepts, s’exercer pour comprendre, s’auto-évaluer et apprendre en faisant [Corbière et Choquet 2005]. La particularité de FreeStyleLearning est que cet environnement laisse les apprenants libres dans leur parcours des activités.

```

...
<Activity Title="Voir les Objectifs" Type="Abstract-scenario">
<Relationship Title="Use" Concept="VideoIntro"/>
</Activity>
...
<Resource Title="VideoIntro">
<ObservedUse Title="Managing">
...
<Track Title="Stop video">
<Content Category="Value" Title="Date" Type="Text" Begin="1" End="26"/>
<Content Category="Keyword" Title="Task" Type="Text" Begin="38" End="53">
  FreeVideoPlayer</Content>
<Content Category="Keyword" Title="Action" Type="Text" Begin="55" End="59">
  stop</Content>
<Content Category="Value" Title="Time" Type="Text" Begin="74" End="-1"/>
</Track>
</ObservedUse>
</Resource>

```

FIGURE 3.4 – Extrait de la modélisation des traces provenant de FSL

Dans le cadre de l’expérimentation, il a été constaté notamment qu’une partie importante des apprenants ne consultait pas la vidéo introductive dans son intégralité. Nous choisissons de présenter l’exemple de la vidéo introductive pour la mise en œuvre de notre modèle. Au regard du scénario prédictif, les concepts “*activity*” et “*learning object*” d’IMS-LD ont été identifiés comme possédant des instances qui étaient des concepts traçables. Il s’agit de l’activité “*Voir les objectifs*” consistant en la visualisation par l’apprenant d’une ressource vidéo d’introduction aux objectifs d’apprentissage (“*VideoIntro*”).

Ces concepts traçables sont représentés sous la forme d’un fichier XML dont un extrait est présenté dans la figure 3.4. L’activité “*Voir les Objectifs*” est un concept traçable ; elle a une relation d’“*utilisation*” avec le concept traçable “*VideoIntro*” qui est de type “*LearningObject*”. Cette ressource est caractérisée par un usage observé “*utilisation du*

player vidéo”, contenant plusieurs éléments de type trace, permettant de localiser les données significatives concernant l’utilisation du player vidéo dans le fichier de “logs” généré par le dispositif d’apprentissage déployé sur FSL.

Dans cet exemple, nous avons la description d’une trace correspondant à l’arrêt de la lecture de la vidéo (cf. Figure 3.6). La trace est composée de deux mots clés, l’un représentant la tâche de lecture d’une vidéo et l’autre représentant l’action de stopper la lecture. Il s’agit de données textuelles pour lesquelles nous avons les positions précises dans la ligne du fichier de “logs”. Ensuite, nous avons deux valeurs à collecter, tout d’abord la date et l’heure de l’action, puis l’instant d’arrêt sur la ligne de temps de la vidéo. Cette description permet alors de parcourir le fichier de “logs” de FSL afin d’en extraire les données d’observation relatives à un usage observé, mis en corrélation avec un concept traçable du scénario prédictif. Ainsi, sur l’exemple de traces de la figure 3.5, nous avons utilisé ce modèle.

```
[04/12/2002:03:18:55 +0952] [FreeApp] Intro gestartet
[04/12/2002:03:18:55 +0962] [FreeVideoPlayer] start() currentTime=0s
[04/12/2002:03:21:58 +0775] [FreeVideoPlayer] stop() currentTime=182.0s
[04/12/2002:03:22:38 +0982] [FreeVideoPlayer] pause() currentTime=0.0s
[04/12/2002:03:22:39 +0002] [FreeTextStudyManager] Standard-Init der Textstudy
[04/12/2002:03:22:39 +0012] [FreeTextStudyManager] in showactualPage vor if mit
ende = false & shownNode: Heberger soi-même son site
[04/12/2002:03:22:39 +0022] [FreeNotesManager] [Store] Notiz zu Réaliser un
serveur HTTP.Etude de textes.Heberger soi-même son site vorhanden?
[04/12/2002:03:22:39 +0032] [FreeNotesManager] Nein.
[04/12/2002:03:22:39 +0032] [FreeNotesManager] [NotesManager] NoteButton
umschalten auf: Existiert nicht!
[04/12/2002:03:22:39 +0052] [FreeTextStudyManager] LOAD FILE
[04/12/2002:03:22:39 +0072] [FreeTextStudyManager] Lade TS-Datei: D:\FSL\
Granulat\services\txtStudy\txt-0.htm
[04/12/2002:03:22:39 +0082] [FreeLinkManager] Setze LinkViewButton für Topic:
txtStudy.Heberger soi-même son site auf false
[04/12/2002:03:22:39 +0153] [FreeApp] textStudy gestartet
```

FIGURE 3.5 – Extrait de traces provenant de FSL

```
[04/12/2002:03:21:58 +0775] [FreeVideoPlayer] stop() currentTime=182.0s
```

FIGURE 3.6 – Trace FSL d’arrêt de lecture de la vidéo

Ce qui nous a permis d’extraire la trace de la figure 3.6 et d’obtenir les données suivantes :

```
Date of the track: 04/12/2002:03:21:58
Duration of the video: 182.0s
```

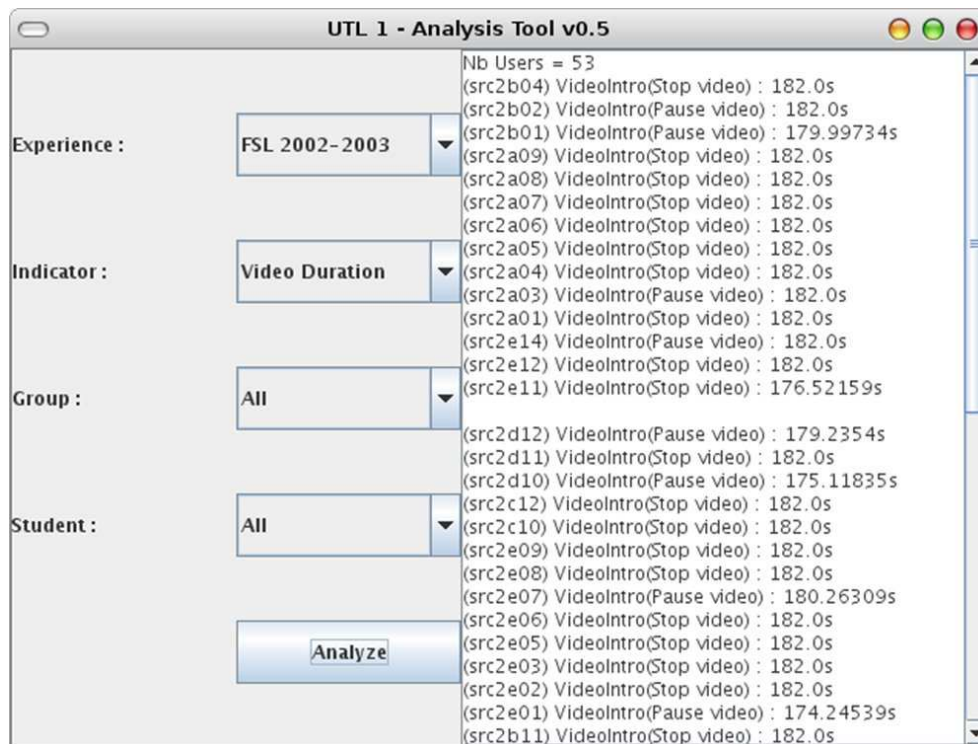


FIGURE 3.7 – Interface de consultation des indicateurs

Ces données ne sont pas directement exploitables par l’enseignant/concepteur sous cette forme. L’approche développée dans cette étude préparatoire permet néanmoins de corréler des éléments de traces générées par le dispositif d’apprentissage avec des aspects du scénario pédagogique prédictif. Ici, les données indiquant l’instant où chaque étudiant a arrêté la vidéo sont reliées à la ressource “VideoIntro” (élément “Track”), elle-même utilisée par l’activité “Voir les Objectifs” (élément “Relationship”).

Dans l’extrait qui suit, nous obtenons le résultat produit par un outil d’analyse (cf. Figure 3.7) permettant la génération de l’indicateur “nombre d’étudiants qui n’ont pas terminé la vidéo d’introduction”. Les fonctions développées dans cet outil peuvent être enrichies ou réutilisées dans d’autres applications, elles reposent sur un format unique indépendant du format des traces générées par les dispositifs d’apprentissage. En effet, elles utilisent directement les valeurs obtenues après transformation.

```
Nb Users = 53
(src2b04) VideoIntro(Stop video) : 182.0s
(src2b02) VideoIntro(Stop video) : 182.0s
(src2b01) VideoIntro(Stop video) : 179.99734s
...
-----
Number of Students with video not completed : 11
```

Sur la base de ce résultat, la ressource “VideoIntro” a été modifiée par l’enseignant, de manière à ne pas prendre en compte les dernières secondes de la vidéo, dont le contenu n’est pas signifiant.

3.6 Synthèse

Notre objectif dans cette étude préparatoire était de voir s'il était possible de s'abstraire du format de représentation des traces dans les EIAH pour baser la création des indicateurs sur la sémantique des informations en prenant en compte les connaissances métier de l'enseignant/concepteur. C'est ce que nous avons montré dans un contexte d'EIAH scénarisés dont les traces sont générées dans des fichiers textuels (structurés ou non) ou des bases de données. Il s'agit d'une approche différente de celles existantes qui intègrent la collecte des données d'observation directement dans leur système de traçage et qui se retrouvent fortement couplées aux EIAH existants.

Cette étude très technocentrée sert de point de départ à la définition d'un processus plus complet autour de l'observation. Il est nécessaire de se rapprocher de l'utilisateur afin de spécifier au mieux ses besoins d'observation tant du point de vue de leur calcul que de leur manipulation et de leur interprétation. C'est pour cela que nous vous présentons dans le chapitre suivant le processus d'ingénierie de l'observation.

Chapitre 4

L'ingénierie de l'observation par la prescription

Sommaire

4.1	Introduction	65
4.2	Présentation de notre approche	66
4.3	La chaîne éditoriale de l'observation prescrite	67
4.3.1	Les rôles concernés	67
4.3.2	Description de la chaîne éditoriale	69
4.4	Un enjeu important : La capitalisation	71
4.5	Les objectifs d'observation	73
4.6	Scénarios d'utilisation	73
4.6.1	Un exemple en réingénierie	74
4.6.2	Un exemple en régulation	77
4.7	Synthèse	78

4.1 Introduction

La question de recherche générale que nous avons choisie de traiter est la suivante : *“Comment aider l’enseignant/concepteur à améliorer sa situation d’apprentissage grâce à l’observation de cette dernière ?”*. Pour cela, nous nous sommes intéressés au processus d’observation et plus particulièrement, nous avons défini un cadre méthodologique de l’ingénierie de l’observation par la prescription. En effet, les travaux dans le domaine de l’observation traitent le plus souvent du processus d’analyse uniquement [Romero et al. 2010]. Ce processus est certes important, car il définit la manière dont sont calculés les indicateurs, mais le plus souvent ces indicateurs sont “subis” par les usagers. Nous avons choisi de nous intéresser au processus d’ingénierie de l’observation qui va de l’agent humain exprimant son besoin d’observation jusqu’à l’agent humain (le plus souvent, il s’agit de la même personne) qui exploite le résultat de l’observation. Nous envisageons donc l’observation au sens large, au travers d’un cycle d’ingénierie/réingénierie qui s’étend de l’expression du besoin jusqu’à l’opérationnalisation et le retour d’expérience sur l’observation. On parle alors d’ingénierie de l’observation.

Nous avons situé notre activité dans le cadre du projet REDiM qui lui-même se focalisait sur la ré-ingénierie des EIAH du point de vue du scénario. En effet, en reprenant la description du projet telle qu’elle a été définie par Christophe Choquet [Choquet 2007], ce projet avait pour objectif principal d’intégrer les enseignants et les formateurs dans le processus de conception d’un EIAH. Nous considérons donc des situations d’ingénierie et de réingénierie d’un EIAH où le concepteur est un enseignant ou un formateur, et non pas un chercheur de la communauté EIAH ou un spécialiste du développement d’un EIAH. Dans notre processus d’ingénierie de l’observation, nous considérons que l’enseignant est un prescripteur de l’observation de sa situation d’apprentissage au même titre que le tuteur, le chercheur en EIAH voire l’apprenant. En suivant le paradigme du projet qui consiste à se positionner sur les EIAH scénarisés, nous avons également pour objectif de scénariser l’observation avec des modèles directement manipulables et interprétables par les concepteurs enseignants ou formateurs.

Dans le cadre d’un cycle d’ingénierie/ré-ingénierie de l’observation, nous avons considéré qu’il était important de prendre en compte les aspects liés à la capitalisation, la réutilisation et l’automatisation de l’observation afin de fédérer des communautés de pratiques. En effet, de nombreux projets s’intéressent de manière ad hoc à l’observation pour des cas bien précis (par exemple, les traces d’interaction dans TraViS [May 2009], Tatiana [Dyke 2009], CAViCoLA [CAViCoLA 2006]), or aujourd’hui ce qui est un enjeu primordial, c’est d’économiser le temps des usagers ou faciliter leur activité en partageant le savoir-faire dont ils disposent, et cela est vrai aussi pour l’observation. Prenons par exemple le cas d’un enseignant/concepteur débutant, s’il dispose d’un accès à une banque d’indicateurs compatibles avec son EIAH, il peut facilement scénariser l’observation de sa situation d’apprentissage. La notion d’indicateurs compatibles dans cet exemple, correspond à la mise en correspondance entre les informations servant de base aux indicateurs et celles disponibles dans les traces de l’EIAH.

Nous proposons dans ce chapitre un processus d’ingénierie de l’observation en EIAH par la prescription qui intègre les acteurs humains de l’EIAH dès la réflexion autour de

l'observation.

4.2 Présentation de notre approche

De notre point de vue, pour que l'observation soit la plus proche de ce que recherche l'utilisateur, il est nécessaire d'impliquer ce dernier dans le processus d'ingénierie. Nous considérons donc l'utilisateur au cœur du processus d'ingénierie de l'observation. Contrairement aux autres approches pour lesquelles les solutions de suivi et d'analyse de traces proposées sont toujours composées soit de solutions de type "fouilles de données", soit de solutions "clé en main" et qui le plus souvent ne satisfont pas tous les usagers, nous avons choisi de baser notre analyse sur l'expression du besoin de l'utilisateur. La seconde contrainte que nous souhaitons prendre en compte correspond aux aspects liés à la capitalisation et la réutilisation du savoir-faire mis en œuvre.

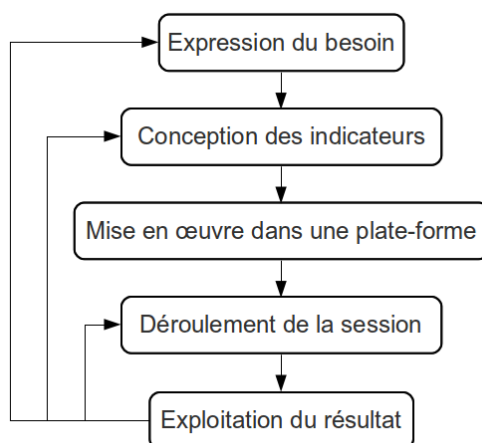


FIGURE 4.1 – Processus d'ingénierie de l'observation

Notre processus d'ingénierie de l'observation repose sur cinq phases (cf. Figure 4.1). **La phase d'expression du besoin** (ou prescription) correspond à la définition de ce qu'il est souhaitable d'observer. Cette phase amène à **la phase de conception des indicateurs**. Lors de la phase de conception, il peut être mis en évidence l'impossibilité logique, matérielle ou humaine d'obtenir des observés nécessaires au calcul d'un indicateur. La conception consiste en la formalisation d'un indicateur de manière à le rendre calculable automatiquement à partir des observés. Ensuite, nous avons **la phase de mise en œuvre**, c'est-à-dire la phase d'intégration de l'indicateur avec la plate-forme. Ce qui signifie être capable d'interagir avec l'environnement afin d'avoir accès aux données collectées, mais aussi dans la mesure du possible permettre la présentation du résultat au destinataire directement dans l'EIAH, si cette diffusion est prévue par le scénario d'apprentissage. Une fois cette phase réalisée, il ne reste plus qu'à **dérouler la situation d'apprentissage** et déclencher, lorsque cela s'avère nécessaire, le calcul des indicateurs. Enfin, la dernière phase concerne **l'exploitation du résultat**. Il s'agit à ce niveau : (1) de proposer le résultat de l'indicateur aux destinataires durant le déroulement de la session ; (2) d'exploiter le

résultat dans une phase de réingénierie des indicateurs au niveau de leur conception ; ou (3) de capitaliser la modélisation et le résultat pour l'ingénierie et la réingénierie des scénarios d'apprentissage et d'observation.

Afin de mettre en œuvre ce processus, nous avons souhaité transposer l'approche de composition sémantique [Garlatti et al. 2004] aux traces. En effet, les aspects déclaratifs et génériques du moteur de composition nous ont semblé être une réponse satisfaisante aux problèmes d'automatisation, de réutilisabilité et de capitalisation. Nous avons alors considéré la formalisation du besoin d'observation des prescripteurs comme des documents génériques qui sont ensuite instanciables en fonction de l'environnement. Nous proposons une chaîne éditoriale de l'observation basée sur la prescription (la description est proposée dans la section suivante). Le principe de cette chaîne est de prendre en considération l'observation dès la phase de conception de l'activité pédagogique, jusqu'à son exploitation dans un processus de ré-ingénierie.

4.3 La chaîne éditoriale de l'observation prescrite

4.3.1 Les rôles concernés

Avant de présenter en détail la chaîne éditoriale, il est important de présenter les rôles impliqués dans ce processus. En effet, nous avons identifié trois rôles importants dans notre approche prescriptive de l'observation : le prescripteur, l'analyste et le développeur. Cette séparation est liée aux activités des différents acteurs et aux compétences nécessaires à chaque étape.

4.3.1.1 Le prescripteur

Notre approche se place du point de vue de la personnalisation de l'observation, au sens où nous ne souhaitons pas imposer des indicateurs généralistes "prêts à l'emploi" mais proposer aux usagers de l'observation des indicateurs qui reflètent exactement l'information dont ils ont besoin, dans la forme qui leur fait sens. Dans ce cadre, le prescripteur correspond au rôle le plus important dans notre approche. Il est à la fois le demandeur et le plus souvent le destinataire des indicateurs. Le prescripteur conçoit le scénario d'observation, à ce titre il spécifie ce qu'il est important d'observer, le moment de l'observation ainsi que le destinataire en fonction du scénario d'apprentissage. L'intérêt de baser notre approche d'analyse sur le prescripteur est que ce dernier peut expliquer l'objectif des différentes étapes de son scénario. Il sait précisément ce qu'il compte faire de l'observation. Nous pouvons être tentés de considérer que le prescripteur est l'enseignant, or différents acteurs de la situation d'apprentissage peuvent se placer dans ce rôle, c'est le cas :

- du concepteur pédagogique ou de l'enseignant qui élabore le scénario et qui souhaite observer l'usage des ressources ou l'organisation des apprenants. Il peut aussi souhaiter bénéficier d'indicateurs pour évaluer les apprenants ou son scénario afin de l'améliorer etc.. Il arrive aussi que l'enseignant souhaite mettre à la disposition de ses apprenants des informations sur leurs activités ou celles des autres afin de s'auto-corriger ou de se poser les bonnes questions.

- du développeur de la plate-forme de formation qui peut souhaiter suivre l'usage des différents composants.
- du tuteur qui assiste l'enseignant dans le déroulement de la situation d'apprentissage et qui souhaite évaluer le bénéfice de ses actions de remédiation.
- de l'étudiant qui souhaiterait observer son propre comportement, son propre apprentissage. Il s'agit ici d'une analyse réflexive de l'activité des apprenants.
- de l'administration qui a besoin de mesurer l'impact et l'utilisabilité des moyens mis en œuvre (taux de fréquentation, etc.).
- ou encore du chercheur qui souhaite évaluer l'efficacité des théories qu'il propose.

Tous ces acteurs peuvent justifier d'un besoin d'observation avec à chaque fois un objectif clairement identifié.

Un besoin d'observation n'est pas une donnée très formalisée, mais il s'agit d'indiquer ce que le prescripteur souhaite observer, pourquoi il le souhaite, comment il veut que cela soit présenté et à qui, éventuellement ce qu'il compte en faire et la solution qu'il compte apporter s'il y a lieu d'en apporter une.

4.3.1.2 L'analyste

L'analyste est la personne disposant de suffisamment de connaissances en traitement de l'information pour pouvoir organiser les données, et spécifier le moyen d'observation (par exemple la formule de calcul des indicateurs). Son activité consiste à assister le prescripteur en transformant les besoins d'observation en indicateurs. Cela implique de maîtriser les informations disponibles dans les traces (c'est-à-dire les observés) et de formaliser aussi toutes les données nécessaires au calcul des indicateurs. Il doit interagir avec le prescripteur au sujet de la compréhension et de l'interprétation de ses besoins, mais aussi avec le développeur afin de savoir ce qui est tracé et ce qui peut l'être (moyennant éventuellement du développement) dans l'environnement de formation.

L'analyste est amené à utiliser la liste des observés possibles dans l'environnement de formation. Il s'agit ici des observés qu'il est possible de collecter grâce à l'EIAH, cela correspond à une liste étendue par rapport aux observés disponibles. En effet, les observés disponibles sont directement exploitables alors que d'autres observés nécessitent l'intervention d'un développeur soit pour activer leur traçage, soit pour développer la sonde permettant de les obtenir. Cette liste est le plus souvent fournie par le développeur. Il identifie ensuite les observés dont il a besoin. Quatre options sont possibles :

- (1) les observés répondent à tous les besoins.
- (2) l'analyste ne peut pas tout modéliser à partir des observés existants :
 - (2A) le développeur peut adapter l'environnement afin d'ajouter des sondes.
 - (2B) le développeur ne peut pas adapter l'environnement, mais il est possible d'intégrer des agents humains d'observation.
 - (2C) il est impossible d'observer certains éléments, l'analyste interagit avec le prescripteur afin d'essayer de trouver une solution, qui peut être, par exemple, l'abandon d'un besoin d'observation.

Ensuite, l'analyste combine ces données avec les outils mis à sa disposition pour formaliser le moyen d'observation (le calcul des indicateurs).

Le rôle d'analyste est primordial pour la création des premiers indicateurs. Dès lors

que la base d'indicateurs disponibles sera suffisamment importante, la réutilisation prendra plus d'ampleur que la conception de nouveaux indicateurs. Ce qui fait que le prescripteur gagnera en autonomie dans le processus d'observation.

