

# Modélisation et construction de traces d'utilisation d'une activité d'apprentissage : une approche langage pour la réingénierie d'un EIAH

Christophe CHOQUET, Sébastien IKSAL [LIUM, Laval]

■ **RÉSUMÉ** : Dans le contexte de l'enseignement et l'apprentissage à distance, la réingénierie d'un dispositif d'apprentissage suppose de disposer d'informations sur ses usages. Ces informations proviennent de sources multiples, comme des interviews, des questionnaires, des vidéos, des fichiers de logs. Nous considérons qu'il est important d'interpréter ces traces afin de confronter les intentions du concepteur avec les activités de l'apprenant durant la session. Dans cet article, nous présentons le langage UTL (Usage Tracking Language). Ce langage est générique, et nous présentons ici une instanciation de sa première version sur IMS Learning Design, le langage de modélisation pédagogique que nous avons choisi durant nos trois années d'expérimentation. Pour conclure, nous présentons plusieurs cas d'utilisation de ce langage, en nous inspirant de nos expérimentations.

■ **MOTS CLÉS** : Analyse de traces, réingénierie, langage de modélisation, scénarisation pédagogique.

■ **ABSTRACT** : In the context of distance learning and teaching, the re-engineering process needs a feedback on the learners' usage of the learning system. The feedback is given by numerous vectors, such as interviews, questionnaires, videos or log files. We consider that it is important to interpret tracks in order to compare the designer's intentions with the learners' activities during a session. In this paper, we present the usage tracking language – UTL. This language was designed to be generic and we present an instantiation of a part of it with IMS-Learning Design, the representation model we chose for our three years of experiments. At the end of the paper, we describe several use cases of this language, based on our experimentations.

■ **KEYWORDS** : Tracks analysis, reengineering, modeling language, instructional design.

- [1. Introduction](#)
- [2. Traces et analyse d'usage des EIAH](#)
- [3. Une première proposition : décrire la trace collectée](#)
- [4. De la trace à l'indicateur \(UTL 2\)](#)
- [5. Scénarios d'utilisation](#)
- [6. Perspectives](#)
- [Bibliographie](#)

## 1. Introduction<sup>1</sup>

De nombreux systèmes interactifs sont maintenant disponibles sur le *Web* et la plupart de ces systèmes utilisent des informations collectées sur leur usage pour améliorer leur qualité. Dans le contexte spécifique de l'enseignement et de l'apprentissage à distance, les deux fonctions majeures de l'enseignant – concevoir un cours et assurer le tutorat à destination des apprenants – sont souvent dissociées, et même assumées par des acteurs différents, ce qui entraîne généralement un manque de retours d'usage de l'EIAH à destination du concepteur. Selon nous, le processus de production d'un EIAH doit explicitement intégrer une phase d'analyse des usages, destinée à informer les concepteurs sur la qualité de la situation pédagogique mise en œuvre et donc, les aider à décider de l'opportunité de la réingénierie du dispositif (Corbière et Choquet, 2004). Les outils d'analyse automatique de l'usage d'une application logicielle sont souvent définis par des statisticiens et des informaticiens. Afin de faciliter l'appropriation, la compréhension et l'interprétation des résultats, nous pensons que les concepteurs enseignants qui sont les acteurs principaux du processus de développement d'un EIAH et les mieux à même d'interpréter les usages d'un point de vue pédagogique, devraient être étroitement associés à cette phase d'analyse, qu'elle soit automatique ou non.

La contribution scientifique que nous présentons dans cet article est dans le droit fil de l'approche de l'ingénierie et de la réingénierie des EIAH que nous développons dans le projet REDiM (Réingénierie des EIAH Dirigée par les Modèles). Dans cette approche, nous insistons particulièrement sur la nécessité de disposer d'une description formelle de la vue du concepteur sur la situation pédagogique, en termes de scénarios et de ressources pédagogiques. Cette description, que nous appellerons par la suite le scénario prédictif (Lejeune et Pernin, 2004), peut alors être comparée aux usages observés, que nous appellerons par la suite les scénarios descriptifs, de manière à améliorer la qualité du dispositif d'apprentissage par la réingénierie pédagogique. Quand les concepteurs utilisent un langage de modélisation pédagogique (*Educational Modeling Language*, EML) comme par exemple, *IMS-Learning Design* (Koper et al., 2003) proposé par le consortium *IMS Global Learning Consortium* (IMS, 2007), pour traduire leur intention de conception dans un modèle susceptible de refléter la situation qu'ils veulent mettre en place, les activités des acteurs d'une session d'apprentissage sont décrites par un scénario pédagogique prédictif qui, implicitement, définit un ensemble de besoins d'observation. Ainsi, l'une des difficultés de l'observation et de l'analyse des usages d'un EIAH réside dans la corrélation de ces besoins aux moyens techniques d'observation qui doivent être mis en œuvre dans le dispositif d'apprentissage (pas seulement l'EIAH, l'artefact informatique, mais également toute l'organisation, en incluant notamment les acteurs humains et différents types de vecteurs de collecte de données, par exemple des enregistrements vidéos, des questionnaires, etc.).

Nous avons fait l'hypothèse qu'une approche d'Ingénierie Dirigée par les Modèles était pertinente pour impliquer le concepteur dans la modélisation (dimension prescriptive) et dans l'analyse (dimension descriptive) de l'observation de l'utilisation d'un EIAH. Nous considérons donc la modélisation de l'observation comme une succession de transformations entre "ce qu'il est important d'observer" – le modèle métier de l'observation – et "ce qu'il faut collecter" – le modèle de l'observation spécifique au dispositif d'apprentissage. Par analogie, nous considérons l'analyse de l'observation comme une succession de transformations entre "ce qui a été collecté" – la représentation des traces spécifiques au dispositif d'apprentissage – et "ce qu'il est important de percevoir" – la représentation métier des traces.

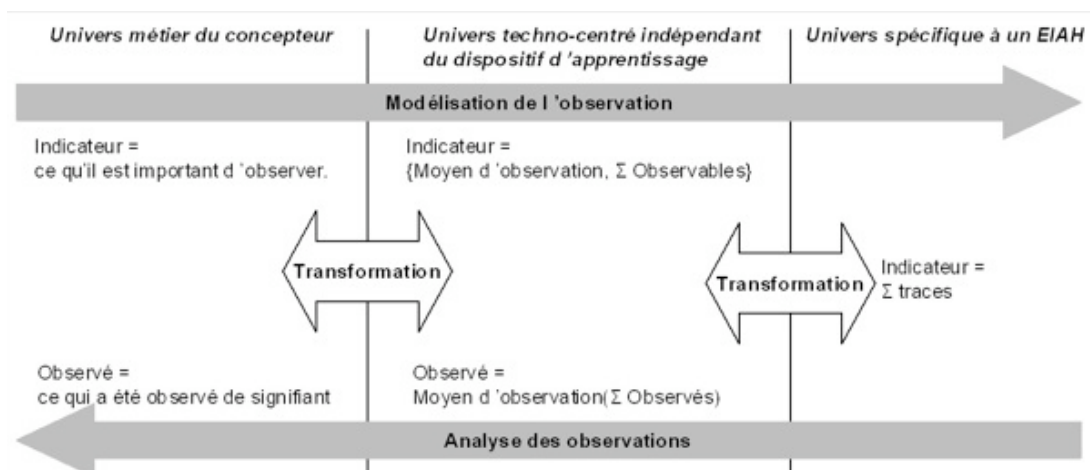


Figure 1 : Approche IDM de la modélisation et de l'analyse de l'observation de l'utilisation d'un EIAH.

La présence des transformations sur la figure 1 précise la nature des interactions entre les rôles typiquement impliqués dans la modélisation et l'analyse de l'observation de l'utilisation d'un EIAH (concepteur, analyste, développeur) :

- en tant que prescripteur, le concepteur définit les indicateurs (ICALTS, 2004) observables qui lui semblent nécessaires pour évaluer la qualité du scénario pédagogique ;
- en tant qu'utilisateur, le concepteur interprète les résultats d'analyse, de manière à décider de la nécessité d'une réingénierie du scénario ;
- le développeur et l'analyste comprennent la nature des indicateurs et négocient avec le concepteur les moyens d'observation et les observables associés ;
- une fois ceux-ci décrits, analyste et développeur opérationnalisent les moyens d'observation et les observables. Ceci les amène à déployer des moyens de collecte spécifiques au dispositif d'apprentissage.

Nous proposons dans cet article le langage UTL pour "*Usage Tracking Language*", dont l'objectif est de permettre aux acteurs intervenant dans le cycle de vie d'un EIAH de décrire les traces d'usage et leur sémantique, ainsi que la définition des besoins d'observation et des moyens à mettre en œuvre pour l'acquisition des données à collecter. Ce langage veut être indépendant tant des technologies utilisées par l'EIAH que du langage de modélisation pédagogique utilisé pour scénariser la situation pédagogique. Ainsi pour un EIAH donné, ce langage doit être instancié sur le langage de modélisation pédagogique et sur les différents formats des traces collectées. UTL permet la structuration des données observées, depuis les données brutes, collectées par le dispositif d'apprentissage pendant une session, jusqu'aux indicateurs établis et/ou calculés à l'aide des données observées, en vue d'analyse la qualité de l'interaction, la nature de l'activité ou l'effectivité de l'apprentissage.

Nous avons actuellement opérationnalisé une première version du langage UTL (Iksal et Choquet, 2005) qui se focalise uniquement sur la transformation des traces par ajout de sémantique. Tout système nécessitant d'analyser le comportement de l'utilisateur utilise soit des techniques automatiques et issues du "*data-mining*" (Mostow, 2004), soit une expertise humaine. Ces méthodes sont couramment utilisées pour construire des représentations de l'usager, des profils, et/ou pour adapter le contenu ou l'interface à l'utilisateur du système (Zheng et al., 2002). Elles sont fondées sur une analyse mathématique et statistique (Bazsalisca et Naim, 2001) et, bien souvent, ont la particularité de travailler les données brutes de manière à en abstraire des indicateurs de plus haut niveau, par exemple pour identifier un mot-clé fréquemment utilisé dans un texte, ou une séquence d'interactions récurrente. Notre objectif est d'aider à la réingénierie d'un EIAH par l'analyse du comportement de l'usager, par exemple l'apprenant ou le tuteur, afin d'améliorer le scénario et les ressources pédagogiques déployés par le dispositif d'apprentissage. Notre proposition consiste à diriger cette analyse par les modèles : l'analyse des données est alors guidée par les besoins d'observation, eux-mêmes identifiés à partir d'un objectif pédagogique.