4.3.1.3 Le développeur

La présence d'un développeur dans les rôles liés à l'observation est nécessaire lorsque la mise en place de l'observation suppose d'intervenir sur l'environnement ou de concevoir des outils informatiques supplémentaires (par exemple pour l'implémentation d'algorithmes complexes lors des calculs d'indicateurs). Le développeur est celui qui peut intervenir dans l'environnement technique de la formation (en EIAH, il s'agit de la plate-forme de formation). Ce dernier est capable d'analyser le fonctionnement de la plate-forme et de dresser la liste de ce qui est traçable par défaut, de ce qu'il peut tracer en intervenant sur le code source de la plate-forme (si ce code source est ouvert), et de dresser la liste de ce qui n'est pas traçable. A partir de ces renseignements, il peut soutenir le travail de l'analyste et débloquer si nécessaire certaines situations.

4.3.2 Description de la chaîne éditoriale

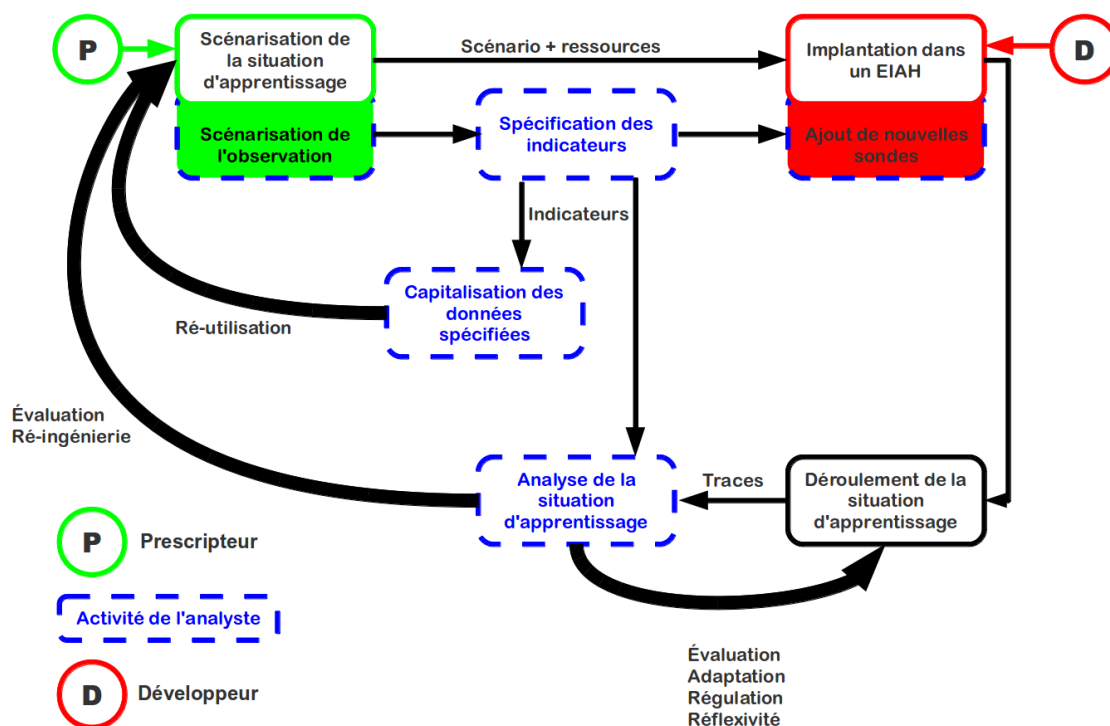


FIGURE 4.2 – Chaîne éditoriale de l'observation

Dans les situations d'apprentissage médiatisées, l'enseignant est aussi concepteur au sens où il est amené à scénariser, de façon formelle ou non, le déroulement de sa situation d'apprentissage. Ce scénario est ensuite implanté dans un environnement, soit directement

importé lorsque le langage de scénarisation est directement assimilable par l'EIAH, soit l'environnement est adapté de façon à pouvoir exécuter le scénario. Nous avons souhaité enrichir ce processus "classique" (Figure 4.2) [Iksal 2011a][Iksal 2011b]. En effet, comme nous l'avons expliqué dans la section précédente, nous considérons l'analyse de la situation d'apprentissage au travers des besoins d'observation exprimés par un prescripteur. La scénarisation de la situation d'apprentissage doit inclure désormais la scénarisation de l'observation. Cela signifie que le prescripteur a pour mission de spécifier ses besoins d'observation. Un besoin d'observation répond aux questions suivantes :

- **Quoi ?** Il est question ici d'indiquer ce que le prescripteur souhaite observer. En plus d'une dénomination sémantiquement claire, il est important de proposer une explication qui sera utilisée par l'analyste pour formaliser l'observation sous la forme d'un indicateur.
- **Quand ?** Il s'agit du moment auquel le prescripteur souhaite que l'observation soit calculée, pendant la session, après la session, à la demande d'un acteur, suite à l'action d'un acteur, à un instant précis du déroulement de la situation.
- **Pourquoi ?** Le prescripteur doit indiquer son objectif d'observation afin de faciliter la capitalisation de son indicateur et sa réutilisation. En effet, souhaite-t-il évaluer un apprenant, une ressource ? Prévoit-il la ré-ingénierie, l'amélioration de sa situation, ou encore la validation de son approche pédagogique ? Envisage-t-il de fournir des informations aux acteurs pendant la session, inclut-il de la réflexivité dans sa situation d'apprentissage ? Ou encore est-ce pour adapter en session la situation d'apprentissage ?
- **Sous quelle forme ?** Afin de faciliter l'interprétation du résultat de l'observation, il est nécessaire de savoir comment le prescripteur souhaite le voir retranscrire. Il peut par exemple demander à ce que ce résultat soit retranscrit automatiquement dans le langage de scénarisation pédagogique qu'il a utilisé pour prescrire la situation d'apprentissage. Comme souvent il n'y a pas de langage formel, les enseignants/concepteurs utilisant leur propre modèle de représentation, il est nécessaire que cette description de la forme du résultat soit la plus ouverte possible. Afin d'être au plus proche de la représentation de l'indicateur que souhaite le prescripteur, nous devons prendre en compte à la fois le moyen de perception (tableau de valeurs, graphique, etc.), ainsi que la forme de l'information (une valeur unique, un ensemble de caractéristiques, etc.).

Zendagui, dans ses travaux de thèse [Zendagui 2010], s'est intéressé à l'exploitation de l'univers métier pour exprimer les besoins d'observation au travers d'un outil graphique généré grâce à l'univers métier. Il a proposé ensuite une démarche permettant de convertir ce besoin d'observation dans un langage plus formel.

L'étape suivante dans la chaîne éditoriale consiste en la spécification des indicateurs et des données nécessaires à leur calcul. Pour cela, les trois rôles présentés précédemment sont amenés à collaborer. En effet, la formalisation des données est une activité qui relève essentiellement de l'analyste, mais ce dernier peut être amené à consulter le prescripteur pour clarifier le besoin d'observation, et à collaborer avec le développeur afin de savoir si

une information est disponible ou s'il faut envisager de faire évoluer l'EIAH. Dans le cas de l'impossibilité d'obtenir une information au travers de l'EIAH (information absente et impossibilité technique), l'analyste sera amené à négocier avec le prescripteur soit pour mettre de côté le besoin d'observation (solution peu satisfaisante), soit pour trouver une solution externe (éventuellement humaine) permettant d'obtenir l'information attendue. Nous pensons ici par exemple à un observateur doté d'un questionnaire à remplir durant la session, ou encore à un enregistrement audio ou vidéo qu'il faudra exploiter ensuite manuellement. L'utilisation de ce type d'artefacts implique quasi-systématiquement un traitement post-session et l'impossibilité de fournir ces indicateurs au cours de la session. Dans la figure 4.2, les rectangles décrivent les activités de chacun des rôles, et les flèches représentent le flux d'information, par exemple, les données issues de l'analyse de la situation d'apprentissage peuvent servir lors de la ré-ingénierie du scénario ou lors de l'adaptation du déroulement de la situation d'apprentissage. Cette chaîne éditoriale fait abstraction des destinataires de l'observation. En effet, nous avons souhaité représenter le processus du point de vue de la prescription de l'observation et le destinataire n'est pas acteur de cette prescription, sauf s'il occupe le rôle de prescripteur.

4.4 Un enjeu important : La capitalisation

La notion de communauté s'est considérablement développée ces dernières années, et de plus en plus de personnes souhaitent fédérer leur savoir-faire [DPULS 2005a], leur expérience. Dans le domaine de l'analyse de traces, et par conséquent, dans le domaine de l'analyse de traces d'apprentissage, il est important de proposer à la communauté des indicateurs performants et pertinents selon les situations. Le projet DPULS a notamment proposé des patrons sur l'analyse d'usage en EIAH. Dans ce projet, les patrons sont des descriptions. Ils n'ont pas été conçus pour une utilisation automatisée, mais ils ont permis de mettre en évidence l'intérêt que portent les différents acteurs à ces indicateurs. La base de connaissance du projet n'a pas évolué, essentiellement en raison de ce manque d'automatisation. En effet, les utilisateurs du navigateur de patrons pouvaient trouver un indicateur intéressant et lire sa description (par exemple : évaluation de l'organisation collective d'une activité synchrone), mais ils ne disposaient pas des outils permettant de le valuer. Ce que nous souhaitons mettre en place avec notre approche formalisée, c'est la mise à disposition d'indicateurs directement utilisables et par conséquent l'un de nos objectifs est de faciliter la capitalisation et la réutilisation de ces indicateurs. La capitalisation a de nombreux intérêts que nous présentons ici. Elle permet :

- de réutiliser un indicateur dans une session différente avec le même EIAH. Ici, l'indicateur est réutilisé tel quel.
- de réutiliser le moyen d'observation décrit dans l'indicateur lors de la création d'un indicateur similaire (par exemple réutiliser le calcul de fréquence d'un évènement dans deux EIAH différents). Dans cette proposition, la capitalisation facilite la création de nouveaux indicateurs. Le projet SBT s'est aussi intéressé à cet usage de la capitalisation [Djouad et al. 2009].

- de créer des “patrons d’observation”. Un patron d’observation regroupe l’indicateur ainsi que une ou plusieurs solutions à mettre en œuvre selon les valeurs de l’indicateur. Il est alors possible de proposer des solutions selon les types de réactions (régulation, adaptation, etc.) et surtout de capitaliser les évaluations et les discussions sur chacune des solutions, de façon à partager l’expérience avec la communauté. Nous souhaitons aussi associer au patron d’observation les différentes possibilités de visualisation de l’indicateur et comme pour les solutions, nous considérons que la communauté peut réagir, évaluer et discuter autour des choix de visualisation.
- d’aider dans la création des scénarios pédagogiques ou d’observation. En effet, à partir du moment où le concepteur d’un scénario pédagogique dispose d’un accès aux indicateurs ou plus largement aux patrons d’observation, il a la possibilité de bénéficier de l’expérience d’autres concepteurs et de corriger son scénario pédagogique avant la première utilisation. Du point de vue de la connaissance métier, il peut alors consulter les besoins d’analyse et les solutions correctives apportées. Il évite du coup les écueils rencontrés par les autres. Du point de vue de l’observation, si les moyens d’observation sont opérationnalisés, il peut alors réutiliser des indicateurs dans son propre scénario d’observation soit directement si l’EIAH est identique, soit en les adaptant, ce qui lui simplifie la construction de son scénario d’observation. L’adaptation pourrait porter, par exemple, sur l’utilisation d’un acteur humain en cas d’absence de la trace dans l’environnement.
- d’élucider de nouveaux indicateurs, en testant ou combinant des indicateurs existants. L’élucitation consiste à faire émerger de nouveaux indicateurs à partir de la base existante. Il peut être question de réutiliser des indicateurs non prévus dans le scénario d’observation et de rechercher si des informations pertinentes sur la session d’apprentissage émergent. On se rapproche dans ce cas de la fouille de données d’une certaine façon car l’utilisateur ne sait pas a priori ce qu’il cherche.
- de créer une communauté, de fédérer les usagers autour de l’observation. Cela permet donc de renforcer des liens et d’aider l’enseignant et/ou le tuteur dans leurs activités. Nous visons une capitalisation à destination des usagers en priorité et non à destination des chercheurs comme dans le cas de DataShop [Koedinger et al. 2010] ou MULCE [Reffay et al. 2008]. Ce qui implique : (1) la prise en compte en cours de session des besoins des usagers. En effet, ces entrepôts ont été conçus pour une utilisation après la session d’apprentissage, les données sont filtrées puis intégrées dans ces outils, ensuite, des algorithmes sont disponibles pour fouiller les données de la session. (2) que les mécanismes de perception soient interprétables par l’usager, en effet, les enseignants ou les apprenants n’ont pas les mêmes facilités d’interprétation que les chercheurs.

Notre proposition repose alors sur la définition d’une représentation formelle (un langage) extensible, ouverte et interrogeable. Notre approche se différencie des autres notamment par la prise en compte de l’usager dans la spécification et aussi par les possibilités offertes d’intégration dans des EIAH, comme nous l’avons mis en œuvre dans l’environnement Hop3X du LIUM. Nous présentons notre solution dans le chapitre suivant.

4.5 Les objectifs d'observation

Nous avons identifié différents objectifs d'observation liés aux indicateurs, ces objectifs concernent différents acteurs de la situation pédagogique. Dans le cadre du projet DPULS [DPULS 2005a], trois objectifs ont été proposés :

- *Evaluation de l'apprenant* (Learner assessment) : les indicateurs servent à évaluer l'apprenant pendant ou après la session d'apprentissage.
- *Régulation de l'apprentissage* (Learning regulation) : les indicateurs sont utilisés en aide à la décision lors de la régulation d'une session d'apprentissage.
- *Validation du système* (System validation) : l'indicateur permet de valider/invalidier les fonctionnalités et/ou l'usage du système.

En nous basant sur cette première liste d'objectifs, nous l'avons étendue de la façon suivante :

- *La régulation* : comme pour DPULS, l'indicateur est utilisé en session pour réguler la situation d'apprentissage.
- *L'évaluation* : il s'agit ici d'utiliser les indicateurs pour évaluer les acteurs en général de la situation d'apprentissage, mais aussi les ressources et les composants du système.
- *L'adaptation* : les indicateurs peuvent être utilisés en session, de façon automatique, pour adapter la situation d'apprentissage au lecteur. Pour cela, le prescripteur devra prévoir des règles d'adaptation associées aux différentes valeurs d'un indicateur.
- *La réflexivité* : souvent lorsque les indicateurs sont proposés, ils sont présentés au concepteur d'une activité ou d'un outil. De plus en plus, il est intéressant de diffuser les indicateurs auprès des usagers directement afin qu'ils prennent conscience de leur activité, de leur apprentissage. C'est le cas dans le système TraViS [May 2009]. Nous considérons ici une présentation de l'analyse de l'activité de l'utilisateur mais aussi une présentation de l'analyse de celle des autres afin que l'utilisateur puisse se situer et mener une réflexion sur son activité propre. La réflexivité peut porter sur un individu ou un groupe, et concerner différents rôles : le tuteur, l'apprenant ou l'enseignant.
- *La réingénierie* : souvent un indicateur est utilisé afin d'améliorer l'activité proposée ou l'outil développé. Il est question d'évaluer une ressource, un scénario, voire de valider des hypothèses liées au scénario pédagogique. C'est pour cela que la réingénierie est un objectif d'observation.

4.6 Scénarios d'utilisation

Dans cette section, nous allons donner deux exemples basés sur notre approche d'ingénierie de l'observation par la prescription. Un premier exemple en réingénierie qui a servi de base à nos travaux, puis un exemple en régulation qui sera réifié dans le chapitre 6 car

il a servi à valider notre proposition.

4.6.1 Un exemple en réingénierie

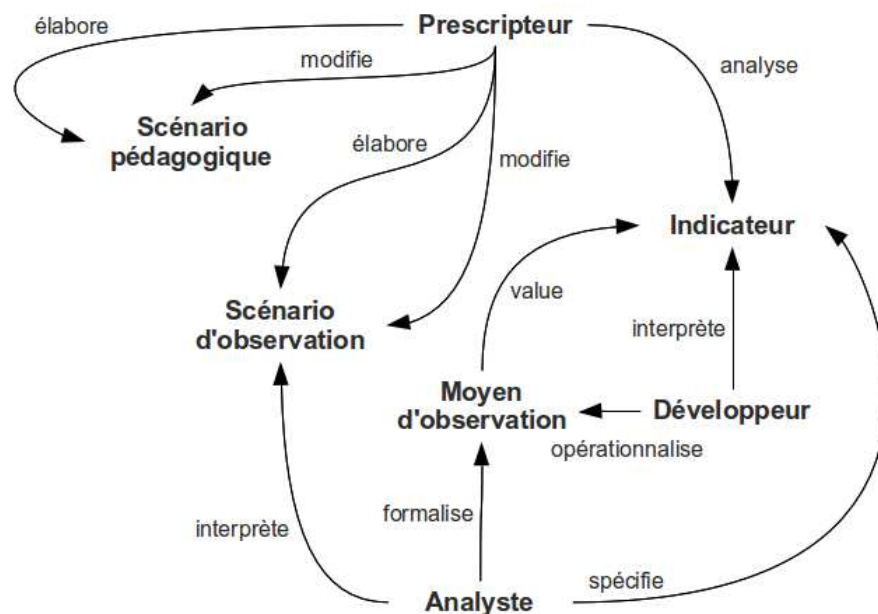


FIGURE 4.3 – Modèle de tâches de la réingénierie

L'objectif de la réingénierie d'un scénario pédagogique est de le faire évoluer suite aux constats réalisés pendant ou après une session d'apprentissage. Le prescripteur (enseignant ou concepteur pédagogique) élabore son scénario et dans le même temps spécifie ses besoins d'observation (par exemple dans son langage métier). La spécification des besoins d'observation passe par une explication détaillée du besoin, la liste des concepts du scénario pédagogique concernés par cette observation, la spécification du résultat attendu (ensemble de valeurs, format de représentation, etc.) ainsi qu'une justification du besoin. Cette justification est utile pour la capitalisation de l'indicateur qui sera créé. L'analyste interprète ces besoins pour formaliser des indicateurs. Il identifie les observés dont il a besoin puis formalise les indicateurs et toutes les données nécessaires au calcul des indicateurs. Le développeur intervient si nécessaire afin de compléter les liaisons entre l'analyseur de traces et l'environnement.

En cours de session ou après la session d'apprentissage, les indicateurs sont valués et transmis au prescripteur, ce dernier étudie les informations qu'il reçoit et en tire des conclusions sur son scénario et sa situation d'apprentissage.

Dans ce scénario type, le prescripteur a exprimé ses besoins avant la session d'apprentissage. Il est possible qu'il définisse des besoins après la session, l'inconvénient de cette situation est l'impossibilité d'intervenir sur l'environnement afin d'intégrer de nouvelles sondes.

Nous avons collaboré avec le laboratoire SYSCOM de l'Université de Savoie afin d'étudier la faisabilité de notre approche sur un environnement indépendant de nos activités

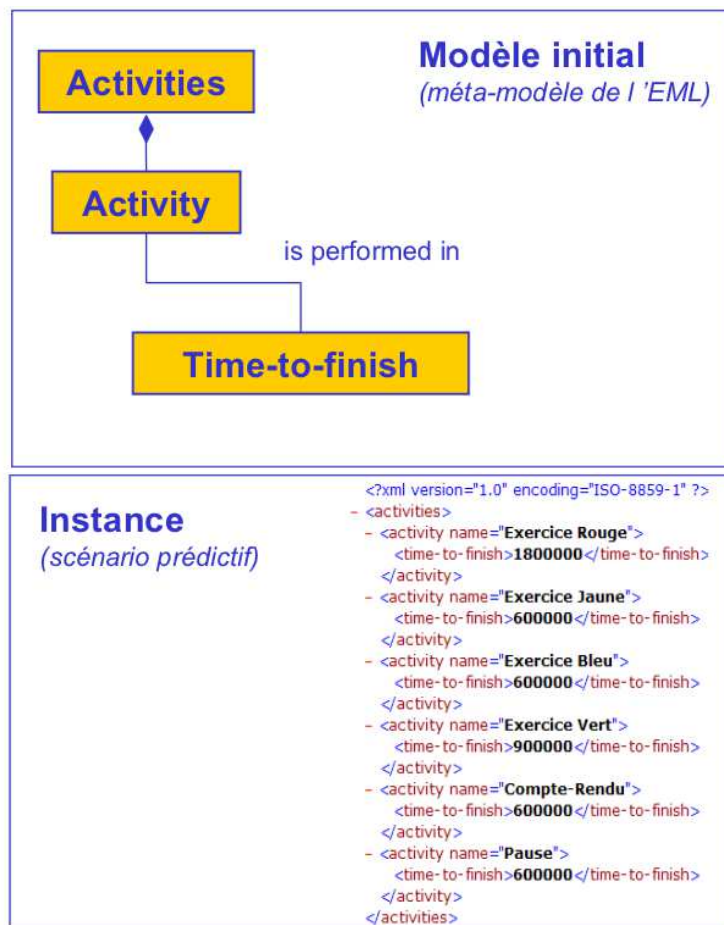


FIGURE 4.4 – Modèle de scénario et son instance

[Iksal et al. 2007]. Nous avons utilisé les traces d'un exerciceur de modélisation UML utilisé dans le cadre de la formation de DUT SRC de l'IUT de Chambéry. Les étudiants avaient cinq activités à réaliser (4 diagrammes UML et un compte-rendu), ils étaient libres (ordonnancement des activités, pauses, etc.) mais leur session était observée au travers d'un enregistreur de frappes et d'un logiciel leur permettant d'indiquer l'activité qu'ils étaient en train de réaliser. Nous disposons de la liste des activités ainsi que d'une durée estimée de réalisation, cf. Figure 4.4 (les activités portaient des noms de couleur afin de ne pas influencer les étudiants dans l'ordre de réalisation des activités, il s'agissait d'une volonté de l'enseignant).

Plusieurs indicateurs ont été modélisés et évalués :

- le dépassement du temps de réalisation d'une activité en moyenne : il s'agit ici de mettre en évidence qu'en général, les étudiants ont dépassé le temps estimé de réalisation d'une activité particulière.
- le dépassement du temps de réalisation d'une activité par étudiant : il est intéressant ici de mettre en évidence le ou les étudiants qui peuvent éprouver des difficultés à réaliser l'activité dans les temps.
- le parcours d'un étudiant : cet indicateur permet d'étudier l'évolution d'un étudiant

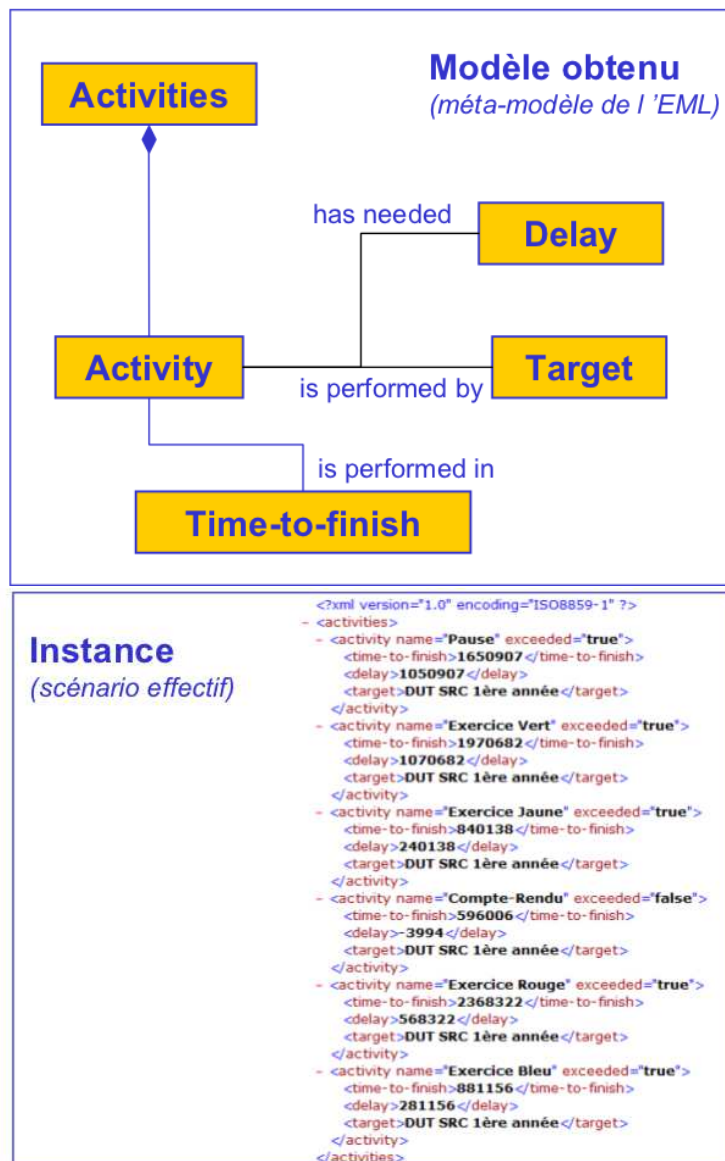


FIGURE 4.5 – Réingénierie du modèle de scénario et de son instance

- dans la liste des activités préconisées pour ce cours.
- l'état de la classe à un instant t : il s'agit ici d'obtenir une "photo" de la classe à différents instants de la session d'apprentissage.
 - la détection des périodes d'inactivité et de leurs durées : il s'agit de mettre en évidence un problème dans la réalisation du scénario par la détection des périodes d'inactivité et surtout de leurs durées. Est-ce que l'étudiant éprouve des difficultés à réaliser l'activité ? Est-ce une durée "normale" à ce stade de l'exercice ? etc.. L'indicateur va classer les périodes d'inactivité en fonction de la durée dans des catégories définies au préalable par l'enseignant.
 - le scénario le plus probable : on cherche à ce niveau à mettre en évidence ce qui correspondrait au scénario réalisé par la majorité des apprenants. En fait, il s'agit

de repérer les différentes étapes du parcours de chaque étudiant et de comptabiliser le nombre de passages sur chacune d'elles. En listant les étapes les plus parcourues, il est possible de déduire le scénario le plus probable.

Afin de représenter les résultats des indicateurs, nous avons été amenés à faire évoluer les modèles, en intégrant deux nouvelles caractéristiques (cf. Figure 4.5), le retard de l'étudiant ainsi que sa promotion.

4.6.2 Un exemple en régulation

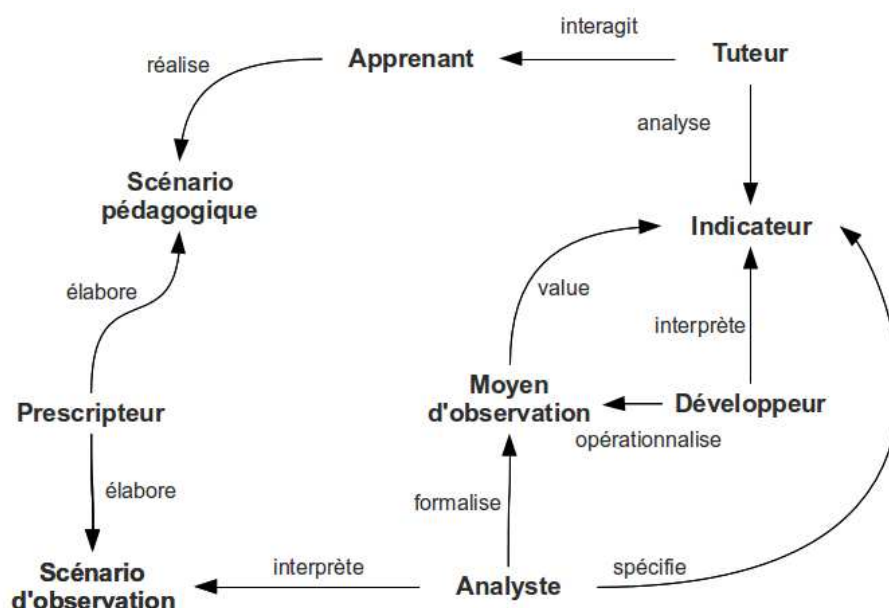


FIGURE 4.6 – Modèle de tâches de la régulation

L'objectif de la régulation d'une situation d'apprentissage est de suivre l'activité des apprenants et d'intervenir de façon à corriger les défauts et à prévenir ou résoudre les points de blocage des apprenants. Le prescripteur (enseignant ou concepteur pédagogique) élabore son scénario et dans le même temps spécifie ses besoins d'observation en fonction de l'activité du tuteur. Cela signifie (éventuellement en accord avec le tuteur) que le prescripteur puisse identifier dans son scénario les éléments à suivre plus particulièrement. La spécification est identique à celle décrite dans l'exemple précédent. En ce qui concerne l'analyste et le développeur, ils opèrent de la même façon en formalisant les indicateurs avec éventuellement une modification de l'environnement.

En cours de session, les indicateurs sont valués et transmis au tuteur, ce dernier étudie les informations qu'il reçoit et réalise des actions de régulation de la situation d'apprentissage. Lorsque cela est possible, le tuteur peut solliciter l'intégration de nouveaux indicateurs (à condition que la réutilisation soit possible, c'est-à-dire que les indicateurs soient prêts à l'emploi et que les sondes soient disponibles dans l'environnement).

Une mise à l'essai [Peña de Carrillo et al. 2011] a été réalisée au sein de l'Université du Maine, elle est présentée dans le chapitre 6 sur l'instrumentation de notre proposition

de formalisation du processus d'ingénierie de l'observation : l'environnement UTL (Usage Tracking Language), présenté au chapitre suivant.

4.7 Synthèse

L'analyse de traces est souvent portée par des systèmes développés ad hoc pour des EIAH particuliers. Depuis quelques années, de plus en plus de travaux se sont focalisés sur des systèmes capables de s'abstraire des formats de représentation des traces spécifiques aux EIAH, c'est le cas par exemple d'IDLS [Bousbia 2011], de SBT [Djouad et al. 2010] ou de TrAViS [May et al. 2008]. Nous avons proposé dans ce chapitre une approche indépendante des formats de représentation, mais surtout une ingénierie de l'observation. La spécificité de nos travaux concerne la prise en compte de l'utilisateur dans ce processus d'ingénierie dès la phase de réflexion sur l'observation. En effet, nous avons pris le parti de considérer que l'utilisateur est capable d'exprimer son besoin d'observation et donc de participer à la chaîne éditoriale de l'observation. Cette expression du besoin d'observation est une prescription de ce que l'utilisateur souhaite obtenir. Nous accordons une importance aussi au rôle de l'analyste qui est nécessaire à l'élaboration d'indicateurs complexes. Toutefois, un enseignant peut être amené à prendre le rôle de prescripteur et de l'analyste s'il en a les capacités.

Dans ce chapitre, nous avons aussi proposé deux scénarios mettant en œuvre notre approche avec deux objectifs différents. Nous retrouvons dans ces deux scénarios un noyau commun concernant la spécification des données, par contre ils diffèrent au niveau de l'utilisation de l'observation. En ce qui concerne les acteurs, le prescripteur est systématiquement l'enseignant/concepteur. C'est lui qui élabore le scénario pédagogique. Nous avons ajouté dans le scénario de régulation, l'apprenant et le tuteur qui sont directement concernés par l'utilisation des indicateurs.

Cette chaîne éditoriale doit être vue comme un cadre méthodologique qui organise les travaux de recherche que nous menons et que nous encadrons. Dans le chapitre suivant, nous présentons le langage de représentation de données d'observation que nous proposons pour spécifier, calculer et capitaliser les indicateurs et les données nécessaires à leur création.

Chapitre 5

Méthodologie de spécification des données d'observation

Sommaire

5.1	Introduction	81
5.2	Représenter les données d'observation : Le modèle DGU . . .	81
5.3	Spécification déclarative des données d'observation	83
5.3.1	Le modèle conceptuel	84
5.3.2	Les modèles d'information	86
5.3.3	De la spécification à l'opérationnalisation : DCL4UTL	93
5.4	Synthèse	96

5.1 Introduction

La mise en œuvre de la chaîne éditoriale de l'observation présentée au chapitre précédent nécessite une méthodologie de spécification des données d'observation. Cette chaîne éditoriale décrit le processus de gestion de l'observation depuis l'expression des besoins par le prescripteur jusqu'à l'exploitation des résultats de l'observation par différents acteurs selon l'objectif préconisé. Nous avons choisi une approche prescriptive pour spécifier l'expression du besoin d'observation sous la forme d'indicateurs capitalisables et réutilisables. Dans ce chapitre nous présentons les actions de recherche menées à la suite de la proposition du modèle sémantique de la trace présentée au chapitre 3. Nous avons souhaité répondre à la question de recherche : *“Comment formaliser le besoin d'observation et le calcul de l'indicateur correspondant afin de le rendre automatisable, réutilisable et adaptable ?”*. Pour cela, nous proposons une méthodologie de description de données basée sur les différents rôles de notre processus d'ingénierie par la prescription. Nous avons conçu un langage de représentation de données associé à cette méthodologie. Ce langage est intéressant à différents niveaux :

- il est conçu autour du besoin et de l'objectif d'observation du prescripteur ;
- il prend en compte la connaissance métier du prescripteur ;
- il est indépendant d'un langage de modélisation pédagogique, d'un format de représentation de traces lié à une plate-forme, d'un type d'entrepôt de données et du langage de programmation associé ;
- il est ouvert ;
- il est opérationnalisé ;
- il est capitalisable et adaptable ;
- il peut être utilisé en cours de session d'apprentissage.

Le savoir-faire obtenu dans le calcul de l'indicateur est alors générique car nous basons notre description sur la sémantique de l'information extraite de la trace plutôt que sur des transformations successives de la trace générée par le système. Ce langage constitue la fondation de nos travaux futurs. Cette deuxième étape dans l'expression des données montre la faisabilité d'une telle approche, l'étape suivante sera de décrire le moyen d'observation (le calcul des indicateurs) sous la forme d'un langage d'expression sémantique utilisable directement par l'utilisateur en nous basant sur le vocabulaire métier décrivant chaque information. = Dans un premier temps, nous présentons la méthodologie de représentation de données appelée DGU (Defining-Getting-Using), puis nous présentons le langage de spécification de données basé sur cette méthodologie.

5.2 Représenter les données d'observation : Le modèle DGU

Chacun des trois rôles proposé dans la chaîne éditoriale possède son propre point de vue sur la donnée d'observation. Le prescripteur est à l'origine de cette donnée, il la décrit et la nomme, puis établit l'utilisation envisagée (l'objectif d'observation). C'est ce qui fait toute l'importance de cette donnée. L'analyste assure le soutien sur la spécification du

moyen d'observation avec l'aide du développeur. A ce niveau, il est question de spécifier ce qui est nécessaire à la création d'une donnée et la façon dont elle doit être évaluée. Nous n'utilisons pas le terme "calculer" dans ce cas, car nous considérons qu'un agent humain peut être amené à valuer une donnée au travers de son expertise.

Selon le processus que nous proposons dans la figure 4.1, trois étapes sont importantes pour la modélisation de l'observation. L'expression du besoin permet d'explicitier ce que le prescripteur souhaite observer. La conception du besoin est une étape de modélisation "technique" du moyen de collecte ou de calcul de l'observation. Enfin, l'exploitation du résultat correspond à l'utilisation qui est faite de l'observation. Notre approche étant basée sur la prescription de l'observation, notre modélisation est "descendante", c'est-à-dire que nous partons de l'énoncé du besoin pour arriver à la spécification du moyen d'observation. Nous proposons alors de modéliser une donnée d'observation selon trois facettes (cf. figure 5.1) :

- la facette définition (Defining) permet de modéliser le besoin d'observation, elle apporte la sémantique de l'observation, c'est-à-dire la réponse à la question "quoi ?" ;
- la facette obtention (Getting) permet de modéliser le moyen d'observation, elle correspond aux aspects techniques de l'observation, elle répond à la question "comment ?" ;
- la facette utilisation (Using) précise l'utilité de la donnée d'observation et décrit l'utilisation qui en est faite. Cette facette concerne l'objectif de l'observation et la présentation du résultat. Dans le cas d'un EIAH scénarisé avec un langage de modélisation pédagogique, il est concevable d'indiquer ici comment intégrer le résultat de l'observation dans le même langage de façon à faciliter la comparaison entre le scénario d'apprentissage prescrit et ce qui pourrait être perçu comme un scénario d'apprentissage effectif.

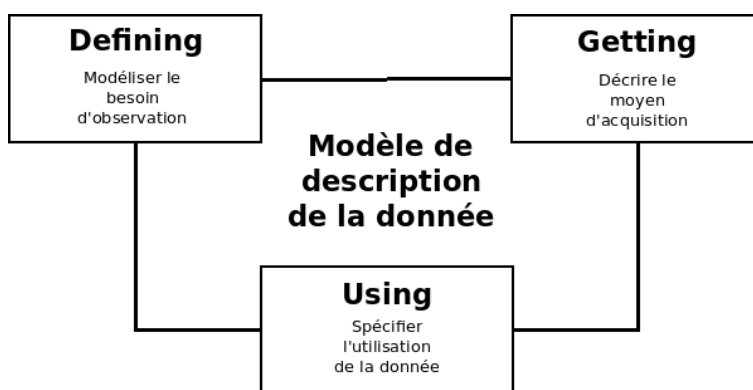


FIGURE 5.1 – Le modèle DGU

En ce qui concerne les rôles, le prescripteur intervient sur les facettes "définition" et "utilisation". Tout d'abord, en tant que demandeur d'une observation, il spécifie son besoin d'observation. Ensuite, il indique dans la facette "utilisation" le lien avec les concepts du

scénario d'apprentissage ainsi que l'objectif d'observation. Le but de ces deux facettes est d'embarquer toute la sémantique de la donnée ainsi que la justification de son intérêt grâce aux connaissances métier du prescripteur.

L'analyste prend en charge la description du moyen d'observation dans la facette "obtention". Il s'agit dans cette description d'indiquer comment est obtenue l'observation ainsi que son ou ses auteurs. Le développeur intervient sur les deux facettes "obtention" et "utilisation". Il est présent, par exemple, en soutien pour l'analyste sur l'expression du moyen et les possibilités offertes par la plate-forme. Il élabore le modèle de représentation du résultat en fonction des attentes du prescripteur.

Ces points de vue peuvent s'appliquer à toutes les données du processus d'observation. La trace observée peut être représentée au travers de sa sémantique, de sa description technique (la façon dont l'information est organisée dans la trace) et de son modèle de représentation pour un traitement automatisé dans la chaîne éditoriale de l'observation.

Nous proposons une instanciation de ce modèle de représentation au travers du langage UTL – Usage Tracking Language – qui permet de spécifier de façon déclarative la totalité d'une donnée d'observation automatisable.

5.3 Spécification déclarative des données d'observation

Avant d'instrumenter l'analyse de l'observation, il est nécessaire de modéliser les données utilisées dans ce processus d'analyse. Il est question dans cette section d'une action de recherche avec pour objectif d'étudier la modélisation des données d'observation en fonction de l'expression des besoins du prescripteur (la prescription). Ce qui nous amène à une approche complémentaire à celle de SBT [Djouad et al. 2009] dans laquelle la modélisation des données s'appuie sur une représentation des traces premières transformées jusqu'à obtention d'une trace suffisamment abstraite pour servir d'indicateur (approche ascendante, depuis la trace vers l'indicateur). Afin de modéliser les données sous une forme déclarative, donc textuelle et évolutive, nous avons proposé un langage. Nous considérons que l'utilisation d'une forme déclarative textuelle est un compromis intéressant pour la capitalisation des données, et son évolutivité permet de prendre en compte la nécessité d'adaptation des données selon le contexte de la situation d'apprentissage. Le langage UTL (Usage Tracking Language) permet de modéliser un observable et, plus largement, toute donnée impliquée dans la modélisation de l'observation [Choquet et Iksal 2007a]. Il instrumente la description des méthodes d'analyse de l'observation en vue de leur capitalisation grâce à un ensemble d'éléments spécifiques pour le moyen d'observation et regroupés dans une section du langage nommée UTL/P (UTL/Patron).

Le langage que nous proposons permet de :

- capitaliser le savoir-faire en analyse et en observation d'une session. Notre objectif est de capitaliser les indicateurs et le moyen d'observation permettant de les valuer. Actuellement, nous n'envisageons pas, comme c'est le cas dans MULCE, de capitaliser l'organisation autour de l'observation comme les protocoles d'expérimentation, le point de vue organisationnel de la formation avec un EIAH etc..
- définir le besoin d'observation, et avec l'assistance du développeur, de spécifier les moyens techniques à mettre en œuvre pour l'acquisition des données à observer,

- structurer des données générées par le dispositif apprentissage depuis la donnée brute jusqu'à l'indicateur dans une forme indépendante des formats des traces générées.

5.3.1 Le modèle conceptuel

Ce langage a été conçu comme une extension de la proposition faite dans le chapitre 3. Le modèle conceptuel que nous proposons intègre une spécification de la donnée d'observation en plus des descriptions des traces et des concepts traçables définis au préalable. Afin de proposer une description capitalisable du savoir-faire, nousinstancions le modèle DGU (Defining-Getting-Using) sur les différentes données d'observation que nous avons identifiées. Chaque donnée, au sens du langage UTL, est caractérisée selon les trois facettes du modèle DGU. Le modèle conceptuel d'UTL est structuré en trois parties :

- la partie “patron” (UTL/P) permet de décrire la structure d'un observable. Cette partie est principalement dédiée à la capitalisation des savoir-faire embarqués dans les techniques d'analyse de l'observation de l'utilisation d'un EIAH.
- la partie “scénario” (UTL/S) permet de lier la description des indicateurs au scénario pédagogique particulier du dispositif d'apprentissage. Ce lien sémantique se fait par instanciation de cette partie sur le méta-modèle d'expression pédagogique adopté par le concepteur et sur le scénario de la situation pédagogique. Cette méthode de transformation est décrite dans le chapitre 3.
- la partie “trace” (UTL/T) permet de lier la description d'une donnée à collecter avec la donnée effective qui devra être observée dans le dispositif d'apprentissage. Ce lien fonctionnel se fait par instanciation de cette partie sur les formats particuliers d'enregistrement des traces du dispositif. Cette méthode de transformation est décrite dans le chapitre 3.

Nous avons organisé le modèle conceptuel selon ces trois parties car cela permet de différencier les différents liens et usages. UTL/P correspond à la description de la donnée qui est stable et complètement réutilisable. En ce qui concerne les deux autres parties, elles peuvent être instanciées plusieurs fois. Il s'agit d'une similarité avec le concept informatique de plugin. En effet, si la donnée est utilisée avec un autre méta-modèle d'expression pédagogique, il est possible de décrire les concepts traçables de ce nouveau méta-modèle et de les relier avec les données décrites avec UTL/P. De la même façon, si l'usager change de plate-forme, il suffit de donner la nouvelle description des traces (UTL/T) et de conserver la sémantique embarquée utilisée par UTL/P.

En ce qui concerne les données d'observation, nous avons décidé de les catégoriser. Etant donnée la nature assez variées de ces données, nous n'avons pas trouvé pertinent de nommer chacun de ces objets de la même façon. Contrairement à SBT où l'analyse part de la trace qui est transformée et qui reste une trace, nous avons souhaité suivre les travaux menés dans différents projets européens ([DPULS 2005a], [ICALTS 2004], etc.). Nous avons retenu deux types principaux de données impliquées dans l'analyse de l'observation : la *donnée dérivée* et la *donnée primaire* (cf. Figure 5.2).

- Les *données primaires* ne sont pas calculées ou établies avec l'aide d'autres données.

Elles sont de trois types : brute, de production ou additionnelle.

- Les *données dérivées* sont calculées ou établies à l'aide des *données primaires* et/ou d'autres *données dérivées*.