La section 2 de cet article présente différents points de vue sur les traces et l'analyse d'usage des EIAH, au travers notamment des projets européens du réseau d'excellence Kaleidoscope.

Puis, dans la section 3, nous présentons le langage UTL, tout d'abord la première version qui est opérationnalisée. Cette version permet de décrire la sémantique des traces collectées par une plate-forme d'enseignement à distance et de corréler ces traces à des besoins d'observation définis par le scénario pédagogique décrivant la situation pédagogique déployée par la plate-forme. Ce langage peut être instancié à la fois sur l'EML utilisé pour décrire le scénario pédagogique et sur le format physique de représentation des traces adopté par la plate-forme.

Dans la section 4, nous détaillons une seconde version du langage dont l'objectif est la capitalisation et la réutilisation du savoir-faire en analyse d'usage des EIAH. Il s'agit d'une version enrichie notamment grâce aux différents projets présentés dans la section 2.

Dans une cinquième section, nous présentons quelques cas d'utilisation qui éclairent sur les possibilités d'UTL. Puis, nous concluons cet article avec quelques perspectives à court et moyen terme sur nos travaux.

Tous les exemples présentés dans cet article sont extraits de nombreuses expérimentations que nous avons menées avec nos étudiants de l'Institut Universitaire de Technologie de Laval ces dernières années. Ces expérimentations ont toutes consisté en l'utilisation d'un dispositif d'apprentissage de la programmation d'un serveur HTTP, constitué de six activités. L'artefact informatique repose sur l'environnement *Free Style Learning* (Brocke, 2001), implanté sur la plate-forme *Open-USS* (Grob et al., 2004), où les apprenants sont libres de choisir une activité parmi les six proposées. Ainsi, si les concepteurs du dispositif ont bien modélisé le scénario prédictif de la situation pédagogique, ce scénario n'était pas prescriptif et ne contraignait pas les choix des apprenants. Par contre, après chaque expérimentation, le scénario prédictif a été comparé avec les scénarios descriptifs obtenus, ce qui a conduit à plusieurs réingénieries du dispositif initial.

## **2. Traces et analyse d'usage des EIAH**

Certains travaux européens sont centrés sur la problématique de l'observation d'une session d'apprentissage et ont diffusé des résultats pertinents sur la représentation, l'acquisition et l'analyse d'une trace d'usage. La plupart de ces travaux (DPULS, 2005a), (ICALTS, 2004), (TRAILS, 2004) ont été menés dans le cadre du Réseau d'Excellence

Européen *Kaleidoscope* (Kaleidoscope, 2004) et ont inspiré notre proposition de modèle conceptuel d'une trace d'usage.

Le projet TRAILS (*Personalized and Collaborative Trails of Digital and Non-Digital Learning Objects*) s'est intéressé aux parcours (*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, c'est-à-dire une collection d'événements observés temporellement situés (Prié et al., 2007). 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. Cette approche est proche de celle de (Champin et al., 2003) et de (Egyed-Zsigmond et al., 2003), qui considèrent les traces d'utilisation dans un hypermédia comme une séquence d'actions, et les utilisent pour identifier l'objectif général de l'utilisateur. Dans nos travaux, nous ne restreignons pas le sens du terme trace à une séquence d'actions, mais 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. Pour faciliter cette compréhension, nous proposons que les résultats de cette analyse, pour être pertinents pour le concepteur, soient représentés comme un scénario descriptif (TRAILS utilise le terme de parcours émergent – *emergent trail*) qui peut être comparé avec le scénario prédictif (TRAILS utilise le terme de parcours planifié – *planned trail*).

Le projet ICALTS (*Interaction and Collaboration Analysis' supporting Teachers and Students' Self-regulation*) et le projet IA (*Interaction Analysis – Supporting participants in technology based learning activities*) qui l'a prolongé, "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" (IA, 2006). Ainsi, la définition des besoins d'analyse de l'usage 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". Nos travaux partagent entièrement cette position.

Le projet DPULS (*Design Patterns for recording and analyzing Usage of Learning Systems*) s'est attaché à l'étude de l'analyse de l'usage d'un EIAH dans des contextes de formation en entreprise ou académique, dans 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 de conception, (DPULS, 2005b), qui formalise sous la forme de patrons réutilisables des solutions éprouvées à des problèmes d'analyse des usages d'un EIAH. Ce projet a également proposé une définition plus large pour le concept d'indicateur qui est vu comme une variable significative 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, 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 méta-donnée 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.

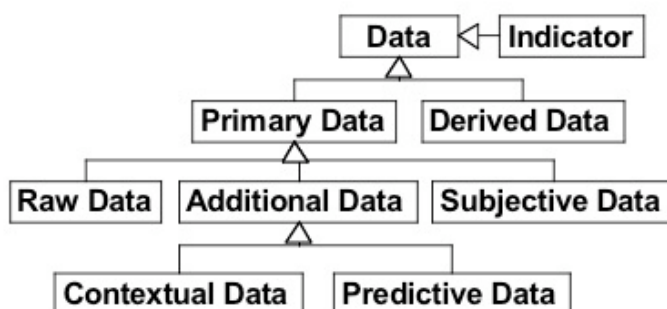


Figure 2 : La typologie des données dans DPULS

Le projet SBT (Système à Base de Traces) (Settoui et al., 2007) considère les traces comme une séquence temporelle d'observés toujours accompagnée d'un modèle capitalisable. Cette approche par modèles 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). L'approche ne retient que trois types de transformations : sélection, réécriture de motifs et fusion temporelle. Nous partageons l'approche par modèles mise en avant dans le projet SBT, mais nous considérons par contre qu'un indicateur n'est pas systématiquement

temporellement situé et que son établissement est une transformation à partir d'autres indicateurs et données (au sens de DPULS).

### **3. Une première proposition : décrire la trace collectée**

Pour instrumenter la transformation des traces générées par un dispositif d'apprentissage en données brutes (une donnée brute étant une donnée significative extraite de la trace, cf. la définition proposée par le projet DPULS) exprimées dans un format indépendant de ce dispositif d'apprentissage, nous avons décidé de définir un langage, nommé UTL 1<sup>2</sup> (Usage Tracking Language, version 1), permettant de décrire cette transformation (Iksal et Choquet, 2005). L'idée fondatrice d'UTL est que chaque trace d'utilisation prise en considération lors de la modélisation de l'observation concourt à témoigner d'un usage observé, et que leur analyse consiste à les mettre en relation avec le scénario pédagogique prédictif afin de pouvoir les analyser. UTL 1 ne considère que des traces collectées de manière automatique. De plus, il ne permet pas la description des moyens d'observation et ne supporte donc pas la modélisation de l'observation sous forme d'indicateurs : il est centré sur la transformation des traces dans un format indépendant du dispositif d'apprentissage, et sur leur association, en tant que témoin d'un usage observé, à une représentation du scénario pédagogique indépendante du méta-modèle d'expression pédagogique utilisé par le concepteur. Nous avons traduit UTL 1 sous la forme d'un schéma XML et ce langage peut 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. Les méthodes d'analyse permettant d'établir des indicateurs sont, dans le contexte d'utilisation de ce langage, réifiées par des outils acceptant en entrée des traces exprimées dans un format indépendant du dispositif d'apprentissage et une représentation du scénario pédagogique prédictif indépendante du méta-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 méta-modèle d'expression pédagogique différent.

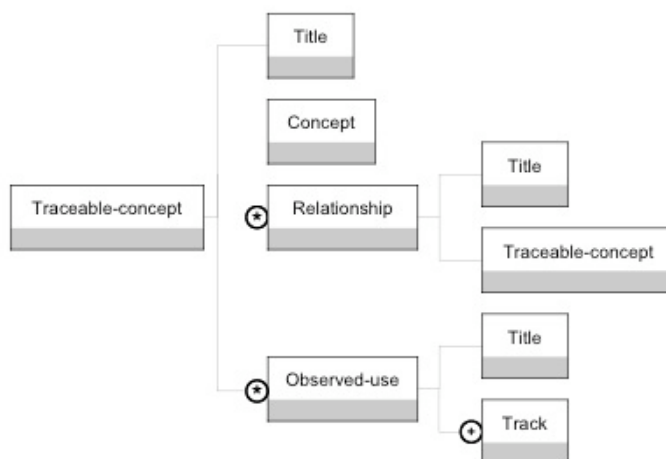
Dans un premier temps, nous présentons cette première version du langage UTL et dans un deuxième temps, nous proposons un exemple d'utilisation.

#### **3.1. Description du langage UTL 1**

Cette section présente les modèles d'information des types de données manipulées par la première version du langage UTL. UTL 1 est composé de deux parties. Une première partie, nommée UTL/S, est dédiée à la représentation de la transformée du scénario pédagogique. Cette transformation consiste (1) à décrire l'ensemble des *concepts traçables* du scénario pédagogique en leur associant un usage observé puis (2) à identifier dans le méta-modèle d'expression pédagogique employé par le concepteur les concepts candidats à l'observation, en fonction des objectifs d'observation que le concepteur se fixe. Cette transformation du scénario pédagogique se fait par interaction entre le concepteur et le développeur. Une seconde partie, nommée UTL/T, est dédiée 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 à la donnée correspondante dans les traces générées par l'EIAH. Des exemples de ces opérations de transformation sont donnés en section 3.2.

Les figures 3 et 4 représentent les modèles d'information de ces deux parties, UTL/S et UTL/T. Nous avons adopté un formalisme analogue à celui du modèle d'information de IMS Learning Design (IMS/LD, 2003).

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.



**Figure 3 : Le modèle d’information d’un concept traçable dans UTL 1**

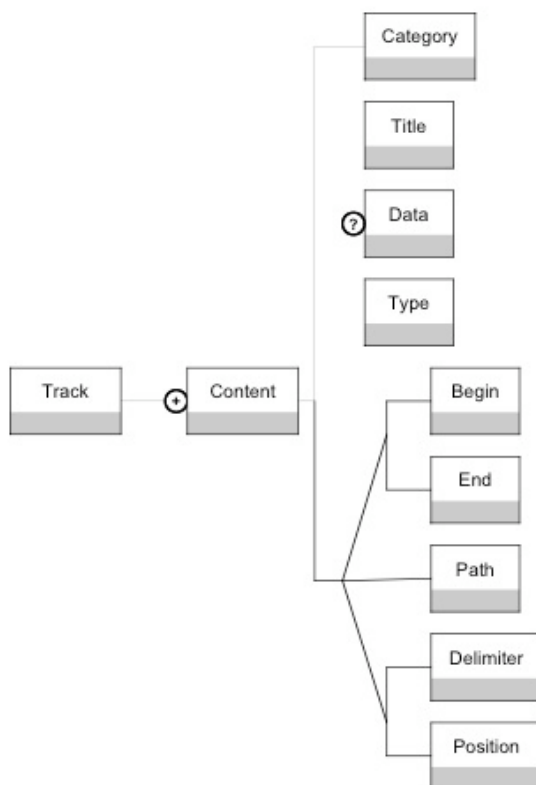
L’élément *titre* (title) permet de donner un nom au concept traçable (ou à la relation).