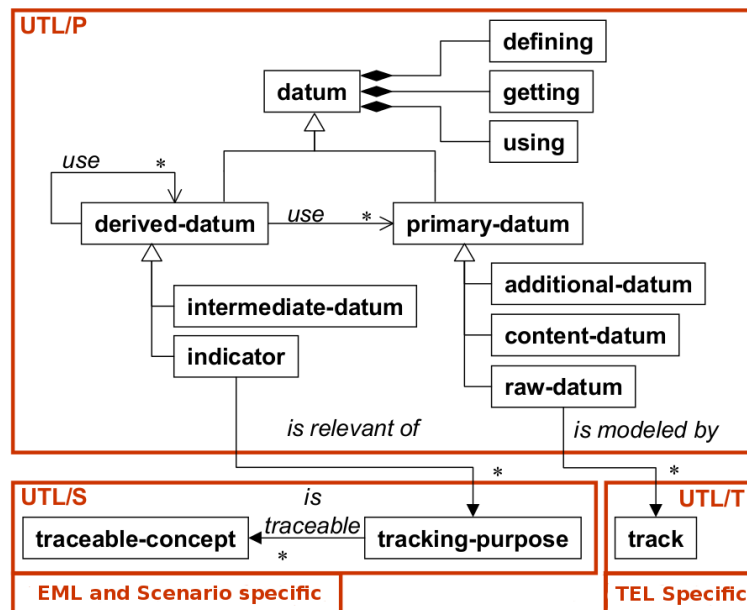


FIGURE 5.2 – Modèle conceptuel du langage UTL

Nous n'avons pas conservé la typologie définie dans DPULS, car si l'on prend en compte la définition de la donnée dérivée, l'indicateur est une donnée dérivée. De plus, nous avons souhaité séparer la donnée brute au sens d'observé (la trace) de ce qui est produit concrètement par l'apprenant (un rapport, un fragment de texte) et constitue un rendu pendant la session. Voici les définitions des différents types de données :

- une *donnée brute* (*raw-datum*) est obtenue par transformation de la trace générée par l'environnement : elle est localisée dans la trace par la modélisation de l'observation, extraite de cette trace après la collecte et représentée dans un format indépendant du dispositif d'apprentissage. Elle peut être collectée avant, pendant ou après la session d'apprentissage par le dispositif d'enseignement, par exemple un fichier de "logs" enregistré par le système, un enregistrement vidéo de l'apprenant pendant la session, les informations recueillies par un questionnaire avant ou après la session, l'ensemble des messages postés dans un forum. Chaque donnée brute est modélisée à l'aide d'un objet *trace* (*track*) selon le même principe que celui employé avec notre premier modèle (cf. chapitre 3). Nous considérons tous les types de données brutes qu'ils soient exploitables directement ou par une interprétation humaine (enregistrement audio ou vidéo).
- une *donnée de production* (*content data*) est produite volontairement par les acteurs de la session d'apprentissage (apprenant, tuteur et/ou enseignant). Ces données sont principalement des travaux réalisés par les étudiants et destinés à être évalués, mais

peuvent également être, par exemple, un rapport établi par le tuteur sur la qualité de l'activité ou d'une ressource pédagogique.

- une *donnée additionnelle* (additional data) est une donnée liée à la situation pédagogique et utilisée pour l'établissement des données dérivées. Ces données sont de natures très variables, par exemple la valeur d'un champ d'une méta-donnée caractérisant une ressource, une taxonomie, une donnée ad hoc, la liste des apprenants, le sujet distribué, etc..
- un *indicateur* (indicator) est un observable signifiant sur le plan pédagogique, calculé ou établi à l'aide d'observés, et témoignant de la qualité de l'interaction, de l'activité et de l'apprentissage dans un EIAH. Ainsi, un indicateur est défini en fonction d'un *objectif d'observation* (tracking purpose), motivé par un objectif pédagogique, et donc lié à un *concept traçable* (traceable concept) du scénario pédagogique prédictif. L'ensemble de ces *concepts traçables* constitue l'instanciation d'UTL sur le métalangage d'expression pédagogique.
- une *donnée intermédiaire* (intermediate datum), au contraire d'un indicateur, n'a pas de signification pédagogique en soi. Elle est cependant nécessaire à la construction d'un ou de plusieurs indicateurs.

Une classification similaire a été adoptée par Bousbia [Bousbia 2011] dans le système IDLS. Il est question effectivement de données comme les données additionnelles, mais elle propose une classification des indicateurs selon leur niveau (haut, intermédiaire et bas), cela sous-entend que toute donnée tracée ou transformée est un indicateur signifiant pour l'utilisateur. Comme nous l'avons déjà expliqué, toutes les données ne répondent pas à un objectif d'observation ou pédagogique, d'où notre classification.

Toutes ces données servent au calcul d'indicateurs, le prescripteur se focalise sur la modélisation d'un indicateur. L'analyste et le développeur collaborent sur les autres types de données, selon ce qui est prévu par le scénario d'apprentissage (données additionnelles et données de production) et par la plate-forme (données brutes). La décision de modéliser des données intermédiaires revient à l'analyste. En effet, il est à même de juger si une donnée qui n'a pas de signification pédagogique peut être réutilisée dans le calcul de différents indicateurs, et dans ce cas, il la modélise en donnée intermédiaire.

5.3.2 Les modèles d'information

Nous avons représenté chaque type de données sous la forme d'un modèle d'information qui est implémenté en XML Schéma. Ce choix technologique est lié aux possibilités d'évolution des schémas qui permettent d'enrichir les modèles d'information. Cela a été le cas, notamment pour l'intégration du langage de combinaison de données DCL4UTL qui permet de décrire le calcul des indicateurs et des données intermédiaires afin d'en capitaliser le savoir-faire. Comme indiqué dans la section 3.3.2, ces modèles d'information utilisent un formalisme analogue à celui d'IMS Learning Design [IMS-LD 2003].

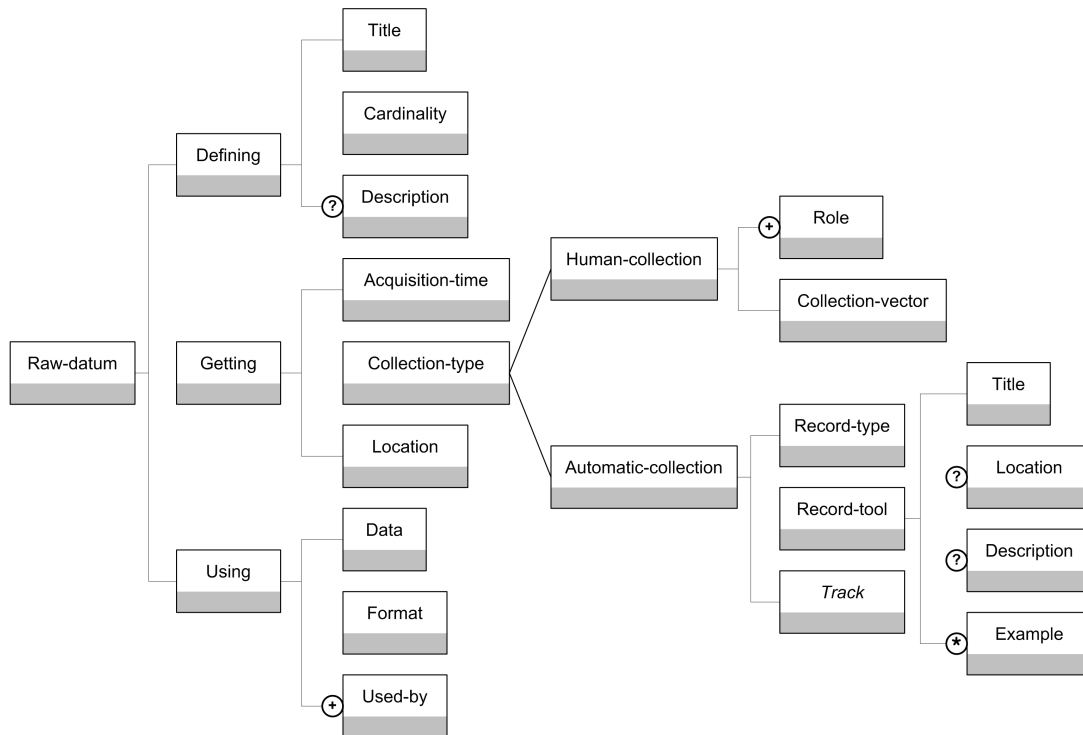


FIGURE 5.3 – Le modèle d'information d'une donnée brute

5.3.2.1 La donnée brute

La donnée brute sert de base au calcul des données d'observation, mais avant tout elle décrit le ou les observés que l'on peut obtenir au sein d'une trace. Pour cela, il est nécessaire de décrire comment l'information peut être obtenue et à quel moment. La figure 5.3 détaille les trois facettes d'une *donnée brute*.

La facette “définition” (Defining) de la donnée brute est décrite par trois éléments : le *titre* (Title) de la donnée doit être concis et sans ambiguïtés pour exprimer la sémantique de la donnée ; la *cardinalité* (Cardinality) de la donnée représente le nombre d'instances possibles de la donnée ; une *description* (Description) plus détaillée peut être ajoutée pour décrire de manière informelle la nature de la donnée. Le titre est un élément important car il permet de retrouver la donnée brute. En effet, il est possible d'avoir plusieurs données brutes pour une même sémantique, chaque donnée correspondant à une trace particulière. Mais lors de l'utilisation d'une base de traces particulière, seule une donnée brute sera instanciée et donc utilisée.

La facette “obtention” (Getting) se concentre sur la description des moyens d'observation, c'est-à-dire de la méthode d'analyse permettant d'établir la donnée. Dans le cadre de la donnée brute, la méthode d'analyse correspond à la méthode d'extraction de l'information dans la trace. Soit cette extraction peut être automatisée (cas des fichiers textuels et structurés par exemple), soit elle est faite manuellement comme pour une vidéo. Elle est composée de trois éléments :

- la *période d'acquisition* (Acquisition-time) prend ses valeurs dans la liste fermée {avant la session (Before-session) ; en cours de session (During-session) ; après la

- session (After-session)}.
- le type de collection (Collection-type) est :
 - un recueil (Human-collection) opéré par au moins un rôle (Role) (par exemple, un observateur de la session) en utilisant un moyen de collecte (Collection-vector) (par exemple une caméra, un papier et un crayon).
 - un recueil automatique (Automatic-collection), caractérisé par :
 - le type d'enregistrement (Record-type) qui prend ses valeurs dans une liste ouverte (par exemple, fichier de log, chat, mail),
 - un outil de collecte (Record-tool) dont la localisation (Location) est connue s'il est déjà développé et utilisable dans l'environnement de formation. Sinon, il est nécessaire d'en donner une description (Description) et/ou quelques exemples (Example) afin d'aider au développement de l'outil.
 - la description de la trace (Track). L'élément "Track" est en italique sur la figure 5.3 pour rappeler qu'il est conforme à la description proposée au chapitre 3.
 - la *localisation* (Location) de la donnée, il s'agit souvent de l'URL du fichier qui la contient ou l'adresse de la base de données.

La facette "utilisation" (Using) comporte trois éléments : *utilisé-par* (Used-by) est proposé afin de faciliter la navigation dans un graphe de dépendances des données ; *contenu* (Data) permet de stocker la donnée après son extraction de la trace source ; *format* décrit comment la donnée est représentée, une fois obtenue (son expression en XML par exemple). Le graphe de dépendances des données a pour objectif de proposer un outil graphique de visualisation des données permettant facilement à l'analyste ou au prescripteur de se représenter les liens entre chaque donnée.

Afin de couvrir une variété de format de trace assez conséquente, nous avons prévu l'intégration des recueils de données réalisées par des agents humains. L'exploitation de ce type de données nécessite un traitement souvent lourd et ne permet donc pas la gestion de ces données en cours de session d'apprentissage.

5.3.2.2 La donnée de production

Les *données de production* sont obtenues lors de la session de formation (rapports, exercices, etc.). Elles sont par conséquent toujours clairement identifiées même celles non prévues dans le scénario prédictif. Ce type de donnée nécessite l'aide d'un agent humain pour soit l'évaluer, soit en extraire l'information pertinente. Si la production est générée par l'EIAH, elle peut être traduite sous la forme d'une donnée brute décrite dans la section précédente. La figure 5.4 détaille les trois facettes d'une *donnée de production*.

Pour les mêmes raisons que la *donnée brute*, la facette "définition" de la *donnée de production* est composée d'un *titre* et d'une éventuelle *description*. Ces informations portent la sémantique de la donnée.

La facette "obtention" correspond au moyen de collecte de la donnée. Nous disposons de deux types d'information, tout d'abord l'auteur de la donnée et l'endroit où elle est mise à disposition, ensuite l'identification de la personne capable de rendre cette donnée exploitable dans le calcul d'indicateurs (par exemple en l'évaluant). Cette facette est caractérisée par :

- la *localisation* de la production,

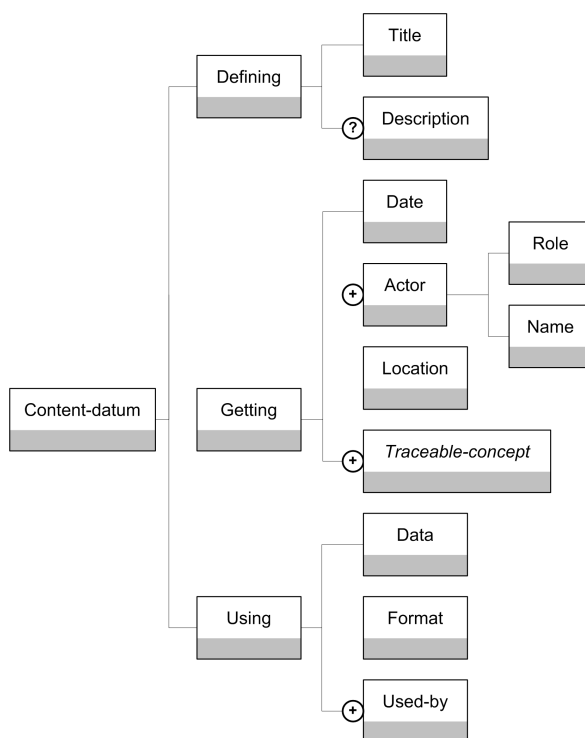


FIGURE 5.4 – Le modèle d'information d'une donnée de production

- la *date* de la production,
- les *acteurs* (Actor) concernés par cette donnée. Nous avons tout d'abord l'auteur de la donnée, mais aussi le ou les évaluateurs qui ont pour mission d'extraire de la donnée des informations exploitables dans la création d'indicateurs (par exemple en donnant une note ou une mention sur un rapport).
- au moins un *concept traçable* (Traceable-concept) du scénario pédagogique. Ces concepts servent à associer la production au scénario pédagogique pour renforcer la description sémantique avec les connaissances métier du prescripteur.

La facette “utilisation” est composée du *contenu* de la donnée, de son *format*, ainsi que de la liste des données qui l'utilisent, comme dans le cas de la *donnée brute*.

5.3.2.3 La donnée additionnelle

Les *données additionnelles* sont variées et nombreuses (une liste d'apprenants, un scénario d'apprentissage formalisé, une ontologie, une taxonomie du domaine, un curriculum académique, etc.). Tout l'intérêt de cette donnée est d'apporter des compléments d'information utiles et formalisés au calcul des indicateurs et qui sont disponibles avant la session d'apprentissage. Nous avons fait le choix dans un premier temps de ne prendre en compte que des données formalisées car ces données ne nécessitent pas d'évaluation et sont la plupart du temps très bien formatées. Nous pourrions à l'avenir prévoir l'intervention d'un agent humain pour l'extraction du contenu si nécessaire.

La figure 5.5 détaille les trois facettes d'une *donnée additionnelle*. La facette “définition” possède les caractéristiques d'une donnée brute pour des raisons identiques, c'est-à-dire

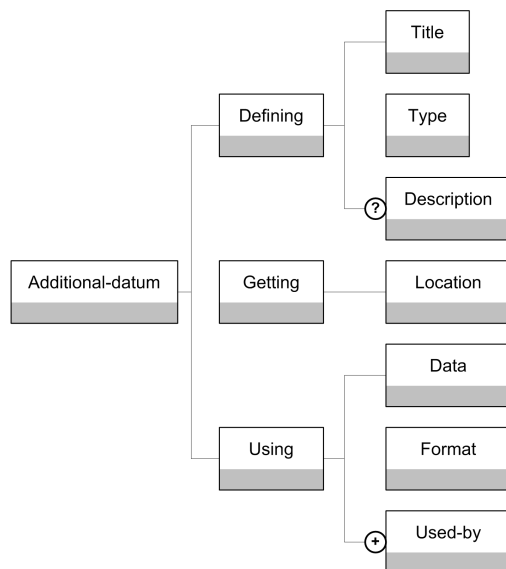


FIGURE 5.5 – Le modèle d’information de la donnée supplémentaire

spécifier la sémantique du contenu de la donnée. Toutefois, face à la diversité des données additionnelles, nous avons ajouté un élément *type* permettant de catégoriser la donnée. Nous n’avons pas actuellement défini de catégories particulières, nous laissons le champ libre pour le prescripteur et l’analyste.

Une *donnée additionnelle* est un élément connu à l’avance et clairement identifié, la facette “obtention” fait donc directement référence à la *localisation* de la donnée.

La facette “utilisation” a pour objectif de décrire le contenu de la donnée, elle est composée du *contenu* de la donnée, de son *format* et de la liste des données qui l’utilisent.

5.3.2.4 La donnée intermédiaire

Après avoir pris en compte les différents types de données primaires, nous avons modélisé la donnée intermédiaire. Il y a une forme de similarité avec le concept d’indicateur intermédiaire proposé par Bousbia et al. [Bousbia 2011]. Il s’agit d’une donnée permettant de calculer une information pouvant être prise en compte dans le calcul de plusieurs indicateurs. Nous ne l’appelons pas indicateur car de notre point de vue, l’indicateur répond à un objectif d’observation du prescripteur. Or, la donnée intermédiaire a pour objectif de faciliter la réutilisation de calculs. La figure 5.6 détaille les trois facettes d’une *donnée intermédiaire*.

Comme pour toutes les autres données, la facette “définition” est composée du *titre* de la donnée, d’une *cardinalité* et d’une éventuelle *description*. Cette facette porte la sémantique de la donnée, même s’il n’y a pas d’objectif d’observation sous-jacent, l’information contenue dans cette donnée a un sens bien précis.

La facette “obtention” caractérise la construction de la donnée, comment la donnée est générée à partir d’autres données appelées *composants* (Component). Ces composants sont utiles à la définition d’un graphe de dépendances de la donnée vis-à-vis de *données primaires* (données brutes, de production ou additionnelles) et/ou de données dérivées

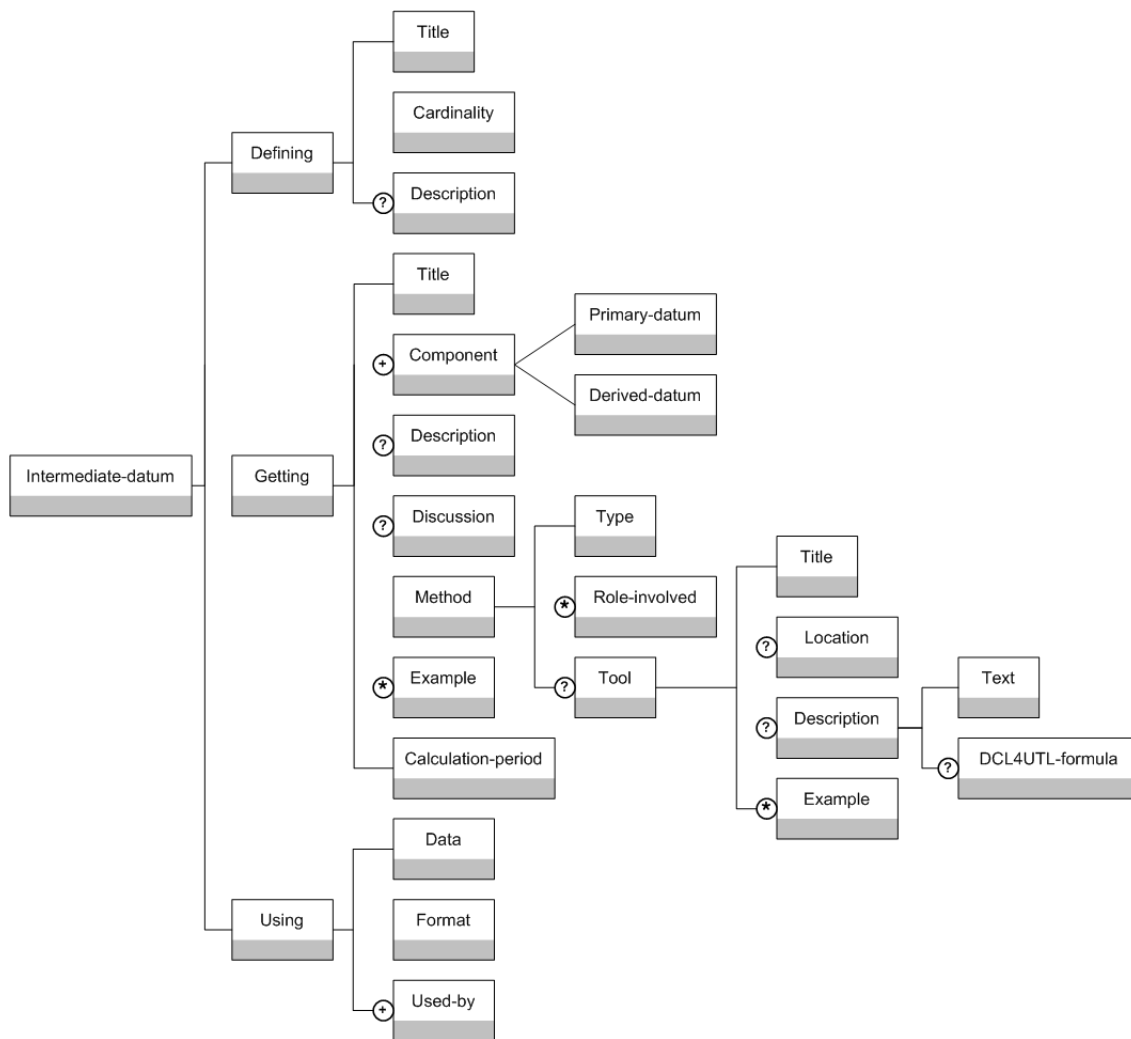


FIGURE 5.6 – Le modèle d’information de la donnée intermédiaire

(données intermédiaires ou indicateurs). C’est cet élément composant qui définit l’ensemble des observables nécessaires à l’établissement de la donnée intermédiaire.

Afin de décrire au mieux ce moyen d’observation capitalisable, la facette “obtention” peut être également composée d’un *titre*, d’une *description* et d’une *discussion* pour préciser la nature de la donnée désirée, et comporte obligatoirement un élément *méthode* (Method) qui peut être du type humain, semi-automatique ou automatique.

Dans le cas où un acteur humain intervient dans l’établissement de la donnée, il faut indiquer son *rôle* (Role-involved). Si la méthode est totalement ou partiellement automatisée, l’outil *support* (Support-tool) doit être indiqué grâce à sa *localisation* dans la mesure du possible, ou par une *description* et quelques exemples. Nous considérons à cette étape qu’un seul outil peut être spécifié pour une donnée intermédiaire. Si toutefois d’autres outils doivent être utilisés, il convient de définir plusieurs données intermédiaires. Le champ *description* se décompose en une information textuelle et/ou une formule de combinaison de données en DCL4UTL, le langage de formalisation de la construction d’une donnée

intermédiaire ou d'un indicateur décrit dans la section 5.3.3.

Comme pour toutes les autres données, la facette "utilisation" représente la donnée une fois évaluée, elle est composée du *contenu* de la donnée, de son *format* et de la liste des données qui l'utilisent.