La description d’un *concept traçable* (Traceable-concept) est composée de toutes les *relations* (Relationship) avec les autres *concepts traçables*. Par exemple, si un concepteur a décrit un scénario avec IMS-LD et si les concepts d’activité et de ressource ont été identifiés comme traçables, l’élément *relation* pourra être utilisé pour indiquer qu’une "activité" X est réalisée en utilisant une "ressource" Y, et son *titre* (Title) pourrait être "utilisation".

L’élément *concept* permet d’associer le *concept traçable* à un concept dont il est l’instance dans le langage de modélisation pédagogique employé par le concepteur. Par exemple, si l’activité "Voir les Objectifs" est définie dans le scénario pédagogique exprimé en IMS-LD, l’élément *concept* du *concept traçable* "Voir les Objectifs" contiendra la valeur "Activity".

La dernière information correspond aux *usages observés* (Observed-use) associés au *concept traçable*. Elle permet la description des relations entre les *traces* et les *concepts traçables*.

Le *titre* (Title) d’un *usage observé* décrit la connexion sémantique entre les différents éléments *trace* (Track), par exemple les réponses d’un étudiant à un QCM ou, comme dans l’exemple présenté dans la prochaine sous-section, le pilotage interactif d’une ressource vidéo.



**Figure 4 : Le modèle d’information d’une trace**

Une trace est généralement composée de plusieurs informations dont seules certaines sont riches en signification pour le concepteur. La partie UTL/T d'UTL permet de représenter la méthode d'extraction de ces données. Le modèle d'information d'une *trace* (cf. figure 4) dans UTL est générique et nous proposons une implémentation qui est compatible avec la majorité des formats de traces collectées de manière automatique.

La description d'une *trace* (Track) se décompose en objets de deux *catégories* (Category): les *mots clés* (Keyword) et les *valeurs* (Value). Ces concepts génériques permettent la description de nombreux formats de traces depuis des fichiers textes structurés jusqu'aux bases de données et aux vidéos (la sous-section suivante donne un exemple pour un fichier de "logs" de type texte).

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 peut prendre pour valeur "Text", "XML" ou "Database". Cette liste est ouverte.

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 par exemple.

En ce qui concerne les fichiers de "logs" sous forme de fichiers texte structurés, UTL propose 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 marqueurs.

La donnée peut être un mot clé, c'est-à-dire un mot ou une phrase toujours présent au même endroit dans la trace, ce qui permet de la retrouver et de l'identifier, ou une valeur, c'est-à-dire une donnée collectée automatiquement par le dispositif d'apprentissage sur l'activité d'un acteur de la session d'apprentissage, comme le temps passé à lire une page, ou le nom d'une page lue, ou encore le résultat à un exercice. Dans les deux cas, la localisation du contenu permet de spécifier la position de la valeur ou du mot clé dans la trace. Des attributs spécifiques ont été prévus :

- *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.

### 3.2. Exemple d'utilisation de UTL 1

Cette sous-section présente un exemple d'utilisation de UTL 1 dans le cadre de 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, présentée en introduction.

Pour ces mises à l'essai, nous avons (1) défini la transformation de IMS-LD en identifiant, avec le concepteur, les concepts du langage dont au moins une des instances était susceptible de devenir un concept traçable, (2) utilisé cette transformation et UTL/S pour décrire les *concepts traçables* du scénario pédagogique prédictif défini par le concepteur, (3) utilisé UTL/T pour décrire les éléments *traces* en les situant dans les fichiers de logs générés par FSL et Open-USS, et (4) défini des outils génériques d'analyse permettant d'établir, à partir des représentations UTL du scénario pédagogique et des traces, des informations significatives pour le concepteur.

Dans son scénario prédictif original, le concepteur a défini une activité "Voir les Objectifs" consistant en la visualisation par l'apprenant d'une ressource vidéo d'introduction aux objectifs d'apprentissage ("VideoIntro"). Ainsi, au regard du scénario prédictif, les concepts "activity" et "learning object" ont été identifiés comme possédant des instances qui étaient des concepts traçables. Nous avons défini le schéma XML dont un extrait est représenté ci-dessous pour opérationnaliser la transformation du scénario pédagogique.

```
<xsd:element name="Activity" type="TraceableConceptType"
substitutionGroup="TraceableConcept"/>
<xsd:element name="LearningObject"
type="TraceableConceptType"
substitutionGroup="TraceableConcept"/>
```

Ce schéma est utilisé pour étendre le schéma XML d'UTL1 de manière à pouvoir représenter, sous la forme d'un fichier XML, les concepts traçables du scénario. Nous présentons un extrait XML de cette instanciation ci-dessous. L'activité "Voir les Objectifs" est un concept traçable; elle a une relation d'utilisation avec le concept traçable "VideoIntro", 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.

```

...
<Activity Title="Voir les Objectifs" Type="Abstract-scenario">
<Relationship Title="Use" Concept="VideoIntro"/>
</Activity>
...
<Resource Title="VideoIntro">
<ObservedUse Title="Managing">
<Track Title="Start">
<Content Category="Value" Title="Date" Type="Text" Begin="1" End="26"/>
<Content Category="Keyword" Title="Task" Type="Text" Begin="33" End="40">FreeApp</Content>
<Content Category="Keyword" Title="Object" Type="Text" Begin="42" End="57">Intro
gestartet</Content>
</Track>
...
<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>

```

Cette instanciation 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 suivant :

```

[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

```

Nous extrayons la trace suivante :

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

Qui nous permet d'obtenir les données suivantes :

```
Date of the track: 04/12/2002:03:21:58
```

```
Duration of the video: 182.0s
```

Ces données ne sont pas directement exploitables par le concepteur sous cette forme. L'approche développée dans cette première version d'UTL 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" (éléments "Track"), elle-même utilisée par l'activité "Voir les Objectifs" (élément "Relationship").

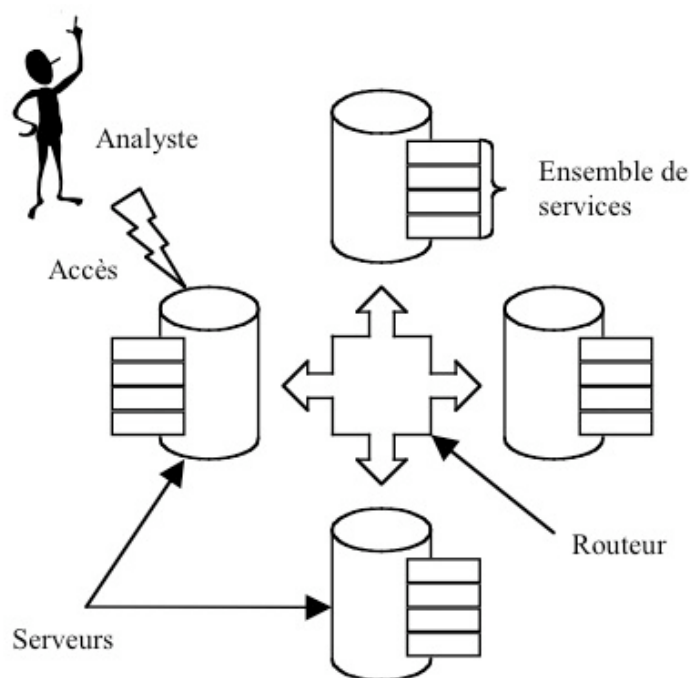


Il est alors possible de développer ou de réutiliser des outils travaillant sur ces données exprimées dans un format indépendant des traces générées par le dispositif d'apprentissage et du langage de modélisation pédagogique. Ces outils, partageant les mêmes formats d'entrée et de sortie, peuvent être combinés entre eux de manière à établir les indicateurs demandés par le concepteur.

Dans l'extrait qui suit, nous obtenons le résultat produit par un outil d'analyse permettant l'établissement de l'indicateur "nombre d'étudiants qui n'ont pas terminé la vidéo d'introduction". Sur la base de ce résultat, la ressource "VideoIntro" a été modifiée, de manière à ne pas prendre en compte les dernières secondes de la vidéo, dont le contenu n'est pas signifiant.

```
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
```

Nous avons développé une architecture démontrant la validité de l'approche pour les outils d'analyse. Cette architecture est basée sur les technologies Java et RMI. Elle permet la création et le partage d'outils d'analyse distribués et enregistrés auprès d'un serveur. La figure 5 schématise cette architecture. La figure 6 présente l'interface d'un outil d'analyse utilisé pour reconstruire la séquence d'activités que chaque étudiant a réalisée, cet outil utilise les résultats d'un outil qui reconstruit la séquence des ressources utilisées puis s'appuie sur la description en UTL 1 du scénario pédagogique pour associer les ressources aux activités. Notons ici que le spectre de l'application de cet outil est limité aux scénarios pédagogiques dont les ressources ne sont utilisées que dans une seule activité.



**Figure 5 : Architecture d'outils d'analyse distribués**

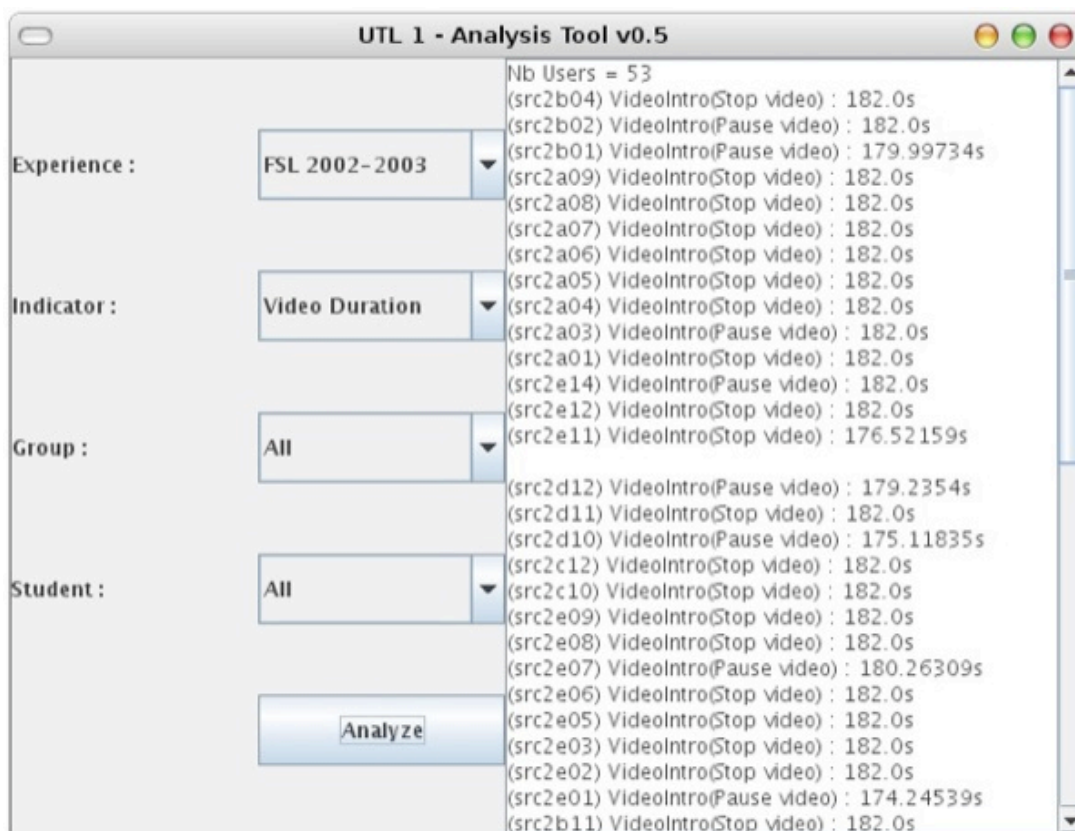


Figure 6 : Interface d'un outil d'analyse exploitant UTL 1

## 4. De la trace à l'indicateur (UTL 2)

Pour instrumenter la modélisation et l'analyse de l'observation dans une perspective de capitalisation, nous avons défini une seconde version du langage UTL ([Choquet et Iksal, 2007](#)) qui permet de modéliser un observable et, plus largement, toute donnée impliquée dans la modélisation de l'observation. Le langage UTL 2 instrumente la description des méthodes d'analyse de l'observation en vue de leur capitalisation en étendant UTL 1 avec un ensemble d'éléments dédiés à cette description et regroupés dans une section du langage nommée UTL/P (UTL/Patron). Cette section n'est, pour le moment, pas opérationnalisée puisqu'elle ne propose pas actuellement de grammaire formelle permettant de spécifier une méthode d'analyse. Elle permet par contre une description structurée des indicateurs dans une forme indépendante des formats de traces générées par un dispositif d'apprentissage et du langage de modélisation pédagogique employé pour décrire le scénario pédagogique.

Nous présentons dans une première sous-section le modèle DGU (Defining, Getting, Using) sur lequel se fondent les descriptions UTL 2. Dans une deuxième sous-section, nous développons le modèle conceptuel d'UTL 2 et les modèles d'information de ses éléments.

### 4.1. Trois points de vue sur les traces : le modèle DGU

Les pratiques consistant à recueillir et exploiter des données acquises pendant l'interaction entre un système informatique et ses utilisateurs sont courantes et ont en commun de composer avec un grand nombre de données. Généralement, il s'agit alors de formuler des hypothèses sur la méthode et l'objectif de l'analyse, d'extraire du vaste ensemble de données recueillies les informations pertinentes, et de les lier entre elles pour établir un résultat.

Dans le cadre de la réingénierie pédagogique, nous proposons de modéliser la trace d'usage en amont de la session d'apprentissage, et donc de considérer les traces comme des objets pédagogiques, de même nature que toute autre ressource, comme un scénario pédagogique par exemple. Si cette position est fréquente dans les systèmes existants, notamment lorsqu'il s'agit de retourner des informations aux apprenants et/ou au système, cela est plus inhabituel pour instrumenter le tuteur dans sa tâche, et rare lorsque le destinataire des traces est le concepteur de l'EIAH.

Dans cet esprit, nous proposons que le concepteur engagé dans un processus d'ingénierie d'un EIAH modélise les besoins d'observation au même moment que la situation pédagogique à mettre en place, et ce, avec des modalités comparables ([Barré et al., 2005](#)). Ces besoins d'observation sont à instancier sur le dispositif d'apprentissage par l'analyste et le développeur qui spécifient ainsi les moyens d'observation à déployer. Enfin, en se fondant sur la modélisation des besoins d'observation, l'analyste et le développeur modélisent l'usage attendu des traces collectées, notamment en termes de reconstruction des scénarios descriptifs à comparer avec le scénario prédictif de la situation pédagogique. Nous proposons ainsi de modéliser une trace suivant trois facettes (cf. figure 7) :

- la facette définition (*Defining*) permet de modéliser le besoin d'observation ;
- la facette obtention (*Getting*) de la trace permet de modéliser le moyen d'observation ;
- la facette utilisation (*Using*) précise l'utilité la donnée d'observation et décrit l'utilisation qui en est faite.

Si ce modèle DGU suppose une approche de modélisation "descendante" de la trace, de l'énoncé du besoin d'observation à l'identification du moyen d'observation, certains dispositifs, tels que les plates-formes disponibles sur le marché, collectent par défaut des données observées. Si ces traces s'avèrent utiles d'un point de vue pédagogique, elles peuvent également être modélisées *a posteriori* selon le modèle DGU. Nous parlons alors ici d'élicitation d'indicateurs pédagogiques.

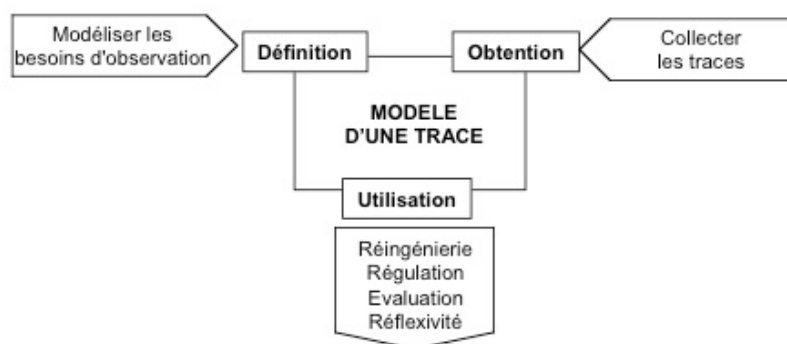


Figure 7 : Le modèle DGU

## 4.2. Description du langage UTL 2

### 4.2.1. Modèle conceptuel du langage UTL 2

UTL 2 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 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 la section 3.1 et exemplifiée dans la section 3.2.
- la partie « trace » (UTL/T) permet de lier la description d'une donnée à collecter à 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 la section 3.1 et exemplifiée dans la section 3.2.

Chaque donnée est décrite conformément au modèle DGU, selon les trois points de vue : définition, obtention, utilisation.

Nous avons identifié deux types principaux de données impliquées dans l'analyse de l'observation : la *donnée dérivée* et la *donnée primaire* (cf. figure 8).

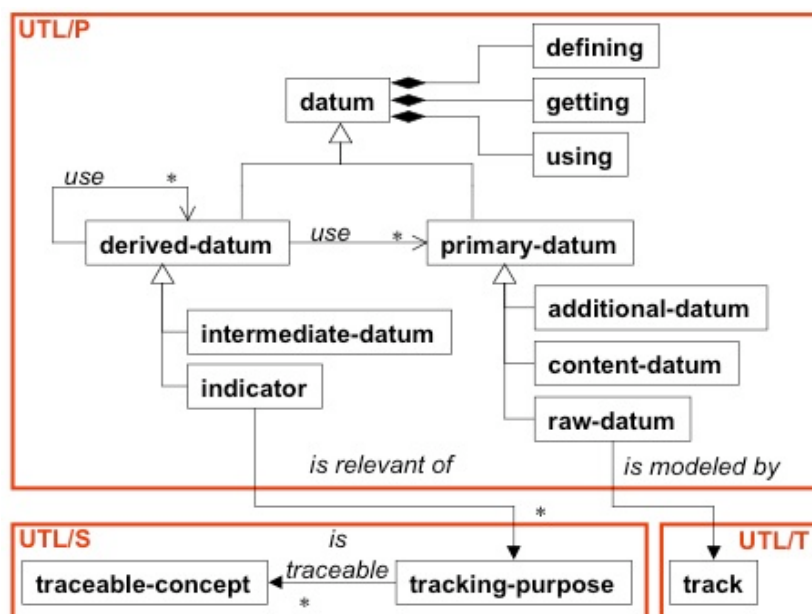


Figure 8 : Modèle conceptuel du méta-langage UTL 2

- 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.
- Une *donnée brute* est obtenue par transformation de la trace : 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 UTL 1 (cf. section 3.1.).
- 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 nature très variable, par exemple la valeur d'un champ d'une méta-donnée caractérisant une ressource, une taxonomie, une donnée *ad hoc*.
- Les *données dérivées* sont calculées ou établies à l'aide des *données primaires* et/ou d'autres *données dérivées*.
- 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 méta-langage d'expression pédagogique (cf. section 3.1).
- 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.

#### 4.2.2. Le modèle d'information de la donnée brute

La figure 9 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.

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. 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 9 pour rappeler qu'il est conforme à la première version d'UTL (UTL 1).

- la *localisation* (Location) de la donnée, il s'agit souvent de l'URL du fichier qui la contient.

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, conformément à la méthode présentée en section 3.1.; *format* décrit comment la donnée est représentée, une fois obtenue (son schéma XML).