5.3.2.5 L'indicateur

L'indicateur correspond à l'expression du besoin d'observation spécifié par le prescripteur. A ce titre, il est associé à l'univers métier du concepteur et répond à un objectif d'observation précis. La figure 5.7 détaille les trois facettes d'un indicateur. Les facettes "définition" et "obtention" sont similaires à celles de la donnée intermédiaire.

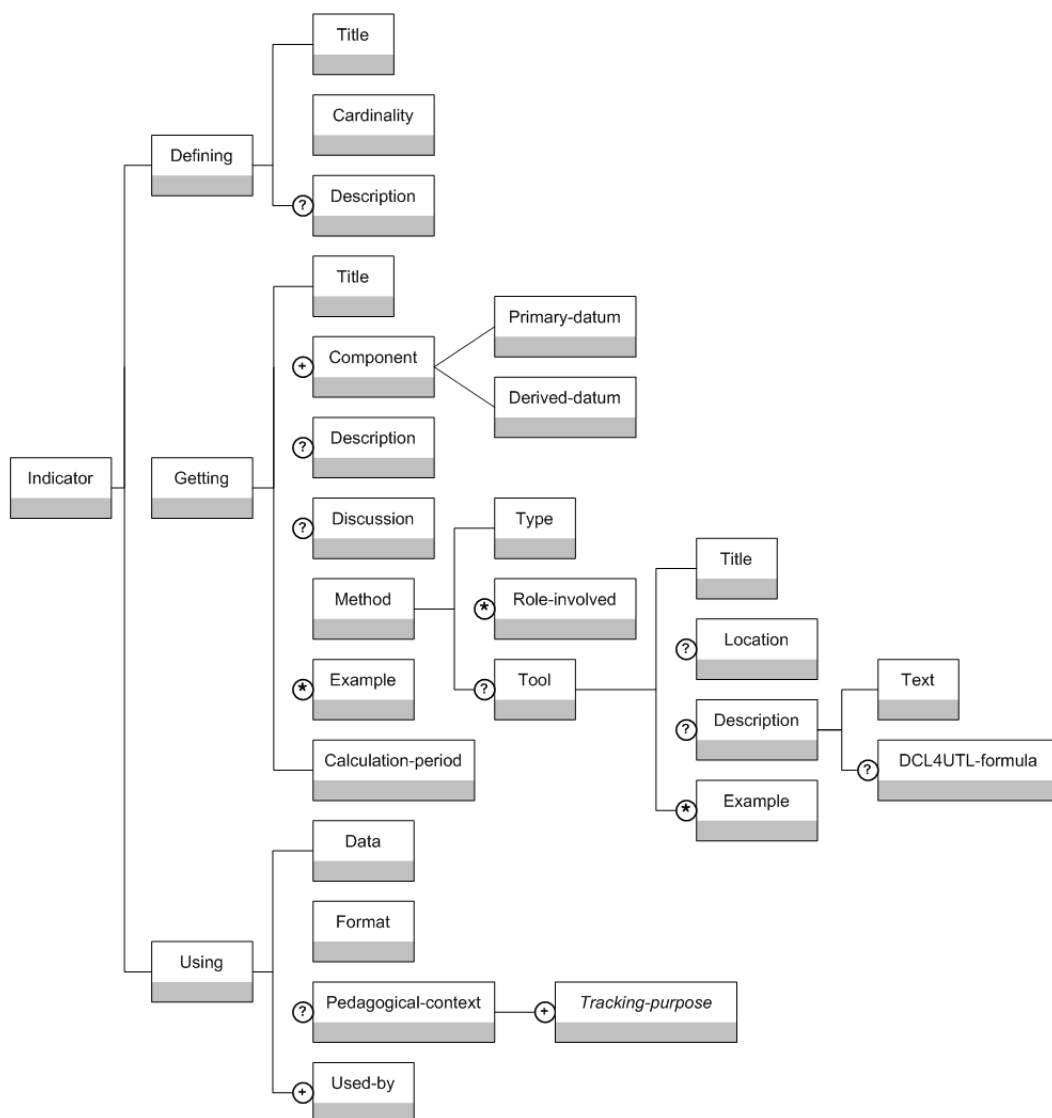


FIGURE 5.7 – Le modèle d'information d'un indicateur

La facette "utilisation" est plus importante que pour les autres données parce qu'elle

décrit l'objectif de l'observation et le lien avec le contexte pédagogique. Elle est caractérisée par quatre éléments :

- l'élément *contenu* (Data) permet de stocker la valeur de l'indicateur, une fois calculé ou établi.
- l'élément *format* (Format) décrit le format de représentation de la donnée.
- l'élément *contexte pédagogique* (Pedagogical-context) définit le ou les *objectifs d'observation* (Tracking-purpose) de l'indicateur. Comme le montre la figure 5.8, l'objectif d'observation permet de comprendre les raisons de la création de cet indicateur. Le prescripteur indique non seulement l'utilisation qui sera faite de l'indicateur mais aussi la cible. L'objectif est décrit par :
 - un ou plusieurs *concepts traçables* (Traceable-concept) qui permettent d'associer l'indicateur à un élément du scénario pédagogique.
 - un *type d'exploitation* (Type). Nous avons défini actuellement cinq types d'exploitation – réingénierie, régulation, évaluation, adaptation et réflexivité. Il s'agit des utilisations les plus courantes des indicateurs, mais nous considérons cette liste ouverte.
 - un ou plusieurs *rôles* (Recipient-role) qui précisent le destinataire de l'indicateur.
 - une *description* qui peut être ajoutée afin de décrire de façon détaillée l'*objectif d'observation*.
- l'élément *utilisé-par* (Used-by) pour le graphe de dépendances des données.

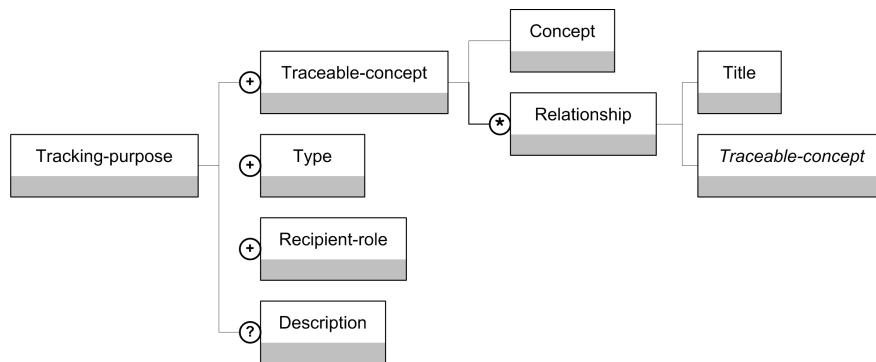


FIGURE 5.8 – Le modèle d'information d'un objectif d'observation

5.3.3 De la spécification à l'opérationnalisation : DCL4UTL

Le langage UTL a été pensé pour être capitalisé. A ce stade, il est possible de capitaliser les descriptions textuelles des moyens d'observation, mais nous restons dans les mêmes critiques que celles émises au sujet du projet DPULS. Afin de rendre la capitalisation plus pertinente, nous avons souhaité proposer une automatisation dans la gestion des données, c'est-à-dire permettre le calcul automatique des indicateurs et des données intermédiaires. La capitalisation de données automatisables facilite la réutilisation et est une motivation supplémentaire pour les usagers. Pour automatiser le calcul tout en restant déclaratif, nous avons proposé dans le cadre de la thèse de Diem Phâm Thi Ngoc, un langage de combinaison de données pour UTL appelé DCL4UTL (Data Combination Language for

UTL). L'objectif de DCL4UTL est de combiner les données UTL afin de produire une nouvelle donnée dérivée, en particulier un indicateur et de capitaliser l'expertise de calcul pour en faciliter la réutilisation [Phâm Thi Ngoc 2011].

UTL étant un langage indépendant du format de représentation et du système de stockage des traces, nous avons considéré qu'il fallait une proposition indépendante d'une représentation technique d'information (base de données). Même si DCL4UTL s'est inspiré des langages de requête classiques (SQL, XQuery, etc.), il reste néanmoins un langage spécifique pour UTL et indépendant d'une architecture de stockage d'information. Nous souhaitons être indépendant de la structure de stockage afin de suivre les évolutions technologiques tout en conservant la maîtrise du langage. L'interpréteur DCL4UTL a pour fonction d'exploiter les règles de calcul et de faire la liaison entre le serveur de stockage des traces et les données UTL.

Comme nous venons de le signaler, notre proposition se base sur les langages de requêtes classiques, au sens où il est nécessaire d'extraire l'information des données formalisée avec UTL, puis de les composer ou de les transformer. La structure du langage est présentée dans la figure 5.9, un exemple de formule de calcul est donné dans la figure 5.10. Un interpréteur a été développé permettant l'automatisation du calcul d'indicateur grâce à ce langage.

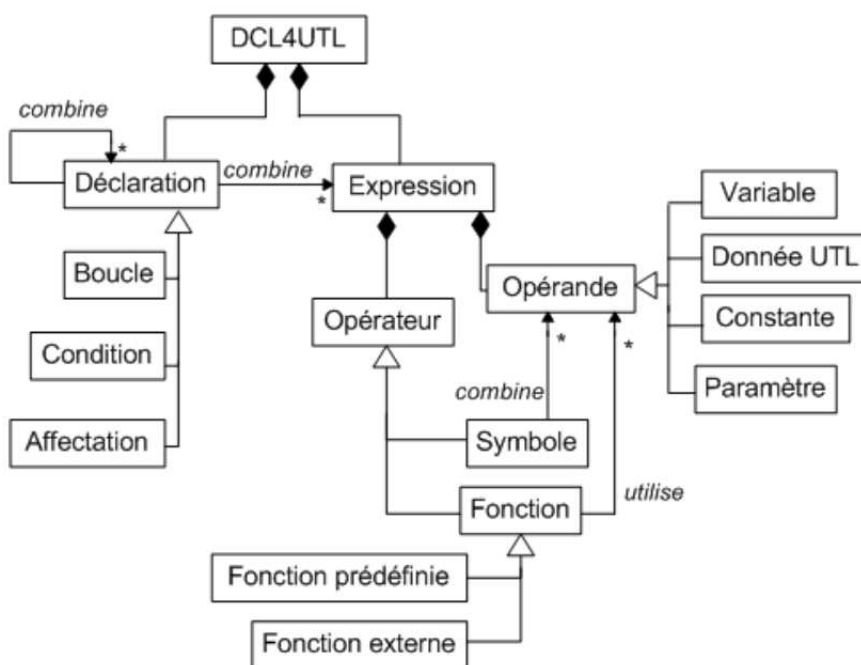


FIGURE 5.9 – Modélisation du langage DCL4UTL

5.3.3.1 Composants du langage

Le langage a été conçu autour d'une grammaire qui est évolutive, actuellement il est composé des structures de contrôle classiques (la boucle "pour", la structure conditionnelle


```

cal {
    for each $i in AD.utilisateur do {
        I.tauxRealisation[patient=i]=
            (I1.nombreBonneActions[patient==i] *100)/12;
    }
}
where
    I = VAPS-I-Taux-Realisation.using.data.indicator,
    I1= VAPS-I-BA.using.data.indicator,
    AD= VAPS-AD-Utilisateur.using.data.additionalDatum

```

FIGURE 5.10 – Exemple de description avec le langage DCL4UTL

et l'affectation), d'un ensemble d'opérateurs mathématiques et logiques (+, -, *, /, <, >, <=, >=, ==, !=, and, or), et d'un ensemble de fonctions internes :

- **first**(*element*) renvoie le premier élément dans un ensemble de données UTL.
- **end**(*element*) renvoie le dernier élément dans un ensemble de données UTL.
- **index**(*element*, *index*) renvoie l'élément à la position *index* dans un ensemble de données UTL, la valeur *index* commence à partir de 1 qui correspond au premier élément de l'ensemble.
- **sum**(*element*) (ou **sumDis**(*element*)) retourne la somme de toutes les valeurs (ou uniquement des valeurs distinctes) d'un élément à valeur numérique.
- **min**(*element*) (et **max**(*element*)) donne la valeur minimale (et maximale) dans l'ensemble d'occurrences d'un élément à valeur numérique.
- **count**(*element*) (ou **countDis**(*element*)) compte le nombre d'occurrences (ou d'occurrences distinctes) d'un élément de donnée UTL.
- **concat**(*str1*, *str2*) concatène une chaîne de caractères *str2* à une autre *str1*.
- **filter**(*crit*) extrait des données dans un ensemble d'éléments de données UTL selon le critère *crit* qui est une condition de sélection sur les données.
- **sort**(*crit*) trie des données selon un ou plusieurs éléments avec un critère croissant (asc) ou décroissant (desc).

5.3.3.2 Les fonctions externes

Dès le début, dans la conception du langage, nous avons souhaité qu'il soit ouvert, c'est-à-dire qu'il permette l'utilisation de fonctions ou services externes. En effet, il peut arriver que le calcul nécessite un algorithme trop compliqué à décrire et surtout ayant été implémenté de façon optimale dans un autre langage. L'idée était donc de permettre la description d'une fonction externe et de l'utiliser comme une fonction interne dans le langage DCL4UTL. Cette fonctionnalité a donc été intégrée, pour des raisons purement techniques, nous avons fait le choix de nous limiter dans un premier temps à l'intégration de bibliothèques JAVA. Dans le cadre de l'environnement d'apprentissage de la programmation Hop3X développé au LIUM, nous avons été amené à utiliser un programme JAVA réalisant l'introspection des programmes développés par les étudiants, les indicateurs se sont alors basés sur ces résultats d'introspection.

5.4 Synthèse

Nous venons de présenter un langage déclaratif permettant de modéliser et de calculer l'ensemble des données d'observation. Ce langage constitue notre réponse à la question de recherche *“Comment formaliser le besoin d'observation et le calcul de l'indicateur correspondant afin de le rendre automatisable, réutilisable et adaptable ?”* (Q2) ainsi qu'une ébauche de réponse à la question *“Comment prendre en compte la connaissance métier dans des indicateurs basés sur des traces spécialisées ?”* (Q3). En effet, nous intégrons des concepts métiers pour la sémantique des données d'observation, mais il est nécessaire de développer ces aspects afin de disposer d'une connaissance métier plus approfondie, comme une ontologie.

En ce qui concerne les évolutions, la première concerne le modèle DGU, en effet, nous avons pris en compte la sémantique du besoin d'observation de manière textuelle, or il serait intéressant de modéliser les concepts métiers et de les manipuler lors de la spécification du besoin (la section Définition). Cela nous permettrait d'être au plus proche de l'activité du prescripteur et de pouvoir retranscrire le résultat de l'observation directement au niveau métier.

L'environnement UTL – dans sa version actuelle – ne couvre pas toute la chaîne de l'observation. En effet, même s'il est conçu de manière déclarative pour faciliter la capitalisation, il est nécessaire (comme cela a été étudié dans le cadre du projet DPULS) de relier les indicateurs à des usages et à des réactions. Il nous semble pertinent de mémoriser la réaction de l'utilisateur sur l'indicateur, par exemple pour le tuteur l'action de régulation qu'il a menée et son point de vue sur le résultat. Cela permettra d'aider les autres usagers dans la conception de leur situation d'apprentissage ou de leur scénario d'observation.

Chapitre 6

Instrumentation et mise en œuvre

Sommaire

6.1	Introduction	99
6.2	Mise en œuvre du processus d'ingénierie	99
6.2.1	Principe général	99
6.2.2	L'éditeur de données	99
6.2.3	L'interpréteur UTL	100
6.2.4	La réutilisation des données	101
6.3	Un exemple concret : Les indicateurs dans Hop3X	102
6.3.1	Présentation de l'environnement	102
6.3.2	La modélisation d'indicateurs	105
6.4	Synthèse	110

6.1 Introduction

Nous venons de présenter l'ingénierie de l'observation par la prescription comme cadre à nos travaux ainsi que la méthodologie de représentation des données d'observation associée. Afin d'aider l'enseignant dans la compréhension de sa situation d'apprentissage, il est nécessaire d'instrumenter le processus d'observation. Ce chapitre a pour objectif principal de présenter la mise en œuvre de notre approche. Pour cela, nous avons conçu des outils permettant de gérer la spécification des données d'observation ainsi que leur utilisation sur une session d'apprentissage.

6.2 Mise en œuvre du processus d'ingénierie

6.2.1 Principe général

La mise en œuvre du processus d'ingénierie de l'observation au travers du langage UTL consiste à modéliser les données grâce à l'éditeur puis à exploiter ces données dans une plate-forme ou grâce à une application séparée si la plate-forme n'est pas modifiable. Le prescripteur définit ses besoins d'observation, ce qui revient à instancier des indicateurs et à renseigner les facettes "Defining" et "Using". La facette "Using" est à renseigner afin de préciser l'objectif d'observation, puis en accord avec le développeur, le prescripteur indique le format de représentation du résultat de l'indicateur. Le développeur de son côté modélise les données brutes en fonction des connaissances qu'il a de l'EIAH et des observés disponibles. L'analyste exploite les informations du prescripteur afin de terminer la modélisation des indicateurs. Pour cela, il sera amené à modéliser des données intermédiaires ainsi que des données de production et des données additionnelles. En effet, l'analyste peut avoir besoin par exemple de la liste des apprenants, du scénario, ou encore de réponses à des exercices qui seront réalisés pendant la session.

6.2.2 L'éditeur de données

Afin de faciliter la prise en main du langage UTL pour la description des indicateurs et des données associées, nous avons développé un prototype d'éditeur qui se présente sous la forme d'un site internet (cf. Figure 6.1). Ce prototype prend en compte la totalité des données modélisées, par contre, il n'a pas été conçu pour utiliser des données instanciées. Afin de rendre possible la navigation dans l'ensemble des données instanciées, il serait intéressant d'adapter le "browser" de patrons dont nous avons eu la responsabilité dans le projet DPULS. L'éditeur pour UTL permet de naviguer dans les collections d'indicateurs, d'en spécifier de nouveaux ou de créer un indicateur sur la base d'un autre. Il a été conçu de façon à gérer différents rôles et à paramétrer l'interface selon ces rôles. Par exemple, le prescripteur pourra avoir une vue réduite des modélisations et l'analyste une vue plus détaillée.

Au sein de l'éditeur, le prescripteur ou l'analyste crée un projet, puis dans le projet, les différents protagonistes modélisent l'ensemble des données d'observation. Par défaut, les données modélisées sont restreintes au projet dans lequel elles ont été conçues. Pour la

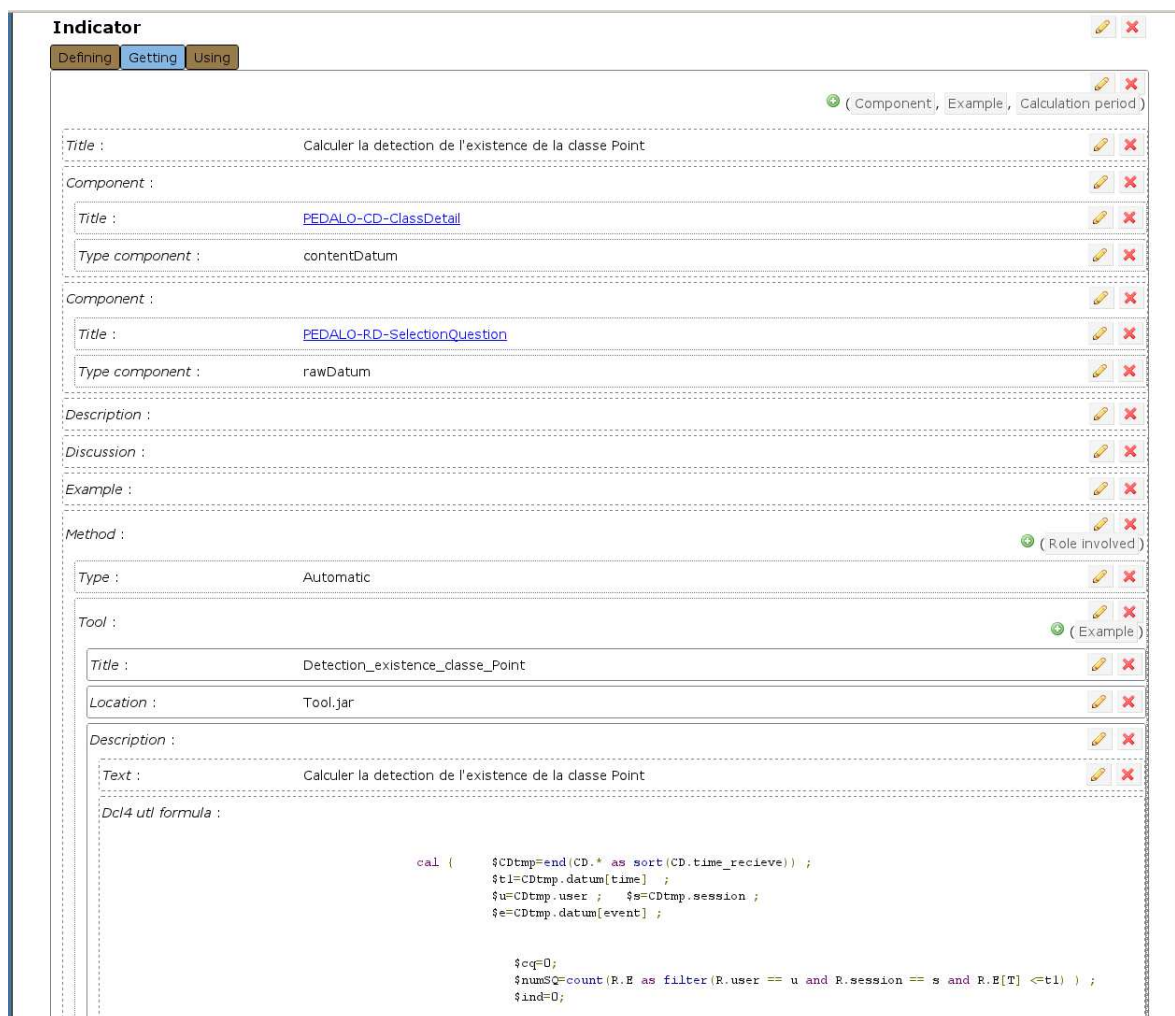


FIGURE 6.1 – Copie d'écran de l'éditeur Edit'UTL sur la partie obtention de l'indicateur

capitalisation, et parce que le langage a été conçu pour être réutilisé, nous avons intégré la possibilité pour le prescripteur de partager ses modélisations de données d'observation en les rendant publiques. Cette fonctionnalité permet à d'autres prescripteurs de construire leurs données en utilisant des données publiques comme base.

Afin de prendre en compte les évolutions possibles du langage, l'éditeur se base sur un schéma XML représentant le langage UTL. Toute modification dans le schéma est prise en compte dans l'interface, tant du point de vue des champs que de la vérification syntaxique en temps réel.