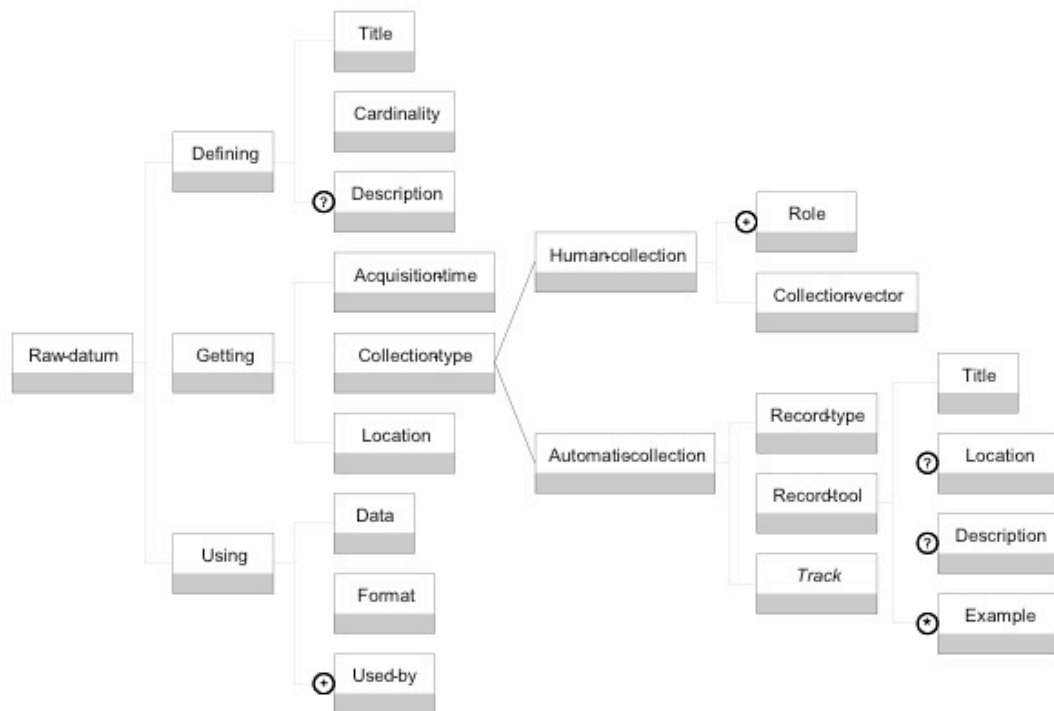


Figure 9 : Le modèle d'information d'une donnée brute

### 4.2.3. Le modèle d'information de la donnée de production

La figure 10 détaille les trois facettes d'une *donnée de production*.

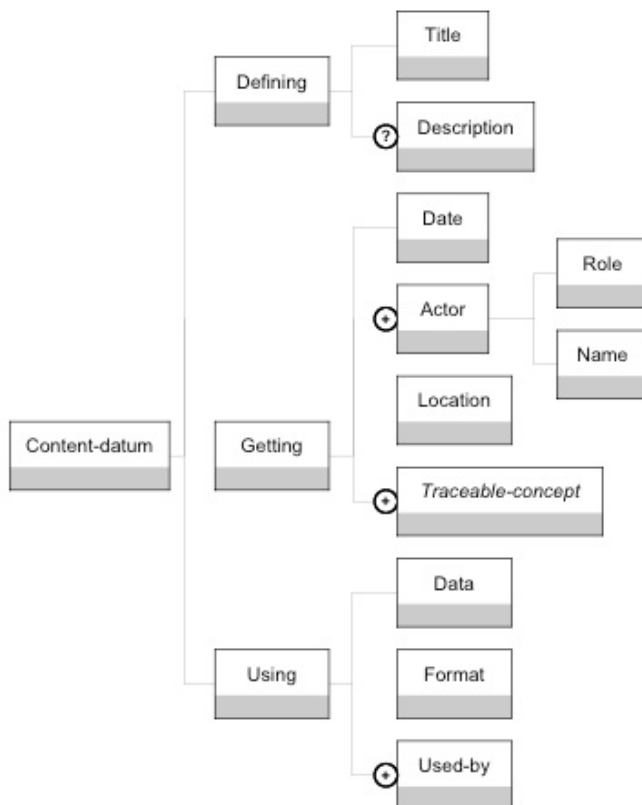
Comme pour la *donnée brute*, la facette "définition" de la *donnée de production* est composée d'un *titre* et d'une éventuelle *description*.

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. La facette "obtention" est caractérisée par :

- la *localisation* de la production,
- la *date* de la production,
- *l'acteur* (Actor) qui a produit cette donnée,
- au moins un *concept traçable* (Traceable-concept) du scénario pédagogique.

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*.



**Figure 10 : Le modèle d'information d'une donnée de production**

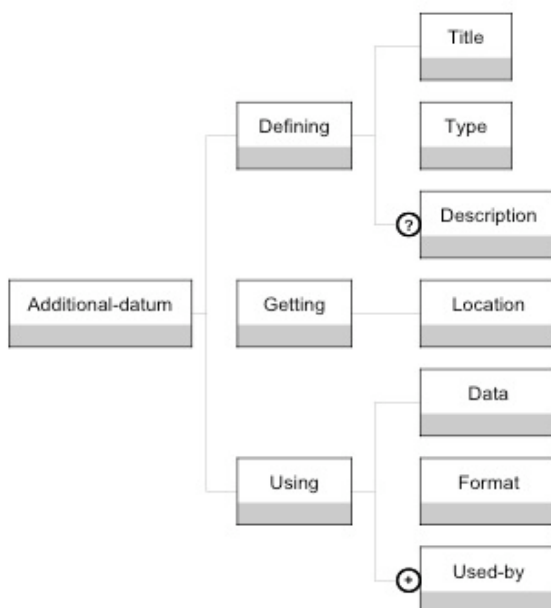
#### 4.2.4. Le modèle d'information de la donnée additionnelle

La figure 11 détaille les trois facettes d'une *donnée additionnelle*.

Les *données additionnelles* sont variées et nombreuses (une ontologie, une taxonomie du domaine, un curriculum académique, etc.), c'est pourquoi la facette "définition" possède en plus du *titre* et de la *description*, un élément *type* (Type) pour classifier la donnée.

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.

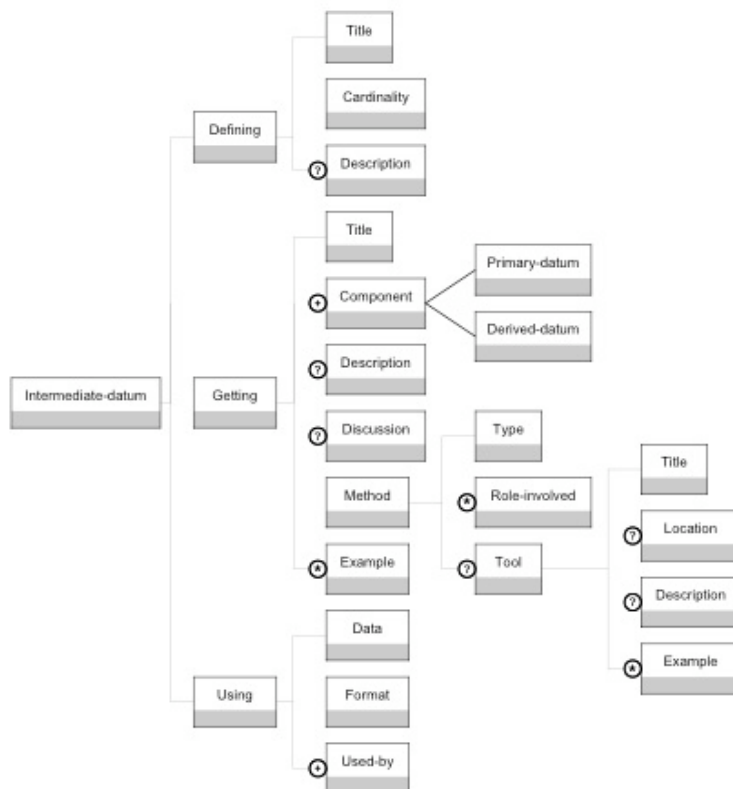
La facette "utilisation" est composée du *contenu* de la donnée, de son *format* et de la liste des données qui l'utilisent.



**Figure 11 : Le modèle d'information de la donnée supplémentaire**

#### 4.2.5. Le modèle d'information de la donnée intermédiaire

La figure 12 détaille les trois facettes d'une *donnée intermédiaire*.



**Figure 12 : Le modèle d'information de la donnée intermédiaire**

La facette "définition" est composée du *titre* de la donnée, d'une *cardinalité* et d'une éventuelle *description*.

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

La facette "obtention" peut être également composée 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 peuvent être utilisés, il convient de définir plusieurs *données intermédiaires*.

La section "utilisation" est composée du *contenu* de la donnée, de son *format* et de la liste des données qui l'utilisent.

#### 4.2.6. Le modèle d'information de l'indicateur

La figure 13 détaille les trois facettes d'un *indicateur*. Les facettes "définition" et "obtention" sont similaires à celles de la donnée intermédiaire.

La facette "utilisation" est caractérisée par quatre éléments :

- l'élément *donnée* (Data) permet de stocker la valeur de l'*indicateur*, une fois calculé ou établi.
- l'élément *contexte pédagogique* (Pedagogical-context) définit le ou les *objectifs d'observation* (Tracking-purpose) de l'*indicateur*. Comme le montre la figure 14, l'*objectif d'observation* est décrit par :

un ou plusieurs *concepts traçables* (Traceable-concept) permettent d'associer l'*indicateur* à un élément du scénario pédagogique.

un *type d'exploitation* (Type). Nous avons défini actuellement quatre types d'exploitation – réingénierie, régulation, évaluation et réflexivité. Nous considérons cette liste ouverte mais, dans le contexte du projet REDiM, nous limitons notre usage du langage à la réingénierie.

un ou plusieurs *rôles* (Recipient-role) précisent le destinataire de l'indicateur.

une *description* peut être ajoutée afin de décrire de façon détaillée l'objectif d'observation.