6.2.3 L'interpréteur UTL

L'instrumentation d'un langage comme UTL nécessite de disposer d'un environnement capable d'interpréter les données et de valuer les indicateurs. L'environnement développé est basé sur le langage JAVA, un analyseur lexical pour le langage DCL4UTL ainsi qu'une base de données XML et son langage de requête XQuery. Il est composé de différents

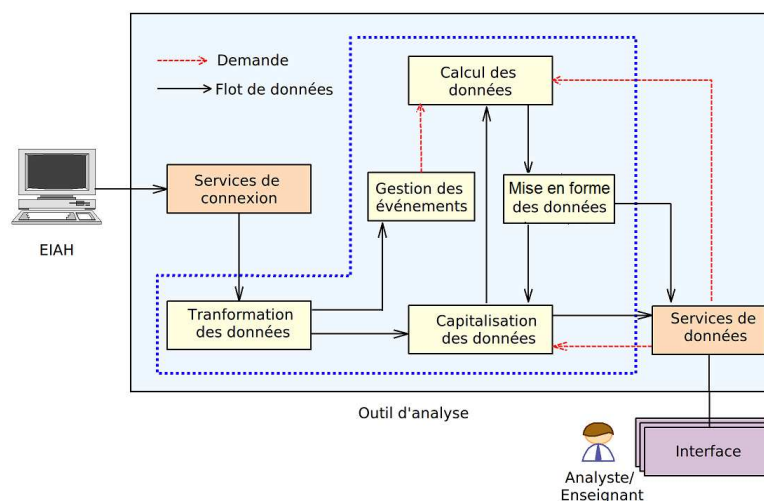


FIGURE 6.2 – Architecture générale de l'interpréteur UTL [Phàm Thi Ngoc 2011]

services (cf. Figure 6.2) dont ceux permettant d'interagir avec l'EIAH, les services de connexion et de données. Au fur et à mesure de la disponibilité des traces, elles sont transformées en données brutes puis capitalisées. Le système gère les événements liés au déclenchement du calcul des indicateurs. Il s'agit d'un système différent d'IDLS, nous n'avons pas la notion de notification. L'indicateur est calculé au moment souhaité par le prescripteur, si au moins une des données nécessaire au calcul n'est pas disponible, l'indicateur n'est pas calculé. Enfin, les données calculées sont mises en forme selon le format décrit dans la représentation de ces données puis capitalisées et envoyées au service de données qui les transmet soit à l'EIAH, soit à un outil externe de perception du résultat des indicateurs.

6.2.4 La réutilisation des données

Comme nous venons de le signaler l'éditeur a été conçu pour partager (rendre public) des modélisations de données d'observation et permettre à d'autres prescripteurs de les exploiter. La réutilisation peut se faire à différents niveaux :

- la réutilisation d'une donnée pour le calcul d'une autre donnée dérivée avec un même EIAH, dans ce cas il n'y a pas de modification, il s'agit d'une réutilisation que nous qualifions de "simple".
- l'exploitation des données d'observation sur un autre EIAH, ici toutes les descriptions sont conservées, il suffit de modifier éventuellement les éléments "track" des données brutes pour les adapter au nouvel environnement. En effet, la donnée brute apporte une valeur signifiante et qualifiée sémantiquement, l'élément "track" associé correspond à la collecte de cette valeur dans le format de traces de l'environnement.
- la réutilisation d'un moyen d'observation est possible lorsque le calcul d'un indicateur est identique à un autre indicateur, par exemple entre les indicateurs "fréquence de connexion au forum" et "fréquence de connexion au test d'auto-évaluation". Le moyen d'observation correspond à la règle de calcul de la donnée dérivée. Pour le

réutiliser, il faut :

- indiquer les modifications sur l'ensemble des données d'observation nécessaire au calcul. Dans l'exemple, il suffit de remplacer la donnée primaire “connexion au forum” par “connexion au test d'auto-évaluation”, le calcul de fréquence reste le même.
- conserver la description du calcul de fréquence qui n'a pas de raison de changer.
- adapter la description du nouvel indicateur. La sémantique portée par l'indicateur est différente.

6.3 Un exemple concret : Les indicateurs dans Hop3X

Le département informatique de l'Université du Maine utilise depuis quelques années un environnement informatisé développé en interne par le LIUM pour l'apprentissage de la programmation. Nous avons décidé de nous associer aux auteurs de façon à répondre aux besoins d'observation liés à cet apprentissage et de mettre en pratique l'environnement UTL. Cette utilisation de notre approche est en fait une instanciation de l'exemple en régulation proposé dans la section 4.6.2 du chapitre 5. Nous commençons par présenter l'environnement, puis nous présentons la méthodologie utilisée pour définir les indicateurs liés à cette activité.

6.3.1 Présentation de l'environnement

Hop3X est un environnement de travaux pratiques pour l'apprentissage de la programmation. Il permet au tuteur de suivre, à distance, l'activité de programmation des apprenants en lui fournissant, d'une part, des informations qualitatives (indicateurs) sur cette activité, et d'autre part, une visualisation en temps réel des productions (code source des programmes). L'environnement est développé en Java et tourne sur différents systèmes (Testé sur MacOS, Windows et Linux).

Dans Hop3X, l'activité de l'apprenant est tracée. Chaque action produit un “Événement XML Hop3X” qui est envoyé au serveur. Ce dernier enregistre les événements pour des utilisation synchrones et asynchrones :

- observer : le serveur envoie tous les événements depuis les apprenants connectés vers les enseignants/tuteurs connectés. Cela permet aux enseignants d'observer l'activité des apprenants ainsi que leurs productions en temps réel.
- recharger : ce processus consiste à reconstruire les productions des apprenants (par exemple des fichiers de code Java). Cela fonctionne lorsque les apprenants et les enseignants/tuteurs sont connectés.
- rejouer : grâce à la reconstruction des productions des apprenants, il est possible pour l'enseignant de rejouer l'activité de l'apprenant au moment de son choix. Il est aussi possible pour l'enseignant de naviguer de compilation en compilation ou d'exécution en exécution.
- simuler : le processus de simulation est un outil à destination des chercheurs, il simule une session en chargeant les traces comme si les apprenants étaient connectés à l'environnement.

- calculer des indicateurs : en utilisant les événements Hop3X, UTL calcule des indicateurs révélant des informations de haut niveau sur les activités des apprenants.

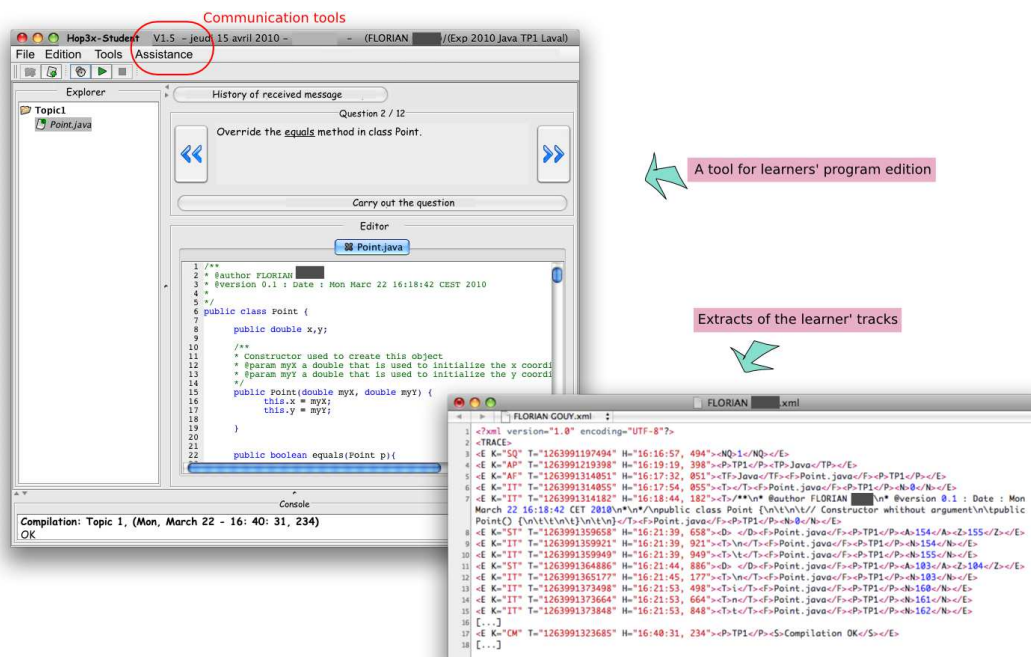


FIGURE 6.3 – Interface Hop3X-Student

L'environnement Hop3X est composé de trois applications :

- Hop3X-Student (Figure 6.3) : elle permet aux apprenants d'éditer, de compiler et d'exécuter du code et des programmes. L'interface permet aussi d'appeler l'enseignant/tuteur afin de solliciter son aide au travers d'un outil de communication audio.
- Hop3X-Server : ce serveur permet la collecte des traces d'interaction des apprenants, de les sauvegarder sous la forme d'événements Hop3X. Il permet le calcul en temps réel des indicateurs.
- Hop3X-Teacher (Figure 6.4) : il s'agit de l'outil de suivi et de régulation de l'enseignant/tuteur. Il permet de gérer un groupe d'apprenants dans une situation de travaux pratiques à distance et synchrone. Il permet aussi aux enseignants/tuteurs de :
 - suivre les activités des apprenants en temps réel via l'interface de visualisation.
 - disposer d'informations de haut niveau sur les productions et les tâches des apprenants grâce aux indicateurs et aux outils de visualisation associés.
 - annoter une partie du programme des apprenants et leur permettre de consulter cette annotation.
 - sélectionner les indicateurs et intervenir au travers d'un outil de communication.
 - rejouer le parcours d'un apprenant en cours de session ou après la session.
 - envoyer un message écrit ou discuter avec les apprenants au travers des outils de communication.
 - visualiser l'historique de leurs interventions au travers d'un outil-mémoire.

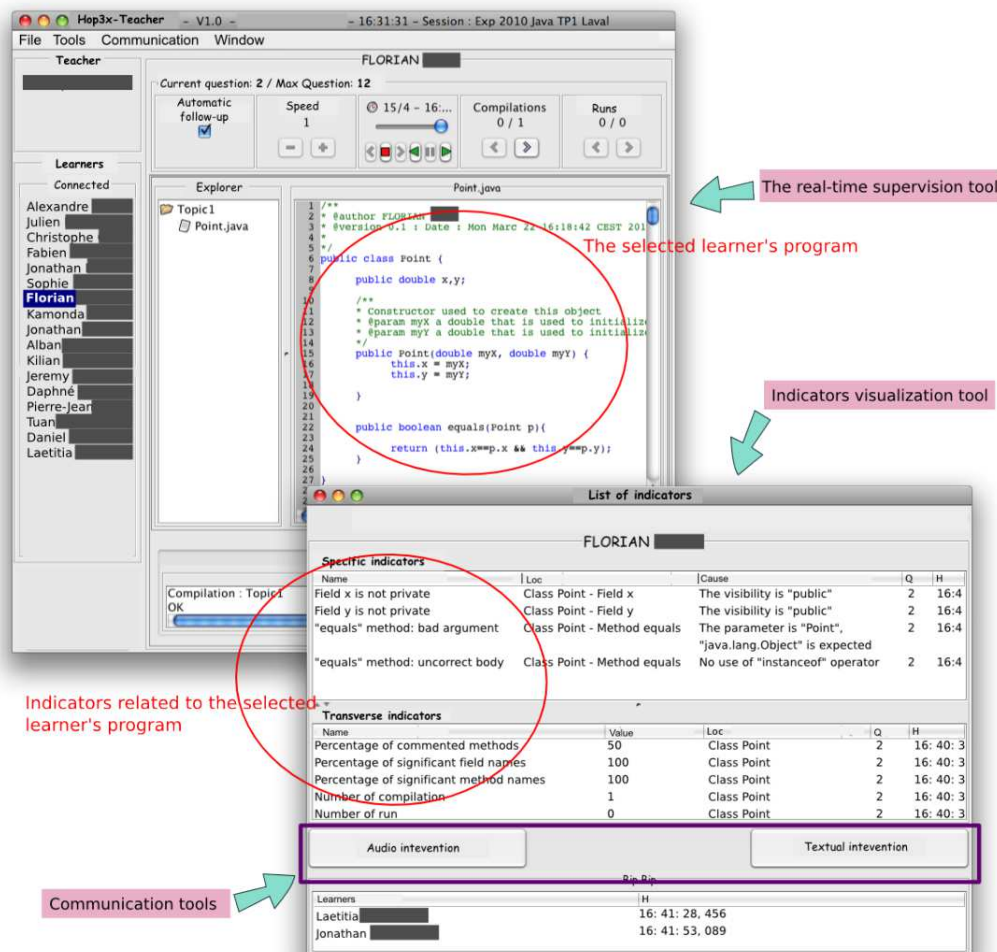


FIGURE 6.4 – Interface Hop3X-Teacher avec les indicateurs venant d’UTL

Pendant la session d’apprentissage, il n’y a pas d’interactions en face à face entre les apprenants et les enseignants/tuteurs. Ces derniers peuvent suivre les tâches et activités des apprenants (les programmes et le code java des apprenants) en temps réel. Afin de réagir et d’interagir avec les apprenants, ils disposent de deux moyens de communication : (i) envoyer un message ; (ii) dialoguer au travers d’un dispositif micro-casque. Ils peuvent associer leurs interventions à un ensemble d’indicateurs calculés par le système. Dans les versions précédentes de Hop3X, le calcul des indicateurs se trouvait dans le code source de l’outil. Chaque modification ou ajout d’indicateur nécessitait de recompiler l’application. L’utilisation d’UTL a permis de s’affranchir de ces contraintes. Les enseignants/tuteurs peuvent être pro-actifs ou réactifs, en fonction des valeurs des indicateurs ou de la sollicitation des apprenants.

Durant les séances de travaux pratiques, les enseignants/tuteurs ont pour tâche de :

- Gérer le groupe d’apprenants, le temps et l’avancement de la séance.
- Assister les apprenants dans l’acquisition de la connaissance et du savoir-faire.
- Guider les apprenants dans l’acquisition de bonnes pratiques de programmation.

6.3.2 La modélisation d'indicateurs

Les indicateurs ont été modélisés dans l'éditeur UTL en fonction des besoins d'observation exprimés par l'enseignant/concepteur. Les besoins d'observation correspondent à deux catégories d'indicateurs, les indicateurs spécifiques à chaque question du sujet et les indicateurs transversaux à toutes les questions. En tout, plus de 80 indicateurs ont été modélisés et exploités lors d'une séance de travaux pratiques. Nous tenons à préciser que le format d'expression des indicateurs n'est pas de notre ressort mais correspond à une demande des développeurs de l'environnement Hop3X pour une meilleure intégration dans leur système. Les auteurs de l'outil tenaient à ce que le format de représentation des indicateurs soit unique, il est de fait complexe et possède un certain nombre d'attributs non utilisés dans le calcul, mais qui doivent néanmoins être renseignés. Cette solution alourdit fortement la syntaxe de la règle de calcul.

Nous allons présenter le sujet de la première séance ainsi que quelques indicateurs demandés par l'enseignant. Ensuite, nous présenterons la méthodologie suivie pour la modélisation des indicateurs.

6.3.2.1 Sujet de la première séance

L'étudiant dispose de douze questions à réaliser dans la séance :

1. Ecrire la classe Point composée de deux champs (valeurs réelles) représentant les coordonnées d'un point en x et y.
2. Redéfinir la méthode equals pour la classe Point.
3. Ajouter un comportement, que l'on appellera distance, permettant à un objet de la classe Point de calculer la distance qui le sépare d'un autre point.
4. Un triangle étant défini par trois points, écrire la classe Triangle.
5. Ajouter un comportement, que l'on appellera périmètre, permettant à un objet de la classe Triangle de calculer son périmètre.
6. Ajouter un comportement, que l'on appellera surface, permettant de calculer la surface d'un triangle.
7. Ecrire la classe Droite, une droite étant définie par deux réels (a et b) représentant les coefficients de l'équation de droite $y=ax+b$.
8. Faire le nécessaire pour qu'il soit possible de créer une droite à partir de deux points (c'est-à-dire la droite qui passe par ces deux points).
9. Ajouter, à la classe Point, un comportement, que l'on appellera estSurDroite, permettant à un point de tester s'il se trouve sur une droite.
10. Ajouter, à la classe Droite, un comportement, que l'on appellera estParallele, permettant à une droite de tester si elle est parallèle à une autre.
11. Ajouter, à la classe Droite, un comportement, que l'on appellera intersection, permettant à une droite de calculer son intersection avec une autre droite.
12. Faire le nécessaire pour qu'il soit possible de créer un Triangle à partir de trois droites.

6.3.2.2 Quelques indicateurs transversaux

1. La fréquence de compilation manuelle (par minute). Cet indicateur a pour but de vérifier si l'apprenant compile régulièrement son programme.
2. La fréquence d'exécution (par minute). Cet indicateur a pour but de vérifier si l'apprenant exécute régulièrement son programme.
3. Le taux de correction des erreurs à la compilation. Cet indicateur a pour but de vérifier si une même erreur revient souvent ou si l'étudiant a assimilé qu'il s'agissait d'une erreur.
4. Le pourcentage de noms de variables significatifs (On se basera sur la longueur des noms). Cet indicateur aide l'enseignant/tuteur à savoir si l'étudiant respecte les règles de programmation. Le calcul de cet indicateur prenant en compte la longueur des noms des variables n'est pas de notre choix. Il s'agit d'une solution préconisée par l'enseignant.
5. Le pourcentage de noms de méthodes significatifs (On se basera aussi sur la longueur des noms).
6. Le pourcentage de variables d'instances privées. Cet indicateur aide l'enseignant/tuteur à connaître le pourcentage de variables d'instances privées par rapport à d'autres variables (variables d'instances publiques et variables de classe).

6.3.2.3 Quelques indicateurs spécifiques

– Pour la question 1 :

1. La détection de l'existence d'une classe s'appelant "Point". Cet indicateur a pour but de vérifier si l'apprenant respecte la consigne et s'il sait comment créer une classe. L'existence d'une classe "Point" est la valeur attendue de cet indicateur.
2. Quelle est la visibilité de la classe "Point" ? "public" est la valeur que l'enseignant veut avoir.
3. Le nombre de champs de la classe Point. La classe Point est composée de deux données représentant les coordonnées d'un point en x et y. Donc, la valeur attendue de cet indicateur doit être 2.
4. Quels sont les types de donnée de ces champs ? Les valeurs des coordonnées d'un point en x et y doivent être réelles. Donc, la valeur attendue de cet indicateur doit être "float" ou "double".
5. Quelle est la visibilité de ces champs ? "private" est la valeur que l'enseignant veut avoir pour cet indicateur.
6. La détection de l'existence d'un constructeur spécifique à 2 arguments.
7. Quels sont les types de donnée des arguments de ce constructeur ?

6.3.2.4 Mise en œuvre des indicateurs avec UTL

Afin d'utiliser UTL dans un EIAH, les étapes suivantes sont à respecter :

1. Collecter les besoins d'observation.
2. Identifier les traces puis modéliser l'information contenue dans ces traces à l'aide des données brutes.
3. Modéliser les autres données primaires (additionnelles, de production).
4. Modéliser, à l'aide de l'éditeur, les différents indicateurs avec les données intermédiaires associées.
5. Collecter les traces.
6. Calculer les indicateurs modélisés.

Avec Hop3X, la collecte des besoins d'observation de l'enseignant s'est faite au travers de la liste des différents indicateurs qui était disponible. En ce qui concerne l'identification des traces, le format des traces d'Hop3X est unique, il s'agit d'une balise XML dont les valeurs des attributs changent selon l'évènement.

6.3.2.5 La description d'un indicateur avec UTL

Nous présentons la modélisation avec UTL de l'indicateur transversal "Calcul du taux de correction des erreurs à la compilation". Cet indicateur nécessite l'utilisation d'une librairie externe, c'est pour cela que nous l'avons sélectionné dans cette présentation. Afin de construire cet indicateur, il est nécessaire de s'intéresser à la façon dont le calcul doit être réalisé, cela permet de mettre en évidence le besoin d'outils externe ainsi que la liste des données brutes nécessaires au calcul. Dans cet exemple, les données brutes nécessaires correspondent aux évènements "changer de question" et "compilation manuelle" qui sont les deux évènements concernés par la compilation. Une fonction externe est nécessaire pour l'instrospection du code Java qu'il n'est pas possible de décrire avec DCL4UTL.

Une fois ces informations regroupées, il est nécessaire de s'interroger sur la nécessité de création de données intermédiaires. En effet, étant donné que la totalité des indicateurs à élaborer est connue, il est possible que certains calculs ou certaines transformation de données soient utiles pour plusieurs indicateurs. Dans ce cas, il est préférable de créer des données intermédiaires, voire des données intermédiaires paramétrables lorsque la règle de calcul est identique à un ensemble de valeurs près.

Title	PEDALO-I-ErrorCorrectionRate
Cardinality	1
Description	Calcul du taux de correction des erreurs à la compilation (Est-ce qu'une même erreur revient souvent ou est-ce que l'étudiant a assimilé qu'il s'agissait d'une erreur?)

TABLE 6.1 – Section Defining de l'indicateur "Calcul du taux de correction des erreurs à la compilation"

Le choix de l'analyste, dans cet exemple, a été de ne pas utiliser de données intermédiaires.

La section Defining

Il s'agit ici d'une description (cf. Table 6.1), généralement donnée par le prescripteur, permettant de retrouver et d'identifier un indicateur. C'est pour cette raison que cette section est relativement simple. Ici, le titre sert à filtrer les données par rapport au nom indiqué, la description donne la sémantique de la donnée, et la cardinalité précise qu'il n'y aura qu'une instance de la donnée dans le résultat.

Title	Calculer le taux de correction des erreurs à la compilation.				
Description					
Component	Title	PEDALO-RD-CompilationManuelle			
	Type component	rawDatum			
Component	Title	PEDALO-RD-SelectionQuestion			
	Type component	rawDatum			
Method	Type	Automatic			
	Tool	Title	utl.external.hop3x.CompilationErrorDetection.numberError		
		Location	Tool.jar		
		description	Text	Calculer le taux de correction des erreurs à la compilation	
			DCL4UTL Formula	Figure 6.4	
Discussion					
Example					
Calculation period	onEvent				

TABLE 6.2 – Section Getting de l'indicateur “Calcul du taux de correction des erreurs à la compilation”

La section Getting

La description du moyen d'obtention de l'indicateur (cf. Table 6.2) correspond ici à une spécification en langage DCL4UTL utilisant deux données brutes (i.e. deux événements Hop3X) qui sont : (i) la demande de compilation de la part de l'apprenant (PEDALO-RD-CompilationManuelle); (ii) le changement de question qui implique compilation et exécution (PEDALO-RD-SelectionQuestion).

Dans ce calcul la variable temps est nécessaire. Les compilations ne sont pas numérotées donc nous devons nous baser sur l'instant de compilation pour identifier les deux programmes à comparer en terme de nombre d'erreurs.

La section Using

Cette description correspond à la spécification de l'indicateur, c'est-à-dire le modèle et non l'instance, par conséquent le champ *Data* n'est pas renseigné, par contre nous y trouvons la syntaxe de formatage du résultat. Il s'agit d'un indicateur élaboré dans un but d'évaluation de l'activité de l'apprenant par l'enseignant ou le tuteur.

La formule de calcul écrite en DCL4UTL (1) (cf. Table 6.4) est assez complexe étant donné le format unique d'expression d'une donnée imposée par les développeurs de l'environnement Hop3X.

Data		
Format	<indicator type="PEDALO-I-ErrorCorrectionRate"><result session="String" student="String" time="String" event="String" value="String" status="String" calculable="String" reference="String" location="String" currentQuestion="String" concernedQuestion="String" category="String " criticalLevel="String"/></indicator>	
Pedagogical context	Tracking purpose	
	Traceable concept	Concept Activity
	Type	Assessing
	Recipient role	Tutor
	Recipient role	Teacher
	Description	
Used by		

TABLE 6.3 – Section Using de l’indicateur “Calcul du taux de correction des erreurs à la compilation”

```

cal {
  $t1 = end(R.E[T] as sort(R.E[T])); $e = R.E[K] as filter(R.E[T] == t1); $v1 = 0;
  $v = R.user as filter(R.E[T] == t1); $s = R.session as filter(R.E[T] == t1); $cq = 0;
  $numQ = count(R1.E as filter(R1.user == v and R1.session == s and R1.E[T] &lt;= t1 ));
  $ind = 0;
  if(numQ &gt;1 ) {
    if (e == "SQ") {ind = numSQ -1;}
    else {ind = numQ;}
    $t3 = index(R1.E[T] as filter(R1.user == v and R1.session == s and R1.E[T] &lt;= t1)
              sort(R1.E[T]), ind );
    cq = R1.E.NQ as filter ( R1.user == v and R1.session == s and R1.E[T] == t3);}
  else {
    if (e == "SQ") {cq = 0;}
    else {
      $t3 = index(R1.E[T] as filter(R1.user == v and R1.session == s and R1.E[T] &lt;= t1)
                sort(R1.E[T]), numQ );
      cq = R1.E.NQ as filter ( R1.user == v and R1.session == s and R1.E[T] == t3);}
    }
  }
  $numCompil = count(R.E as filter(R.user == v and R.session == s) );
  $numCurrent = externe utl.external.hop3x.CompilationErrorDetection
    numberError(R.E.S as filter(R.user == v and R.E[T] == t1));
  if(numCompil &gt; 1) {
    $t3 = index(R.E[T] as filter(R.user == v and R.session == s) sort(R.E[T]), numCompil -1 );
    $numPrv = externe utl.external.hop3x.CompilationErrorDetection
      numberError(R.E.S as filter(R.user == v and R.E[T] == t3));
    if(numPrv == 0) {
      if(numCurrent == 0) {v1 = "Pas d'erreurs entre deux compilations !!!";}
      else { v1 = "Nouvelles erreurs : "; v1 = concat(v1, numCurrent);}
    }
    else {v1 = ((numPrv - numCurrent)/numPrv)* 100; v1 = concat(v1, "%");}
  }
  else {v1 = "Premiere compilation !";}
  I.result[session = s][student = v][time = t1][event = e][value = v1][status = "OK"]
    [calculable = "OK"][reference = ""][location = ""][currentQuestion = cq]
    [concernedQuestion = ""][category = "transversal"][criticalLevel = "" ] = "";
}
where I = PEDALO-I-ErrorCorrectionRate.using.data.indicator,
       R = PEDALO-RD-CompilationManuelle.using.data.rawDatum,
       R1 = PEDALO-RD-SelectionQuestion.using.data.rawDatum

```

TABLE 6.4 – Formule de calcul de l’indicateur “Calcul du taux de correction des erreurs à la compilation”

6.4 Synthèse

Dans le cadre de l'instrumentation du processus d'ingénierie de l'observation, c'est-à-dire de notre proposition de chaîne éditoriale de l'observation, nous avons proposé un premier langage de modélisation sémantique des traces au travers du modèle de représentation de scénario pédagogique utilisé par l'enseignant ou le concepteur. Puis nous avons proposé une extension de ce langage qui correspond d'une part à une instanciation du modèle DGU (Defining-Getting-Using) de représentation de données, et d'autre part à une représentation formelle des données d'observation. Nous avons considéré différents types de données d'observation selon leurs usages et leurs provenances. Chaque donnée est donc modélisée selon les trois points de vue : définition (Defining), obtention (getting) et utilisation (Using). L'environnement UTL permet une spécification déclarative et calculable, il permet de formaliser le moyen d'obtention de la donnée, comme par exemple, la formule de calcul d'un indicateur.

Nous avons conçu un éditeur pour le langage UTL, ce dernier permet de parcourir l'ensemble des données modélisées et de modéliser de nouvelles données. Le langage DCL4UTL couplé à son interpréteur permet l'automatisation du calcul du moyen d'observation. Les données ainsi modélisées sont capitalisées et aussi réutilisables. En ce qui concerne leurs spécifications, elles ne dépendent ni d'un format de trace particulier, ni d'un langage de scénarisation pédagogique. La modification de l'élément *Track* des données brutes permet d'adapter les indicateurs à d'autres environnements de formation. Il s'agit d'une étape qui est prévue mais non outillée pour le moment.

L'expérimentation avec Hop3X a permis de modéliser 83 indicateurs et de recueillir 218773 instances de données brutes pour 6 groupes de 14 étudiants en une seule session d'apprentissage. La session a produit 135581 instances d'indicateurs. En moyenne, environ 3.37 données brutes sont produites par seconde et 2.09 indicateurs sont calculés par seconde. Les interventions du tuteur ont quadruplé, ce qui prouve le bénéfice en terme d'aide à la régulation apporté par l'observation. En fait, le tuteur disposant d'un accès direct aux résultats d'analyse, a pu intervenir plus rapidement, toutefois à raison de 2 indicateurs par seconde, le risque d'une surcharge cognitive n'est pas à écarter. Il serait utile d'étudier de nouvelles modalités de perception des indicateurs ou de filtrer ceux qui sont présentés. Actuellement, tous les indicateurs sont présentés au tuteur.

Dans le chapitre suivant, nous allons introduire les questions de recherche et les évolutions envisagées par rapport à nos travaux.

Chapitre 7

Conclusion et perspectives

Sommaire

7.1 Synthèse des activités menées	113
7.2 Partager et réutiliser au sein de la communauté EIAH	114
7.3 Perspectives à moyen terme	116
7.3.1 L'expression du besoin d'observation dans l'univers métier du prescripteur	117
7.3.2 Le retour d'information ou comment percevoir les indicateurs?	117
7.3.3 La définition de nouveaux indicateurs en session	117
7.3.4 Vers une composition encore plus sémantique	118

Ce chapitre a pour objectif de synthétiser nos résultats par rapport à notre problématique de recherche, puis de poser les questions de recherche qu'il nous semble important de prendre en compte et qui permettent d'identifier des perspectives de recherche dans le domaine de l'observation en EIAH. Enfin, nous présentons les actions de recherche que nous envisageons de mener à moyen terme.

7.1 Synthèse des activités menées

Dans tous les domaines aujourd'hui, on essaye de se rapprocher de l'utilisateur, de lui proposer des services de plus en plus adaptés, voire personnalisés. Les travaux que nous avons menés dans le cadre du Web Adaptatif ont apporté une réflexion, une méthodologie sur le développement d'outils adaptatifs et personnalisables. En particulier, nous avons mis en œuvre les principes du Web Sémantique dans la composition sémantique de documents virtuels personnalisables. Le moteur proposé utilise des ontologies pour construire à la volée des documents selon les caractéristiques des usagers. Nous avons pu mettre en évidence des besoins dans la création et la maintenance des modèles utilisateurs. Cette gestion peut être réalisée directement avec l'utilisateur avec d'une part le risque d'une surcharge cognitive liée à des tâches qui ne relèvent pas de son activité principale, et d'autre part le problème de la fiabilité des informations données par l'utilisateur qui se pose. Une autre solution consiste alors à observer le comportement des usagers afin d'en tirer des conclusions sur ses connaissances et ses compétences. Le deuxième constat concerne l'évaluation des outils et des méthodes mis en œuvre. En effet, il est difficile d'obtenir des retours sur l'utilisation de ces nouveaux outils, soit parce que les outils ne permettent pas d'observer ce qui se passe, soit parce qu'il est nécessaire de se baser sur la perception de l'utilisateur qui n'est pas systématiquement fiable. A ce titre, l'observation du comportement des usagers, de l'acquisition de leurs connaissances et de leur savoir-faire, est devenue un enjeu primordial. Dans le domaine spécifique des EIAH, ce besoin est grandissant, de nombreux projets européens du réseau Kaleidoscope, les différents travaux des équipes nationales et internationales (communauté Educational Data Mining, ateliers traces et EIAH en Rhône-Alpes, école thématique du CNRS, etc.) l'attestent.

Dans le cadre de nos activités au sein du projet REDiM du LIUM, nous nous sommes intéressés à la définition et à l'opérationnalisation du processus d'ingénierie de l'observation pour des EIAH scénarisés. Notre activité se focalisant sur les EIAH scénarisés, nous avons choisi de ne pas aborder le problème de l'observation par l'analyse directe des observés (ou trace brute) qui amène à la découverte d'usages non prévus mais qui font sens à l'utilisateur. Nous avons décidé de prendre en compte le savoir-faire des usagers au travers des besoins d'observation qu'ils expriment. C'est pourquoi, nous qualifions notre approche de prescriptive. De notre point de vue, l'enseignant ou le concepteur qui scénarise sa situation d'apprentissage, peut scénariser aussi l'observation de cette situation et préconiser éventuellement une association entre les deux scénarios.

Les différents résultats que nous proposons actuellement sont tout d'abord une chaîne éditoriale qui décrit notre vision de la démarche d'observation prescriptive en s'appuyant sur trois rôles : le prescripteur, l'analyste et le développeur. Cette chaîne éditoriale intègre l'expression du besoin d'observation et de l'objectif d'observation, la spécification des

indicateurs qui en découlent, mais aussi la capitalisation du savoir-faire afin d'en faciliter le partage et la réutilisation. Il nous a alors semblé nécessaire de concevoir nos modèles de manière à ce qu'ils soient opérationnalisables et automatisables en session.

Nous avons proposé ensuite un modèle de représentation de données d'observation associé à cette chaîne éditoriale, le modèle DGU (Defining-Getting-Using) ainsi qu'un langage formalisant ce modèle de représentation. Le langage UTL (Usage Tracking Language) est un langage instrumenté permettant de spécifier de façon déclarative les indicateurs répondant aux besoins d'observation, ainsi que toutes les données d'observation nécessaires au calcul de ces indicateurs. Le moyen d'obtention des données (par exemple la règle de calcul d'un indicateur) peut aussi être spécifié grâce au langage de combinaison de données DCL4UTL. Ces différents formalismes facilitent la capitalisation et la réutilisation du savoir-faire. Ils ont été conçus de manière à ce que l'expression du savoir-faire en analyse soit indépendant du format de représentation des traces, du langage de scénarisation et du système de stockage (par exemple une base de données XML ou relationnelle) des données d'observation.

Afin d'évaluer le langage UTL, nous l'avons intégré dans l'EIAH Hop3X développé au LIUM [Lekira et al. 2011b]. Cette version d'Hop3X couplée à UTL a été utilisée dans le cadre de plusieurs expérimentations au département informatique de l'Université du Maine ainsi qu'au département SRC de l'IUT de Laval. Toutefois, comme nous l'avons évoqué dans la synthèse du chapitre précédent, cet ensemble d'outils que nous proposons n'est qu'un élément de l'instrumentation de la chaîne éditoriale.

L'enseignant/concepteur est au cœur de notre démarche en tant que prescripteur de l'observation, expert de son scénario d'apprentissage et donc la personne la plus compétente pour spécifier un scénario d'observation. Toutefois, les aspects déclaratifs de notre langage et notamment la déclaration du moyen d'observation subissent la critique d'éloignement de l'enseignant/concepteur. C'est un point sur lequel nous travaillons, tout d'abord avec la spécification des stratégies d'observation, ensuite sur l'évolution de la composition de données vers des langages graphiques de spécification plus facile à appréhender. La thèse de Mohand Akli Ouali porte sur l'étude des stratégies d'observation. La composition de données avec un langage graphique est une perspective à moyen terme. La seconde limite correspond à la validation de la généricité de la partie "Patron" de notre langage. En effet, actuellement les EIAH que nous avons utilisés étaient trop différents pour faciliter la réutilisation. Nous envisageons de collaborer avec d'autres EIAH afin de réutiliser certains indicateurs déjà produits.

7.2 Partager et réutiliser au sein de la communauté EIAH

La question générale sur laquelle repose notre problématique est : **“Comment aider l'enseignant/concepteur à améliorer sa situation d'apprentissage grâce à l'observation de cette dernière ?”**. Afin d'y répondre, nous avons proposé un cadre méthodologique pour l'ingénierie de l'observation par la prescription. Dans ce cadre, l'enseignant/concepteur occupe le rôle de prescripteur afin d'obtenir un processus d'obser-

vation répondant à ses objectifs par rapport à sa situation d'apprentissage. La chaîne éditoriale de l'observation guide nos actions de recherche et nous a amené à identifier une question de recherche pour laquelle, il nous semble nécessaire que la communauté apporte des réponses :

– **Comment partager et réutiliser les modèles outillés de la communauté ?**

Ces dernières années, nous avons constaté une volonté forte de la communauté de chercheurs en EIAH, de capitaliser et de partager sur l'observation. Il suffit de voir tous les travaux menés sur l'analyse de traces, les indicateurs ou les outils de suivi pour comprendre qu'il y a un potentiel sous-utilisé.

Aider l'enseignant à mieux percevoir sa situation passe aussi par une forme de collaboration. La mise en commun des situations observées et des solutions apportées pourrait améliorer l'activité des enseignants et des tuteurs. Ce qui constitue l'un des principaux échecs sur la réutilisation des outils d'analyse aujourd'hui, est l'impossibilité de les réutiliser "facilement". En effet, il serait souhaitable de disposer de services d'analyse adaptables aux différents EIAH existants que l'on pourrait partager entre membres de la communauté que l'on soit chercheur, enseignant ou concepteur de plate-formes.

Actuellement, la communauté dispose de nombreux outils (TraViS, Tatiana, SBT, UTL, etc.), malheureusement sans aucune fédération des projets. Même si ces derniers n'ont pas de liens directs, il pourrait être utile de les recenser dans un premier temps, de les rendre accessibles au travers de services web par exemple. Sébastien George soulève la question de l'interopérabilité entre les systèmes manipulant les traces afin de pouvoir raisonner à des niveaux différents [George 2010]. Il s'agit ici de pouvoir tester un même corpus de traces sur des outils différents. C'est typiquement le but de projets comme MULCE ou DataShop qui sont destinés aux chercheurs uniquement. Plusieurs de ces outils sont complémentaires et pourraient constituer le point de vue national de l'observation en EIAH.

Aujourd'hui, nombreux sont les enseignants qui souhaitent effectivement disposer d'outils d'observation, des tentatives de partage du savoir-faire ont été réalisées dans le cadre du réseau d'excellence Kaleidoscope, mais n'ont pas dépassé le stade de recommandations ou de propositions. C'est le cas de la banque de patrons proposée par le projet DPULS. Nous pensons qu'il est temps aujourd'hui de fédérer cette communauté et de mettre en commun le savoir-faire en observation. Ce savoir-faire est constitué bien sûr des techniques d'analyse, mais aussi de l'utilisation qui est faite de l'analyse. En effet, il est pertinent de disposer des actions de régulation tutorales testées, et évaluées sur certains indicateurs, de disposer aussi des mécanismes de visualisation, d'adaptation et de réingénierie utilisés. La capitalisation pourrait intégrer des discussions de fond sur les choix stratégiques en terme d'analyse afin tout d'abord d'éviter de refaire ce qui existe déjà et d'aider les collègues qui s'initie à l'apprentissage avec un EIAH. Une solution serait, par exemple, de mettre en évidence des risques dès l'étape de scénarisation et d'offrir l'opportunité de corriger ou d'observer grâce aux moyens testés et validés par la communauté.

Le partage des modèles et outils passe aussi par la constitution d'un entrepôt de données d'observation, afin que les chercheurs puissent valider leurs propositions. L'analyse de données d'observation nécessite de pouvoir manipuler des données de provenances diverses et de formats variés. Souvent les chercheurs dans ce domaine sont amenés à élaborer

plusieurs expérimentations afin de disposer de traces suffisantes pour pouvoir concevoir et tester leurs modèles et leurs outils. Il serait intéressant de pouvoir disposer d'entrepôts de données d'observation, c'est-à-dire de traces générées lors d'une session mais aussi de toutes les informations liées à l'EIAH qui pourraient s'avérer utiles comme le scénario, des productions d'apprenants, des réactions tutorales etc.. Les tentatives pour ce type d'approche commencent toujours par la recherche d'un format unique de représentation. Par exemple dans le projet MULCE [MULCE 2009], il est question de créer un corpus d'apprentissage comportant à la fois les données, les productions, le scénario, le contexte de recherche. Cette définition nous semble très intéressante, mais dans le modèle proposé, le choix a été fait d'une part de se limiter aux traces d'interaction et d'autre part de fournir une structure unique afin d'accéder à ces traces au travers d'une plate-forme de partage. De plus, ils considèrent que le corpus d'apprentissage doit inclure le système de fouille de données. Nous pensons qu'il est préférable de conserver les données dans leur format originel afin de permettre à d'autres outils de les manipuler et aussi que les outils puissent être facilement réutilisables directement sur les environnements d'apprentissage. Néanmoins, il est nécessaire de prévoir un mécanisme d'anonymisation des données en se basant, par exemple, sur des profils d'apprenant suffisamment précis mais sans possibilités d'identifier l'apprenant. De plus, les travaux du projet MULCE apportent déjà des solutions pour la notion de droit et de licence d'utilisation.

Dans le cadre de nos activités de recherche, nous envisageons d'étudier au concept de *Patron d'observation* afin de faciliter le partage et la réutilisation des indicateurs. Le patron d'observation comporte non seulement l'indicateur mais aussi les solutions de correction/amélioration/régulation déployées par le concepteur. De façon à créer une communauté d'utilisateurs sur les indicateurs, nous envisageons de coupler le patron avec des évaluations, des critiques sur l'effectivité de l'indicateur et de ses solutions. Cette perspective entre directement dans notre cinquième question de recherche sur la communauté de pratiques exprimée dans la section 1.1. Elle constitue aussi l'objectif du projet OPALE que nous envisageons de déposer à nouveau avec l'IRIT et le LIRIS, dans ce projet nous allons prendre comme base de travail UTL et SBT, ce qui sera un premier rapprochement entre ces deux approches.

7.3 Perspectives à moyen terme

Il s'agit de présenter ici un ensemble de perspectives qui nous envisageons de développer ou que nous avons déjà commencé à prendre en compte dans nos activités. Maintenant que nous disposons d'un environnement d'analyse en partie instrumenté, nous souhaitons développer d'une part, la prescription en étant au plus près des connaissances métier du prescripteur, et d'autre part, l'exploitation des données d'observation. Ces perspectives sont des pistes de réponses aux questions de recherche que nous avons énoncées dans le premier chapitre. En effet, l'amélioration de la compréhension de la situation d'apprentissage passe par une meilleure prise en compte du modèle métier de l'enseignant/concepteur, sachant que celui-ci nécessite de s'intéresser aux spécificités de chacun. Les destinataires de l'observation méritent aussi d'être étudiés de plus près. Il est donc nécessaire d'étudier comment faciliter la perception de l'observation, le contexte est à prendre en compte,

mais il faut surtout éviter la surcharge d'information qui ne peut que nuire au bon déroulement d'une session d'apprentissage. Enfin, notre position quant à l'implication de l'enseignant/concepteur dans le processus d'ingénierie nous amène à étudier des moyens d'élaboration de données d'observation qui permettraient de s'abstraire de l'analyste. Pour cela, nous envisageons d'exploiter la sémantique des données directement dans la spécification des moyens d'observation.

7.3.1 L'expression du besoin d'observation dans l'univers métier du prescripteur

Nous envisageons de revenir sur l'activité du concepteur. Tout d'abord, il est important de faire le lien entre l'expression du besoin d'observation et la formalisation de l'indicateur. Les travaux de Zendagui [Zendagui 2010] en constituent un début prometteur. En effet, il propose un modèle conceptuel des besoins d'observation et un processus d'aide à la spécification de ces besoins. Le modèle conceptuel a pour but principal de structurer les différentes informations nécessaires pour préparer une activité d'observation. Le processus d'aide qui est proposé accompagne l'enseignant-concepteur dans les différentes étapes de la spécification de ses besoins d'observation. Dans ces travaux, un éditeur graphique supportant le processus et le modèle proposés a été développé. Il a proposé une amorce de transformation vers le langage UTL qui mérite d'être étoffée car elle ne prend en compte que la spécification au niveau "Définition" des données d'observation.

7.3.2 Le retour d'information ou comment percevoir les indicateurs ?

Dans la version actuelle de nos résultats, nous n'avons pas encore instrumenté l'exploitation de l'indicateur. En effet, nous mémorisons la ou les valeurs obtenues, mais il est nécessaire désormais de présenter cette information aux usagers. La thèse qui vient de démarrer porte sur la notion de stratégie d'observation : ce qu'il faut observer, comment et quand, comment instrumenter le destinataire de l'observation (notion de tableau de bord d'observation), etc.. La stratégie d'observation est propre au destinataire des indicateurs (dans le cadre des EIAH : l'enseignant, l'apprenant, l'administrateur de la formation, le concepteur), elle est constituée d'un ensemble d'indicateurs avec leurs interfaces de perception, d'un ensemble de règles de composition (d'assemblage) ainsi que du contexte d'utilisation de cette stratégie. La prise en compte du contexte de spécification et d'utilisation des indicateurs au sein même de leur définition devient nécessaire pour en faciliter la réutilisation. La stratégie d'observation doit être capitalisable et extensible pour pouvoir être réutilisée. Une stratégie extensible doit permettre, lors de sa réutilisation, l'ajout d'un ou plusieurs indicateurs dans un contexte donné.