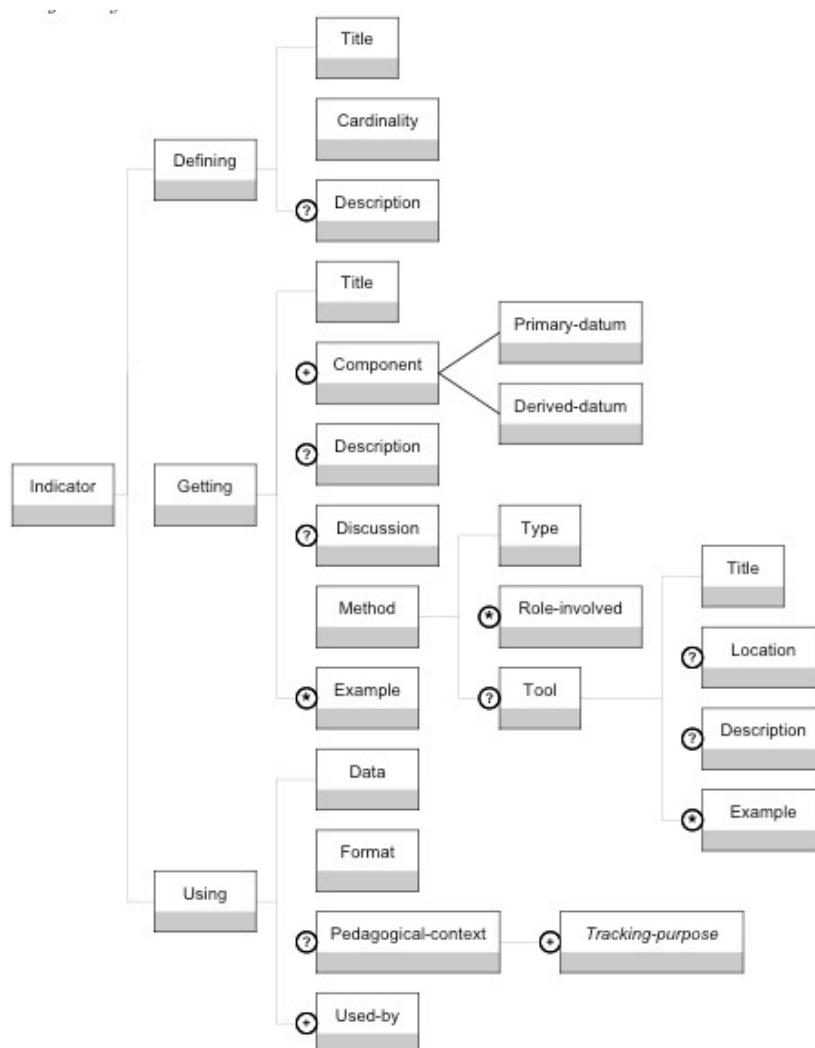


Figure 13 : Le modèle d'information d'un indicateur

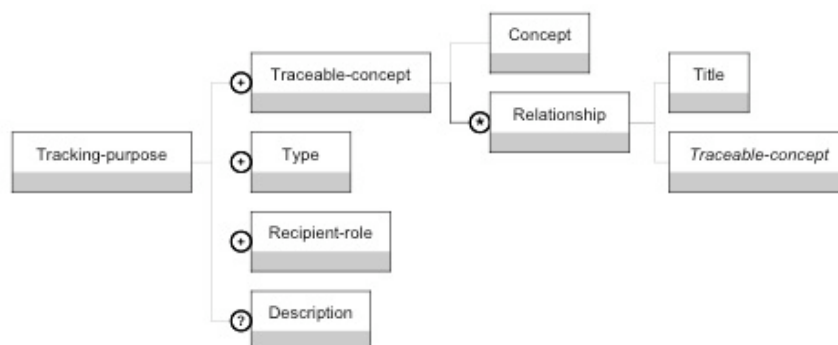


Figure 14 : Le modèle d'information d'un objectif d'observation

## 5. Scénarios d'utilisation

Cette section présente trois scénarios d'utilisation possible du langage UTL. Nous décrivons ces scénarios en prenant des exemples issus de deux expérimentations différentes pour lesquelles nous disposons de trois années de logs. Pour chaque cas étudié, nous disposons du scénario pédagogique prédictif décrit en *IMS Learning Design*. Toutes les ressources utilisées par le scénario sont indexées à l'aide du standard *Learning Object Metadata* (LOM).



2002). Nous utilisons le langage UTL pour apporter de la sémantique à toutes les traces produites par les environnements. La première étape consiste en l'interprétation des traces à partir du modèle du concepteur (i.e. les concepts de modélisation du scénario pédagogique) et de la description sémantique des traces. Dans un second temps, les observations sur l'utilisation du système d'apprentissage sont disponibles pour l'analyse.

## 5.1. Analyse d'utilisation : interprétation des traces

L'analyse automatique de traces commence par une interprétation de ces traces. UTL 1 a été conçu pour ajouter de la sémantique au contenu des fichiers de *logs*. Nous l'utilisons dans un premier temps pour filtrer le contenu de ces fichiers. Cela revient à conserver uniquement les traces considérées comme pertinentes par le concepteur, c'est-à-dire les traces qui sont explicitement décrites avec UTL. Le deuxième usage d'UTL consiste à typer les traces et à extraire les valeurs représentatives de l'activité des acteurs du dispositif. Le résultat de cette étape est une structure de données qui contient les traces interprétables (cf. section 3.3 pour un exemple) et qui est partageable entre les différents services d'analyse. Cette structure de données est aussi disponible pour tout chercheur qui souhaite développer de nouveaux services.

## 5.2. Analyse d'utilisation : calcul des données dérivées

Il existe de nombreuses manières d'utiliser les traces interprétées, par exemple pour évaluer l'utilisation d'une ressource, ou comparer le scénario réalisé par l'apprenant avec le scénario prédit par le concepteur. De plus, différentes interprétations sont possibles avec les mêmes données brutes. Par exemple, les courriels échangés par les différents acteurs de la session d'apprentissage peuvent être analysés (*parsing*) afin de retrouver des marqueurs sémantiques qui peuvent révéler des rôles socio-affectifs dans un groupe (Cottier et Schmidt, 2005), ou visualisés sous la forme d'un graphe orienté entre les acteurs afin de mesurer la cohésion dans un groupe (Reffay et Chanier, 2003). L'utilisation de traces augmentées de sémantique permet la définition de services. Les résultats des services d'analyse sont par exemple : le taux d'utilisation d'une ressource, la performance d'un apprenant, l'émergence d'un rôle (*leader*, etc.), la reconstruction d'un scénario d'apprentissage réalisé, ou encore la détection d'une séquence de ressources utilisées qui n'a pas été prescrite. Afin d'exposer quelques idées sur l'analyse de l'utilisation d'un scénario, nous allons présenter deux cas : l'établissement d'une donnée statistique, et un résultat d'analyse qui doit être retranscrit dans le méta-modèle d'expression pédagogique du concepteur.

### 5.2.1. Premier cas : l'établissement d'une donnée statistique

Ces données sont par exemple : le taux d'utilisation d'une ressource, la note moyenne d'un exercice, ou le temps passé sur une activité particulière (le plus rapide, le temps moyen et le temps le plus long). Nous filtrons les traces en fonction de la sémantique associée et réalisons quelques calculs sur les données obtenues. Dans le cas du taux d'utilisation, une première solution est de compter le nombre d'apprenants pour lesquels nous avons trouvé au moins une trace liée à l'utilisation de la ressource concernée. A l'occasion d'une de nos expérimentations, dans une première phase d'ingénierie, le concepteur a déclaré un besoin d'observation sur l'utilisation d'une ressource particulière. Après la session, l'observation a mis en évidence que de nombreux apprenants ont utilisé moins de 15 secondes certaines ressources en début de session. Ces utilisations révélaient une exploration préalable du dispositif par les étudiants. Dans un second temps, c'est-à-dire dans une phase de réingénierie, nous avons donc modifié les outils d'observation pour isoler les utilisations de ressources de moins de 15 secondes afin de détecter une période d'exploration.

### 5.2.2. Second cas : la retranscription dans le modèle du concepteur

L'un des objectifs principaux de la réingénierie dirigée par les modèles est d'utiliser le même méta-modèle d'expression pédagogique pour la description du scénario prédictif que pour le scénario observé généré à partir des traces produites par le système d'apprentissage. Dans nos premières expérimentations, nous avons utilisé IMS-LD. Le principal intérêt dans l'utilisation d'un modèle commun réside dans la possibilité de comparer les différents scénarios, ce qui a permis aux concepteurs d'identifier des utilisations imprévues des ressources ou des incohérences dans les séquences d'activités. Dans l'une de nos expérimentations, avec FSL, nous avons observé que certains apprenants ont utilisé au début de la session d'apprentissage, un exercice de type QCM, que le concepteur destinait à l'évaluation des acquis en fin d'activité. Les apprenants ont navigué dans la liste de questions afin de découvrir le domaine abordé, de s'auto-évaluer avant de commencer l'activité. Cette observation a mené le concepteur à proposer deux facettes pour l'exercice, une pour l'évaluation, en fin d'activité, et une seconde pour découvrir le domaine, en début d'activité. Pour évaluer la pertinence de ces actes de réingénierie, nous avons considéré deux types de retranscription du scénario observé : une première qui est générée à partir des traces d'un seul apprenant, et une seconde qui représente la combinaison de plusieurs scénarios d'apprenants.

Pour la retranscription du scénario observé d'un seul apprenant, nous utilisons le méta-modèle d'expression pédagogique afin d'identifier les concepts clés, par exemple le concept d'activité dans IMS-LD. Ensuite, nous filtrons les traces en fonction de ce concept et de tous ses composants. Pour finir, nous organisons toutes les instances de ces concepts clés en une séquence qui correspond au scénario observé.

Pour la retranscription d'une combinaison des scénarios de tous les apprenants, il faut disposer de tous les scénarios observés des apprenants. Nous comparons alors les séquences de concepts clés (par exemple, les activités), et nous comparons en profondeur chaque concept. Nous calculons le pourcentage de chaque utilisation ou de la position de

chaque concept dans la séquence. le résultat obtenu est un graphe où chaque relation est qualifiée avec le pourcentage des apprenants qui ont choisi cette direction.

### 5.3. Analyse d'utilisation : des données brutes aux indicateurs

Dans cette partie, nous proposons un cas où le calcul des indicateurs à partir des traces interprétées est conforme aux modèles d'information de la première partie de cet article. Ce cas correspond à un patron de conception proposé par le projet DPULS (DPULS, 2005a) qui est intitulé : "Playing Around with Learning Resources". Ce patron, disponible sur (DPULS, 2005b) propose une approche de détection des apprenants qui "papillonnent" entre les ressources au démarrage de l'activité d'apprentissage. Dans ce patron, nous proposons une solution fondée sur le calcul de deux indicateurs : "la caractérisation de la séquence de ressources" et "la caractérisation du temps de réalisation d'une activité". La méthode d'obtention du premier indicateur est la suivante : la séquence d'utilisation des ressources par l'apprenant est évaluée à "non significative" si le temps passé sur chaque ressource est inférieur à un pourcentage (dans notre cas 10%) du temps estimé de réalisation défini pour chacune des ressources concernées. En ce qui concerne le deuxième indicateur, le temps de réalisation d'une activité est qualifié de "commencement" si la durée effective de l'activité est inférieure à un pourcentage (dans notre cas 10%) du temps estimé de réalisation de l'activité.

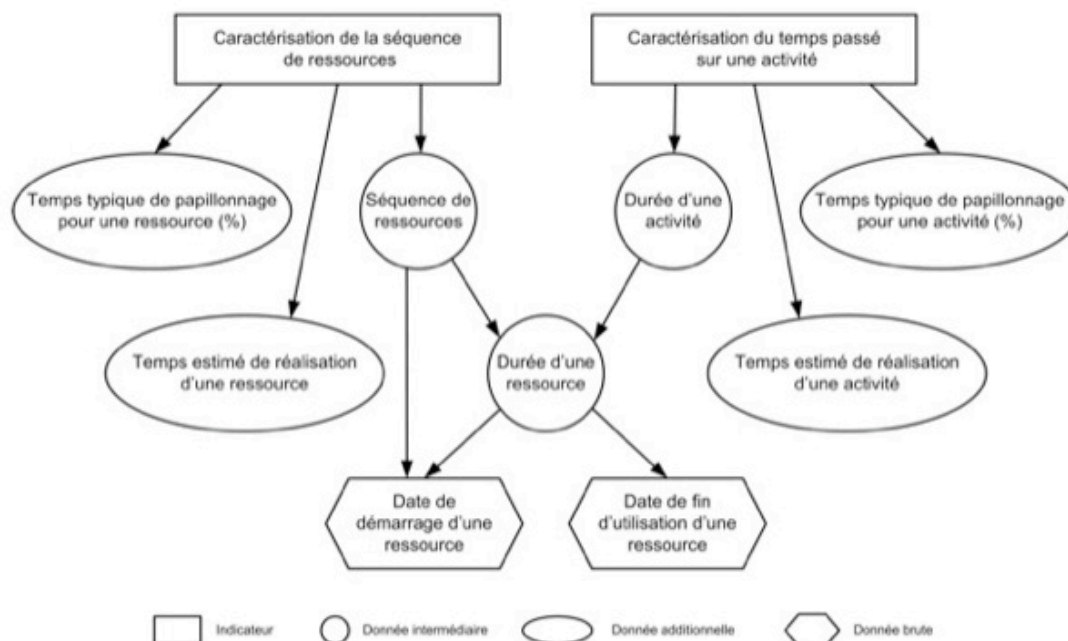


Figure 15 : Cartographie des indicateurs et données utilisées

Dans la figure 15, le graphe représente l'utilisation des différentes données obtenues. Nous trouvons les indicateurs mais aussi les données intermédiaires, des données additionnelles ainsi que des données brutes. UTL 1 est capable d'identifier et d'extraire ces données brutes. UTL 2 permet de prendre en compte ces données brutes lors de la génération d'indicateurs pour le concepteur pédagogique. L'heure de début et l'heure de fin d'utilisation d'une ressource permettent l'évaluation de la durée d'utilisation de la ressource. Les données additionnelles, comme le temps estimé de réalisation, peuvent être extraites du scénario prédictif (par exemple grâce au champ 5.9 du LOM : "Durée d'apprentissage"). Ces données peuvent aussi être données par le concepteur comme le temps typique de "papillonnage" sur une ressource. Il s'agit d'un pourcentage du temps d'utilisation d'une ressource considéré comme un temps minimum en dessous duquel l'utilisation n'est pas à prendre en compte du point de vue de l'apprentissage. Les tableaux ci-après présentent la description de ces données avec UTL 2, conformément aux modèles d'information présentés dans cet article. Le tableau 1 présente la table d'information pour la donnée brute "Date de début d'utilisation d'une ressource", le tableau 2 correspond à la donnée additionnelle "Temps typique de papillonnage sur une ressource", le tableau 3 décrit la donnée intermédiaire "Séquence de ressources" et le tableau 4 présente l'indicateur "Caractérisation de la séquence de ressources".

D	Title	Redim-RD-StartVideo					
	Card.	n					
	Description	Cette donnée correspond à la date de début de l'utilisation de la ressource vidéo.					
G	Acquisition-time	During-session					
	Collection-type	Automatic-collection	Track	Record-type	Log-file		
				Record-tool	Title	Méthodes de génération de traces de FreeStyleLearning	
					Location		
				Content	Category	Keyword	
					Title	Outil	
					Data	FreeApp	
					Type	Text	
					Begin	33	
					End	40	
				Content	Category	Keyword	
					Title	Action	
					Data	Intro gestartet	
					Type	Text	
					Begin	42	
					End	57	
				Content	Category	Value	
					Title	Date	
					Type	Text	
					Begin	1	
End					26		
Location	~exp/StudentID/file.FSL						
U	Data						
	Format	<pre>&lt;rawdatum type= « Redim-RD-StartVideo »&gt; &lt;date&gt; long &lt;/date&gt; &lt;student&gt; string &lt;/student&gt;&gt; &lt;/rawdatum&gt;</pre>					
	Used-by	Redim-ID-SeqRes					
	Used-by	Redim-ID-ResDur					

**Tableau 1 : Table d'information d'une donnée brute**

D	Title	Redim-AD-TypTimeVideo
	Type	Scenario
	Description	Cette donnée donne la durée estimée de consultation de la ressource vidéo.
G	Location	Section métadonnée du scénario
U	Data	182
	Format	Integer, time in seconds
	Used-by	Redim-ID-SeqRes

**Tableau 2 : Table d'information d'une donnée additionnelle**

D	Title	Redim-ID-SeqRes		
	Cardinality	n		
	Description	Calcule les séquences d'utilisation des ressources par étudiant		
G	Title	Filtrage et calcul des séquences d'utilisation des ressources par étudiant		
	Component	Primary-datum	Redim-RD-StartVideo	
		Derived-datum	Redim-ID-ResDur	
	Description	Il s'agit de trier les sessions d'utilisation de ressources par étudiant et par date, de retrouver les durées et de combiner ces données.		
	Method	Type	automatic	
		Tool	Title	Calculation_Resource_Sequence
Location			Redim.jar	
U	Data			
	Format	<pre>&lt;intermediatedatum type= « Redim-ID-SeqRes »&gt; &lt;student name= « string »&gt; &lt;session date= « log »&gt; &lt;resource&gt; string &lt;/resource&gt; &lt;duration&gt; long &lt;/duration&gt; &lt;/session&gt; &lt;/student&gt; &lt;/intermediatedatum&gt;</pre>		
	Used-by	Redim-I-CaracSeqRes		

**Tableau 3 : Table d'information d'une donnée intermédiaire**

D	Title	Redim-I-CaracSeqRes			
	Cardinality	n			
	Description	Evaluation de la pertinence d'une séquence d'utilisation de ressources par étudiant			
G	Title	Comparaison de durées de séquences et de durées préconisées			
	Component	Primary-datum	Redim-AD-TypTimeVideo		
		Primary-datum	Redim-AD-TypTimeGaming		
		Derived-datum	Redim-ID-SeqRes		
	Method	Type	automatic		
		Tool	Title	Calculation_Relevance_Resources_Sequence	
			Location	Redim.jar	
Desc.			La séquence de ressources réalisée par un apprenant est évaluée comme « non significative » si la durée de chaque ressource est inférieure à une fraction (10% par exemple) de la durée préconisée.		
U	Data				
	Format	<pre>&lt;indicator type= « Redim-I-CaracSeqRes »&gt; &lt;student name= « string » evaluation= « string »&gt; &lt;session date= « log »&gt; &lt;resource&gt; string &lt;/resource&gt; &lt;duration&gt; long &lt;/duration&gt; &lt;/session&gt; &lt;/student&gt; &lt;/indicator&gt;</pre>			
	Pedagogical-context	Tracking-purpose	Traceable-concept	Concept	Resource
		Type	Reengineering		
		Recipient-role	Designer		
		Description	Détection des périodes de papillonnage dans l'utilisation des ressources pour mettre en évidence une utilisation normale ou une découverte du système		

Tableau 4 : Table d'information d'un indicateur

Toutes ces descriptions sont exprimées en XML et peuvent donc être utilisées par un système en vue de pré-calculer les données et d'interagir avec l'analyste et/ou le concepteur dans le cas où l'établissement de ces données requiert une intervention humaine.

## 6. Perspectives

Le modèle DGU (*Defining, Getting, Using*) que nous avons présenté, est adapté à la modélisation et à la représentation de ce que le dispositif d'apprentissage doit tracer, en s'appuyant sur le scénario pédagogique prédictif conçu pour la situation d'apprentissage mise en place. Pour chaque donnée, les concepteurs sont amenés à définir ce qu'il faut tracer (i.e. les besoins d'observation), comment tracer (i.e. les moyens d'observation) et pourquoi tracer (i.e. l'objectif d'observation). L'opérationnalisation d'UTL 2 nécessite un langage de description du moyen d'obtention des données, des travaux sont en cours sur le sujet. Lorsqu'il sera opérationnalisé, les données pourront être combinées entre elles afin d'établir des indicateurs de haut niveau significatifs pour le concepteur et/ou l'analyste. Ce langage permet de définir et de mettre en évidence des corrélations entre le scénario prédictif et les scénarios descriptifs reconstruits à l'aide des observations recueillies pendant une session d'apprentissage, et

représentés à l'aide du même méta-modèle d'expression pédagogique.

Une première mise à l'épreuve externe, en collaboration avec l'équipe « observation et trace » de l'Université de Savoie, a été menée, les résultats ont montré l'importance de la scénarisation de l'observation dans le cadre de l'analyse des usages d'un EIAH. D'autres mises à l'épreuve sont à l'étude actuellement.

Cependant, des travaux, par exemple (Seel et Dijkstra, 1997), ont montré que les enseignants et les formateurs – qui selon nous sont les principaux concepteurs potentiels de dispositifs d'apprentissage médiés – ont des difficultés avec les techniques formelles de modélisation pédagogique, tout spécialement lorsqu'il s'agit d'explicitier et de réifier leurs intentions de conception. Si un des intérêts d'UTL est de permettre au concepteur de définir des besoins d'observation (facette *Getting*) et d'explicitier des objectifs d'observation (facette *Using*), puis de négocier avec les développeurs la définition des moyens d'observation (facette *Getting*) d'un indicateur, nous sommes bien conscients que ce langage ne peut être utilisé tel quel par un enseignant ou un formateur ne disposant pas de connaissances particulières en informatique. C'est pourquoi nous travaillons actuellement dans trois directions de recherche que nous considérons intéressantes et complémentaires pour mieux impliquer enseignants et formateurs dans la conception des EIAH.

Nous développons un ensemble de règles et de filtres pouvant être appliqués sur le méta-schéma XML et le méta-modèle d'expression pédagogique qu'utilise le concepteur (Barré et al., 2007). Ces règles et filtres pro-actifs raisonnent sur la structure du méta-modèle d'expression pédagogique et suggèrent au concepteur des besoins d'observation, liés au scénario pédagogique qu'il définit. Si le concepteur y associe un objectif d'observation qu'il juge pertinent, ces suggestions sont des indicateurs à modéliser et à collecter.

Parallèlement, et suite aux travaux menés dans le projet DPULS, nous cherchons à valider l'approche "patron" d'UTL, et nous travaillons donc à l'établissement et à l'exploitation de patrons de conception dédiés à l'analyse des usages. Notons d'ailleurs que les modèles d'information présentés dans cet article permettent d'associer une donnée, par exemple un indicateur, à un patron de conception. Pour cela, il est nécessaire de détailler et décrire la solution d'analyse proposée par un patron de conception pour l'établissement d'un indicateur (problème traité par le patron) avec un objectif d'observation donné (contexte d'application de la solution proposée par le patron).