7.3.3 La définition de nouveaux indicateurs en session

Nous avons considéré jusqu'à présent la prescription des indicateurs avant la session d'apprentissage car il était nécessaire de faire intervenir un analyste sur la spécification

des données. Or, il serait intéressant de permettre aux usagers (enseignant, tuteur et éventuellement étudiants) de combiner des données et de définir de nouveaux indicateurs directement en session. Il est question ici de reprendre les travaux de Noa Randriamalaka sur l'élicitation [Randriamalaka et al. 2007] et de proposer un langage graphique de composition de données d'observation avec évaluation en temps réel.

7.3.4 Vers une composition encore plus sémantique

Actuellement la sémantique est amenée au travers des concepts traçables associés aux données et qui sont établis en fonction du scénario, ainsi que par les différents champs de description associés aux données. Les relations entre les données, notamment lors de la composition, sont figées. En effet, pour spécifier la section "Obtention" d'un indicateur, nous indiquons directement le nom et le type des données nécessaires. L'approche que nous envisageons est similaire à celle du moteur SCARCE. L'idée repose sur l'utilisation d'ontologies qui modéliseraient l'univers métier ainsi que l'univers des plate-formes (ou plus généralement celui des environnements), puis de nous baser sur ces ontologies pour décrire la sémantique des données. Cela permettrait d'indiquer la sémantique de la donnée nécessaire au calcul plutôt que son nom, et rendrait les spécifications de données encore plus génériques. Nous pouvons avoir plusieurs données brutes avec une sémantique identique mais des noms différents car elles ont été conçues pour des projets spécifiques. L'utilisation d'une recherche sur la sémantique des données permettrait de faciliter la réutilisation. En effet, en partant du principe de SCARCE, l'analyste ou directement le prescripteur pourrait regrouper un ensemble de données existantes cohérentes par rapport à son scénario et à son EIAH et l'interpréteur fonctionnerait directement sur ces données. Par exemple, pour un indicateur mesurant le taux d'interaction entre les apprenants, il suffirait de préciser dans les données nécessaires au calcul qu'il faut des données sur "l'utilisation d'un moyen d'interaction" sans préciser lequel, puis dans les données brutes proposer une ou plusieurs données étiquetées "utilisation" et "Moyen d'interaction" pour qu'elles soient prises en compte.

Bibliographie

- [Adam et al. 2005] Adam J.-M., Bessagnet M., Bouzeghoub A., Caron P.-A., Carron T., Choquet C., David B., Derycke A., George S., Iksal S., Labat J.-M., Laforcade P., Le Pallec X., Lecocq C., Luengo V., Nodenot T., Oubahssi L., Peter Y., Rebai I., Ros-selle M. et Vantroys T. (2005). Contributions de l'Action Spécifique «Conception d'une Plate-forme pour la recherche en EIAH» à l'ingénierie des Environnements Informatiques pour l'Apprentissage Humain.
- [Barré et al. 2004] Barré V., Choquet C., Iksal S. et Corbière A. (2004). Usage analysis in an e-learning system : LD representation significance. Dans *IEEE International Conference on Advanced Learning Technologies (ICALT'2004)*, pages 570–574.
- [Bousbia 2011] Bousbia N. (2011). Traces de navigation des apprenants dans un environnement de formation sur le web. Dans *Atelier "Partager des données d'observation pour la recherche en EIAH - traces d'activité d'apprentissage", EIAH'2011*.
- [Bousbia et Labat 2007] Bousbia N. et Labat J.-M. (2007). Perception de l'activité de l'apprenant dans un environnement de formation. Dans *Actes de la conférence Environnements Informatiques pour l'Apprentissage Humain*, pages 233–238. INRP.
- [Bousbia et al. 2010] Bousbia N., Rebai I., Labat J.-M. et Balla A. (2010). Learners' navigation behavior identification based on traces analysis. Dans *User Modeling and User-Adapted Interaction : The Journal of Personalization Research*, volume 20, pages 455–494.
- [Brusilovsky 1996] Brusilovsky P. (1996). Methods and techniques of adaptive hypermedia. *User Modeling and User Adapted Interaction*, 6(2-3) : 87–129.
- [CAViCoLA 2006] CAViCoLA (2006). Computer-based analysis and visualization of collaborative learning activities. consulté le 01/09/07.
- [Choquet 2007] Choquet C. (2007). *Ingénierie et réingénierie des EIAH - L'approche REDiM*. Thèse d'habilitation à diriger des recherches, Université du Maine.
- [Choquet et Iksal 2007a] Choquet C. et Iksal S. (2007a). Modeling tracks for the model driven reengineering of a TEL system. *Journal of Interactive Learning Research (JILR), Special Issue Usage Analysis in Learning Systems : Existing Approaches and Scientific Issues*, 18(2) : 161–184.
- [Choquet et Iksal 2007b] Choquet C. et Iksal S. (2007b). Modélisation et construction de traces d'utilisation d'une activité d'apprentissage : une approche langage pour la réingénierie d'un EIAH. *Sciences et Technologies de l'Information et de la Communication pour l'Education et la Formation (STICEF)*, 14.

- [Corbière 2006] Corbière A. (2006). *Analyses des apports du méta-standard, ODP-RM à la communauté EIAH, Instances sur un système de formation*. Thèse de Doctorat, Université du Maine.
- [Corbière et Choquet 2005] Corbière A. et Choquet C. (2005). ODP-RM : Un cadre de réingénierie des systèmes de formation. Dans *Sciences et Technologies de l'Information et de la Communication pour l'Éducation et la Formation*, volume 12.
- [David 2005] David J.-P. (2005). D32.3.1 (final) – state of art of tracking and analysing usage. Report, Kaleidoscope - JEIRP DPULS.
- [Dean et al. 2002] Dean M., Connolly D., van Harmelen F., Hendler J., Horrocks I., McGuinness D., Patel-Schneider P. et Stein L. (2002). OWL web ontology language 1.0 reference. Working draft, W3C, Url : <http://www.w3.org/TR/2002/WD-owl-ref-20020729/>.
- [Dimitracopoulou et al. 2004] Dimitracopoulou A. et al. (2004). D26.1.1. : State of the art on interaction and collaboration analysis. Technical report, Kaleidoscope Deliverables, accès (restreint). consulté le 01/09/07.
- [Djouad et al. 2009] Djouad T., Mille A., Reffay C. et Benmohammed M. (2009). Ingénierie des indicateurs d'activités à partir de traces modélisées pour un environnement informatique d'apprentissage humain. Dans *Sciences et Techniques de l'Information et de la Communication pour l'Éducation et la Formation*, volume 16.
- [Djouad et al. 2010] Djouad T., Settouti L., Prié Y., Reffay C. et Mille A. (2010). Un système à base de traces pour la modélisation et l'élaboration d'indicateurs d'activités éducatives individuelles et collectives. mise à l'œuvre sur moodle. Dans *TSI*, volume 29, pages 721–741.
- [DPULS 2005a] DPULS (2005a). Design patterns for recording and analysing usage of learning systems. Technical report, Kaleidoscope NoE.
- [DPULS 2005b] DPULS (2005b). DPULS design patterns browser, kaleidoscope project. consulté le 5 avril 2011.
- [Dyke 2009] Dyke G. (2009). *Un modèle pour la gestion et la capitalisation d'analyses de traces d'activités en interaction collaborative*. Thèse de Doctorat, Ecole Nationale Supérieure des Mines de Saint-Etienne.
- [France et al. 2007] France L., Heraud J.-M., Marty J.-C. et Carron T. (2007). Visualisation et régulation de l'activité des apprenants dans un EIAH tracé. Dans *Conférence EIAH 2007 (Environnements Informatiques pour l'Apprentissage Humain)*, pages 179–184.
- [Frias-Martinez et al. 2006] Frias-Martinez E., Chen S. Y. et Xiaohui L. (2006). Survey of data mining approaches to user modeling for adaptive hypermedia. *IEEE Transactions on Systems, Man, and Cybernetics, Part C : Applications and Reviews*, 36(6) : 734–749.
- [Garlatti et Iksal 2000] Garlatti S. et Iksal S. (2000). Concept filtering and spatial filtering in an adaptive information system. Dans Brusilovsky P., Stock O. et Strapparava C., éditeurs, *Adaptive Hypermedia and Adaptive Web-Based Systems*, pages 315–318. AH 2000, Springer Verlag (LNCS 1892).

- [Garlatti et al. 2004] Garlatti S., Iksal S. et Tanguy P. (2004). SCARCE : An adaptive hypermedia environment based on virtual documents and the semantic web. Dans Chen S. Y. et Magoulas G. D., éditeurs, *Adaptable and Adaptive Hypermedia Systems*, pages 206–224. Idea Group Inc.
- [Gaudio et Talavera 2006] Gaudio E. et Talavera L. (2006). Data mining to support tutoring in virtual learning communities : Experiences and challenges. Dans Romero C. et Ventura S., éditeurs, *Data mining in e-learning*, pages 207–226. WIT Press.
- [Gendron et al. 2012] Gendron E., Pourroy F., Carron T. et Marty J.-C. (2012). Towards a structured approach to the definition of indicators for collaborative activities in engineering design. Dans *Journal of Engineering Design*. Taylor & Francis.
- [George 2010] George S. (2010). *Interactions et communications contextuelles dans les environnements informatiques pour l'apprentissage humain*. Thèse d'habilitation à diriger des recherches, INSA de Lyon / Université Claude Bernard (Lyon 1).
- [Guéraud et al. 2004] Guéraud V., Adam J.-M., Pernin J.-P., Calvary G. et David J.-P. (2004). L'exploitation d'objets pédagogiques interactifs à distance : le projet FORMID. Dans *Revue des Sciences et Technologies de l'Information et de la Communication pour l'Éducation et la Formation (STICEF)*, volume 11.
- [Hall et al. 2009] Hall M., Frank E., Holmes G., Pfahringer B., Reutemann P. et Witten I. H. (2009). The WEKA data mining software : an update. *SIGKDD Explor. Newsl.*, 11(1) : 10–18.
- [Hand et al. 2001] Hand D., Mannila H. et Smyth P. (2001). *Principles of data mining. Adaptive computation and machine learning*. MIT Press.
- [Heraud et al. 2005] Heraud J.-M., Marty J.-C., France L. et Carron T. (2005). Une aide à l'interprétation de traces : application à l'amélioration de scénarios pédagogiques. Dans *Conférence EIAH 2005 (Environnements Informatiques pour l'Apprentissage Humain)*, pages 237–248.
- [IA 2005] IA (2005). Interaction analysis. Technical report.
- [ICALTS 2004] ICALTS (2004). Interaction & collaboration analysis' supporting teachers & students' self-regulation. Technical report.
- [Iksal 2011a] Iksal S. (2011a). A declarative and operationalized language for learning systems analysis : UTL. Dans *I-Society 2011*, pages 62–66, London.
- [Iksal 2011b] Iksal S. (2011b). Tracks analysis in learning systems : A prescriptive approach. Dans Akmayeva G., Shoniregun C. et Kinshuk, éditeurs, *International Journal for e-Learning Security*, volume 1, pages 3–9. Infonomics Society.
- [Iksal et al. 2004] Iksal S., Barré V., Choquet C. et Corbière A. (2004). Comparing prescribed and observed for the re-engineering of elearning systems. Dans *IEEE Sixth International Symposium on Multimedia Software Engineering (MSE)*, pages 106–109.
- [Iksal et al. 2007] Iksal S., Carron T. et Marty J.-C. (2007). Modéliser les expérimentations basées sur les traces : Une étude de cas avec le langage utl. Technical report, LIUM, SYSCOM.

- [Iksal et Choquet 2005a] Iksal S. et Choquet C. (2005a). An open architecture for usage analysis in a e-learning context. Dans *ICALT 2005*, pages 177–181.
- [Iksal et Choquet 2005b] Iksal S. et Choquet C. (2005b). Usage analysis driven by models in a pedagogical context. Dans *Workshop UALS in AIED 2005*, pages 49–56.
- [Iksal et Garlatti 2002] Iksal S. et Garlatti S. (2002). Adaptive special reports for on-line newspapers. Dans Mizzaro S. et Tasso C., éditeurs, *Proceedings of the AH'2002 Workshop on Personalization Techniques in Electronic Publishing*, pages 31–43. Universidad de Malaga.
- [IMS-LD 2003] IMS-LD (2003). *IMS Learning Design*. IMS Global. consulté le 01/09/07.
- [Jermann et al. 2001] Jermann P., Soller A. et Mühlenbrock M. (2001). From mirroring to guiding : A review of the state of art technology for supporting collaborative learning. Dans *European Perspectives on Computer-Supported Collaborative Learning*, pages 324–331.
- [KALEIDOSCOPE 2004] KALEIDOSCOPE (2004). Kaleidoscope project. consulté le 01/09/07.
- [Klinger 2006] Klinger E. (2006). *Apports de la réalité virtuelle à la prise en charge de troubles cognitifs et comportementaux*. Thèse de Doctorat, Enst Paris.
- [Koedinger et al. 2010] Koedinger K., Baker R. S., Cunningham K., Skogsholm A., Leber B. et Stamper J. (2010). *A Data Repository for the EDM community : The PSLC DataShop*, chapitre 4, pages 43–56. Data Mining and Knowledge Discovery. Romero, C., Ventura, S., Pechenizkiy, M., Baker, R.S.J.d.
- [Lassila et Swick 1999] Lassila O. et Swick R. (1999). Resource description framework (RDF) model and syntax specification. Technical report, W3C. Url : <http://www.w3.org/TR/1999/REC-rdf-syntax-19990222>.
- [Le Robert 2011] Le Robert C. (2011). *Le Petit Robert : Dictionnaire alphabétique et analogique de la langue française*. Le Robert.
- [Lekira et al. 2011a] Lekira A., Després C. et Jacoboni P. (2011a). An indicator-based approach to support teachers' activities. Dans *The 11th IEEE International Conference on Advanced Learning Technologies (ICALT'11)*, Athens - Georgia (USA).
- [Lekira et al. 2011b] Lekira A., Després C., Jacoboni P., Choquet C., Iksal S., Py D. et Pham Thi Ngoc D. (2011b). Using indicators during synchronous tutoring of practical work. Dans *The 11th IEEE International Conference on Advanced Learning Technologies (ICALT'11)*, Athens - Georgia (USA).
- [Lund et Mille 2009] Lund K. et Mille A. (2009). Traces, traces d'interactions, traces d'apprentissages définitions, modèles informatiques, structurations, traitements et usages. Dans Marty J.-C. et Mille A., éditeurs, *Analyse de traces et Personnalisation des EIAH*, Traité Informatique et Systèmes d'Information, chapitre 1, pages 21–56. Lavoisier-Hermès.
- [Martel et al. 2006] Martel C., Vignollet L., Ferraris C. et Durand G. (2006). LDL : a language to model collaborative learning activities. Dans Pearson E. et Bohman P., éditeurs, *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2006*, pages 838–844. ED-MEDIA 2006.

- [May 2009] May M. (2009). *Using Tracking Data as Reflexive Tools to Support Tutors and Learners in Distance Learning Situations : an Application to Computer-Mediated Communications*. Thèse de Doctorat, INSA de Lyon.
- [May et al. 2008] May M., George S. et Prévôt P. (2008). Tracer, analyser, et visualiser les interactions des apprenants pendant les activités de communications médiatisées. Dans *Actes du colloque JOCAIR : Journées Communication et Apprentissage Instrumentés en Réseau*, pages 251–263.
- [Mazza et Milani 2004] Mazza R. et Milani C. (2004). GISMO : a graphical inter-active student monitoring tool for course management systems. Dans *Technology Enhanced Learning '04 International Conference*.
- [Merceron et Yacef 2003] Merceron A. et Yacef K. (2003). A web-based tutoring tool with mining facilities to improve learning and teaching. Dans *Proceedings of 11th International Conference on Artificial Intelligence in Education*. F. Verdejo and U. Hoppe.
- [Merceron et Yacef 2005] Merceron A. et Yacef K. (2005). TADA-Ed for educational data mining. Dans *Interactive Multimedia Electronic Journal of Computer-Enhanced Learning*, volume 7.
- [MOODLE 2012] MOODLE (2012). Moodle : Open source course management system. <http://moodle.org/>.
- [Mostow et Beck 2006a] Mostow J. et Beck J. (2006a). Some useful tactics to modify, map and mine data from intelligent tutors. Dans *Special Issue on Educational Applications of the Journal Natural Language Engineering*, volume 12, pages 195–208.
- [Mostow et Beck 2006b] Mostow J. et Beck J. (2006b). Some useful tactics to modify, map and mine data from intelligent tutors. *Nat. Lang. Eng.*, 12(2) : 195–208.
- [MULCE 2009] MULCE P. A. (2009). Notion de corpus d'apprentissage (LETEC). Technical report, ANR.
- [Ouali 2012] Ouali M.-A. (2012). Modélisation outillée d'une stratégie contextualisée d'observation dans le cadre d'un environnement d'apprentissage. Dans Carron T. et Joiron C., éditeurs, *Actes des quatrièmes rencontres des Jeunes Chercheurs en EIAH*, pages 87–92.
- [Peña de Carrillo et al. 2011] Peña de Carrillo C., Choquet C., Després C., Iksal S., Jacoboni P., Lekira A., Ouraiba E. et Pham Thi Ngoc D. (2011). Engineering and reengineering of technology enhanced learning scenarios using context awareness processes. Dans Graf S., Lin F., Kinshuk et McGreal R., éditeurs, *Intelligent and Adaptive Learning Systems : Technology Enhanced Support for Learners and Teachers*, pages 34–58. IGI Global.
- [Phâm Thi Ngoc 2011] Phâm Thi Ngoc D. (2011). *Spécification et conception de services d'analyse de l'utilisation d'un environnement informatique pour l'apprentissage humain*. Thèse de Doctorat, Université du Maine.
- [Phâm Thi Ngoc et al. 2010a] Phâm Thi Ngoc D., Iksal S. et Choquet C. (2010a). Learning tracks' analysis with DCL4UTL : An instantiation with the division of labor indicator. Dans *Proceedings of Global Learn Asia Pacific 2010*, pages 3334–3343. AACE.

- [Phâm Thi Ngoc et al. 2010b] Phâm Thi Ngoc D., Iksal S. et Choquet C. (2010b). Re-engineering of pedagogical scenarios using the data combination language and usage tracking language. Dans *The 10th IEEE International Conference on Advanced Learning Technologies (ICALT' 2010)*. IEEE Computer Society Publications.
- [Phâm Thi Ngoc et al. 2009] Phâm Thi Ngoc D., Iksal S., Choquet C. et Klinger E. (2009). Utl-cl : A declarative calculation language proposal for a learning tracks analysis process. Dans *The 9th IEEE International Conference on Advanced Learning Technologies (ICALT' 2009)*, pages 681–685.
- [Randriamalaka 2008] Randriamalaka N. (2008). *Assistance à la modélisation et à l'analyse des traces d'usage dans un EIAH : une approche par la réingénierie*. Thèse de Doctorat, Université du Maine.
- [Randriamalaka et Iksal 2006] Randriamalaka N. et Iksal S. (2006). Patterns approach in the re-engineering process of learning scenario. Dans *The 6th IEEE International Conference on Advanced Learning Technologies (ICALT'2006)*, pages 799–803.
- [Randriamalaka et al. 2007] Randriamalaka N., Iksal S. et Choquet C. (2007). Indicators' elicitation process for re-engineering of learning scenario : Tracks approach based on usage tracking language. Dans *The 7th IEEE International Conference on Advanced Learning Technologies (ICALT'07)*, pages 492–496.
- [Rasmussen et al. 2011] Rasmussen M.-H., Mostow J., Tan Z.-H., Lindberg B. et Li Y. (2011). Evaluating tracking accuracy of an automatic reading tutor. Dans *SLaTE : ISCA (International Speech Communication Association) Special Interest Group (SIG) Workshop on Speech and Language Technology in Education*.
- [Reffay et al. 2008] Reffay C., Chanier T., Noras M. et Betbeder M.-L. (2008). Contribution à la structuration de corpus d'apprentissage pour un meilleur partage en recherche. Dans *Revue Sciences et Technologies de l'Information et de la Communication pour l'Education, Numéro spécial EPAL*, volume 15, pages 185–219. STICEF.
- [Romero et al. 2010] Romero C., Ventura S., Pechenizkiy M. et Baker R. S., éditeurs (2010). *Handbook of Educational Data Mining*. Data Mining and Knowledge Discovery. CRC Press.
- [Settouti et al. 2011] Settouti L., Guin N., Luengo V. et Mille A. (2011). Adaptable and reusable query patterns for trace-based learner modelling. Dans *EC-TEL, Sixth European Conference on Technology Enhanced Learning : Towards Ubiquitous Learning.*, pages 384–397.
- [Settouti et al. 2010] Settouti L., Guin N., Mille A. et Luengo V. (2010). A trace-based learner modelling framework for technology-enhanced learning systems. Dans *The 10th IEEE International Conference on Advanced Learning Technologies*, pages 73–77.
- [Settouti 2011] Settouti L. S. (2011). *Systèmes à Base de traces modélisées - Modèles et langages pour l'exploitation des traces d'Interactions*. Thèse de Doctorat, Université Claude Bernard Lyon 1.
- [Settouti et al. 2007] Settouti L. S., Prié Y., Marty J.-C. et Mille A. (2007). Vers des systèmes à base de traces modélisées pour les EIAH. Technical Report Rapport de recherche RR-LIRIS-2007-016, LIRIS.

- [Stamper et al. 2011] Stamper J., Koedinger K., Baker R. S., Skogsholm A., Leber B., Demi S., Yu S. et Spencer D. (2011). Managing the educational dataset lifecycle with datashop. Dans *Proceedings of the 15th International Conference on Artificial Intelligence in Education*, pages 557–559. Kay, J., Bull, S. and Biswas, G.
- [Tchounikine 2004] Tchounikine P. (2004). Platon-1 : quelques dimensions pour l'analyse des travaux de recherche en conception d'eiah. rapport de l'action spécifique "fondements théoriques et méthodologiques de la conception des eiah". Technical report, Département STIC du CNRS.
- [TRAILS 2004] TRAILS (2004). Personalised and collaborative trails of digital and non-digital learning objects. Technical report.
- [vom Brocke 2001] vom Brocke J. (2001). Freestyle learning - concept, platforms, and applications for individual learning scenarios. Dans *46th International Scientific Colloquium*.
- [Zendagui 2010] Zendagui B. (2010). *Support à la spécification des besoins d'observation dans un contexte de ré-ingénierie des scénarios pédagogiques : une approche dirigée par les modèles*. Thèse de Doctorat, Université du Maine.