Enfin, comme nous l'avons déjà souligné, nous veillons à être indépendant à la fois du méta-modèle d'expression pédagogique utilisé pour décrire le scénario d'apprentissage, et des formats de représentation utilisés par le dispositif d'apprentissage pour les traces. Ce choix a été fait pour rester pertinent avec plusieurs contextes de développement, mais aussi parce que nous pensons que l'émergence d'un standard, et donc l'appropriation généralisée d'un méta-modèle d'expression pédagogique par les concepteurs pédagogiques, n'est pas proche. Nous pensons, en nous appuyant sur nos expérimentations (El Kechaï et Choquet, 2007), que le – souvent long – apprentissage d'une technique et/ou d'un méta-modèle d'expression pédagogique, notamment en considérant la nécessaire adhésion à la métaphore et à la sémantique du modèle de ce langage, est un frein à l'implication des enseignants et des formateurs dans la conception des EIAH. C'est pourquoi, dans le contexte particulier de la conception collaborative de scénarios pédagogiques, nous travaillons à permettre aux concepteurs de définir eux-mêmes leur méta-modèle (vocabulaire, syntaxe abstraite, syntaxe concrète) d'expression d'un scénario pédagogique, tout en assurant sa transformation vers un modèle computationnel en XML. Nous avons développé un prototype d'éditeur fondé sur ces idées, l'approche ayant été testée lors d'une expérimentation préalable. Dans ce contexte, nous étudierons les possibilités de modélisation conjointe du scénario pédagogique et des besoins d'observation de la situation d'apprentissage.

## Bibliographie

- BARRÉ V., EL KECHAÏ H., CHOQUET C. (2005). Re-engineering of Collaborative E-learning Systems: Evaluation of System, Collaboration and Acquired Knowledge Qualities. *Workshop "Usage Analysis in Learning Systems", AIED'05*, Amsterdam, Pays-Bas, p. 9-16.
- BARRÉ V., CHOQUET C., IKSAL S. (2007). Observation Scenario Development Using Recommendations, *IEEE International Conference on Advanced Learning Technologies (ICALT'2007)*, Niigata, Japan, p. 605-607.
- BAZSALISCZA M., NAIM P. (2001). *Data Mining pour le Web*. Paris, Eyrolles Eds.
- BROCKE J. V. (2001). Freestyle Learning, Concepts, Platforms, and Applications for Individual Learning Scenarios. *46th International Scientific Colloquium, Ilmenau Technical University, Ilmenau, Allemagne, 2001*
- CHAMPIN P.-A., PRIÉ Y., MILLE A. (2003). MUSETTE: Modeling USEs and Tasks for Tracing Experience. *Workshop 'From Structured Cases to Unstructured Problem Solving Episodes For Experience-Based Assistance', ICCBR'03*, Trondheim, Norvège, p. 279-286.
- CHOQUET C., IKSAL S. (2007). 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*, Vol. 18 n°2, p. 161-184.
- CORBIÈRE A., CHOQUET C. (2004). Re-engineering Method for Multimedia System in Education. *IEEE Sixth International Symposium on Multimedia Software Engineering (MSE)*, Miami, USA, p. 80-87.
- COTTIER P., SCHMIDT C. (2005) Le dialogue en contexte : Pour une approche dialogique des environnements d'apprentissage collectif. *Revue d'intelligence artificielle*, Hermès, Vol. 19 n°1-2, p. 235-252.

- DPULS (2005). *Design Patterns for Recording and Analysing Usage of Learning Systems*. Kaleidoscope Project, <http://www.noe-kaleidoscope.org> (consulté le 5 mai 2007)
- DPULS (2005). *DPULS Design Patterns Browser*. Kaleidoscope Project, <http://lucke.univ-lemans.fr:8080/dpuls/login.faces> (consulté le 5 mai 2007)
- EGYED-ZSIGMOND E., MILLE A., PRIÉ Y. (2003). Club (Trèfle): a use trace model. *5th International Conference on Case-Based Reasoning*, Trondheim, Norvège, p. 146-160.
- EL KECHAÏ H., CHOQUET C. (2007). Reusing Pedagogical Scenarios at a Knowledge Level: a Model Driven Approach, *7th IEEE International Conference on Advanced Learning Technologies (ICALT'2007)*, Niigata, Japan, p. 670-674.
- GROB H. L., BENSBERG F., & DEWANTO B. L. (2004). Developing, Deploying, Using and Evaluating an Open Source Learning Management System. *Journal of Computing and Information Technology*, Zagreb, Croatia, University Computing Centre Eds., Vol. 12 n° 2, p. 127-134.
- IA (2005). *Interaction Analysis – Supporting participants in technology based learning activities*. Kaleidoscope Project, <http://www.noe-kaleidoscope.org> (consulté le 5 mai 2007)
- ICALTS (2004). *Interaction & Collaboration Analysis' supporting Teachers & Students' Self-regulation*. Kaleidoscope Project, <http://www.noe-kaleidoscope.org> (consulté le 5 mai 2007)
- IKSAL S., CHOQUET C. (2005). Usage Analysis Driven by Models in a Pedagogical Context, *Workshop "Usage Analysis in Learning Systems"*, AIED'05, Amsterdam, Pays-Bas, p. 49-56.
- IMS (2006). *IMS Global Learning Consortium*. <http://www.imsglobal.org/> (consulté le 5 mai 2007)
- IMS/LD (2003). *IMS Learning Design*. <http://www.imsglobal.org/learningdesign/index.html> (consulté le 5 mai 2007)
- KALEIDOSCOPE (2004). *Kaleidoscope Project*, <http://www.noe-kaleidoscope.org> (consulté le 5 mai 2007)
- KOPER R., OLIVIER B., & ANDERSON T. (2003). *IMS Learning Design Information Model (version 1.0)*. IMS Global Learning Consortium, Inc.
- LAFORCADE P., & CHOQUET C. (2006). Next Step for Educational Modeling Languages: The Model Driven Engineering and Reengineering Approach, *The 6th IEEE International Conference on Advanced Learning Technologies (ICALT'2006)*, Kerkrade, Pays-Bas, p. 745-747.
- LEJEUNE A., & PERNIN J-P. (2004). A Taxonomy for Scenario-based Engineering. *Cognition and Exploratory Learning in Digital Age (CELDA 2004)*, Lisbonne, Portugal, p.249-256.
- LOM (2002). *Draft Standard for Learning Object Metadata*, IEEE Inc.
- MOSTOW J. (2004). Some useful design tactics for mining ITS data. *Workshop on Analyzing Student-Tutor Interaction Logs to Improve Educational Outcomes, ITS'04*, Maceio, Brésil, p. 20-28.
- REFFAY C., & CHANIER T. (2003). How social network analysis can help to measure cohesion in collaborative distance-learning. *Computer Supported Collaborative Learning Conference (CSCL'2003)*, Bergen, Norvège, p. 343-352.
- SEEL N., & DIJKSTRA S. (1997). General Introduction. *Instructional Design : International Perspectives*. (vol. 2) (pp. 1-13). Hillsdale, NJ, Lawrence Erlbaum Associates.
- SETTOUTI L., PRIÉ Y., MARTY J.-C., MILLE A., (1997). Vers des Systèmes à Base de Traces modélisées pour les EIAH. *Rapport de recherche RR-LIRIS-2007-016*.
- TRAILS (2004). *Personalised and Collaborative Trails of Digital and Non-Digital Learning Objects*. Kaleidoscope project, <http://www.noe-kaleidoscope.org> (consulté le 5 mai 2007)
- ZHENG C., FAN L., HUAN L., YIN L., WEI-YING M., & LIU W. (2002) User Intention Modeling in Web Applications Using Data Mining. *World Wide Web: Internet and Web Information Systems*, Pays-Bas, Springer Eds., p. 181–191.

---

<sup>1</sup> Cet article est une version remaniée et réactualisée d'un article publié en anglais dans la revue JILR (Choquet et Iksal, 2007).

<sup>2</sup> UTL existe en deux versions. Cette version fut la première développée. La seconde version intègre à ce langage des fonctionnalités de capitalisation des savoir-faire en matière de modélisation et d'analyse de l'observation de l'utilisation d'un EIAH. Elle fait l'objet de la quatrième section de cet article.

---

## ■ A propos des auteurs

Christophe CHOQUET est Maître de Conférences en Informatique à l'Université du Maine. Membre du LIUM, il coordonne le projet REDiM (Réingénierie des EIAH Dirigée par les Modèles). Ses travaux actuels portent sur l'analyse des traces d'utilisation dans un EIAH et sur la modélisation collective de scénarios pédagogiques. Il est membre du réseau européen d'excellence Kaléidoscope au sein duquel il a animé le projet DPULS.

**Adresse :** Laboratoire d'Informatique de l'Université du Maine, IUT de Laval, Département Services et Réseaux de Communication, 52, rue des Drs Calmette et Guérin, F-53020 LAVAL Cedex 9

**Courriel :** [Christophe.Choquet@lium.univ-lemans.fr](mailto:Christophe.Choquet@lium.univ-lemans.fr)

**Toile :** <http://www-lium.univ-lemans.fr/~choquet/>

Après avoir fait une thèse dans le domaine du Web Adaptatif et du Web Sémantique, Sébastien IKSAL a intégré le LIUM en 2003 comme Maître de Conférences en Informatique dans le projet REDiM (Réingénierie des EIAH Dirigée par les Modèles). Ses travaux actuels portent sur l'analyse de traces et notamment sur le langage UTL (Usage Tracking language) qu'il propose. Il a collaboré aux projets européens CANDLE (1999-2002) et DPULS (2004-2005) et fait partie du réseau européen d'excellence Kaléidoscope.

**Adresse :** Laboratoire d'Informatique de l'Université du Maine, IUT de Laval, Département Services et Réseaux de Communication, 52, rue des Drs Calmette et Guérin, F-53020 LAVAL Cedex 9

**Courriel :** [Sebastien.Iksal@lium.univ-lemans.fr](mailto:Sebastien.Iksal@lium.univ-lemans.fr)

**Toile :** <http://www-lium.univ-lemans.fr/~iksal/>

---

### Référence de l'article :

Christophe CHOQUET, Sébastien IKSAL, Modélisation et construction de traces d'utilisation d'une activité d'apprentissage : une approche langage pour la réingénierie d'un EIAH, *Revue STICEF*, Volume 14, 2007, ISSN : 1764-7223, mis en ligne le 07/03/2008, <http://sticef.org>

© Revue Sciences et Technologies de l'Information et de la Communication pour l'Éducation et la Formation, 2007