HAL
open science

A propos d'un entrepôt de données universitaire : modélisation des acteurs et méta données

Frédérique Peguiron, Odile Thiery

To cite this version:

Frédérique Peguiron, Odile Thiery. A propos d'un entrepôt de données universitaire : modélisation des acteurs et méta données. [Interne] 2005. inria-00000357

HAL Id: inria-00000357

https://inria.hal.science/inria-00000357v1

Submitted on 27 Sep 2005

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.

A propos d'un entrepôt de données universitaire : modélisation des acteurs et méta données

PEGUIRON Frédérique
Frederique.Peguiron@loria.fr
Et
THIERY Odile
Odile.Thiery@loria.fr

LORIA, Campus Scientifique
B.P. 239
54506 VANDOEUVRE-lès-NANCY CEDEX

Une université abrite différents acteurs qui ont recours à des systèmes de ressources documentaires, des systèmes de production d'information, des systèmes de recherche d'information. L'intérêt que nous portons à la classification des acteurs de l'université, fondée sur notre modèle de représentation de l'utilisateur RU=(T, F, B, A), pour la construction d'un entrepôt de données (ED) et des bases métiers associées, permet de faire évoluer un Système d'Information (SI) en un Système d'Information Stratégique (SIS). Nous travaillons à la conception d'un entrepôt de données de ressources documentaires dans un cadre pédagogique intégrant la modélisation de l'utilisateur. La description de ressources, en vue de leur réutilisation dans des parcours de formation, évoquent les difficultés rencontrées et formulent des propositions pour combler des manques dans les normes existantes et rendre plus opérationnels certains descriptifs. La modélisation des acteurs d'une part et des types de documents d'autre part, permettent d'élaborer des corrélations afin d'améliorer les réponses. La mise en relation des acteurs et des documents est possible par les méta données de l'entrepôt de données et la méta modélisation de l'(ED).Nous élaborons également les méta données propres à l'(ED) qui définissent les méta données structurelles et d'accessibilité propres au système de pilotage. Afin de procéder au mieux au développement de notre contribution au (SIS), la méta modélisation de l'(ED) permet d'élaborer un schéma directeur pour la construction de l'(ED).

Mots-clés : Système d'information, système d'information stratégique, entrepôt de ressources documentaires, classification des utilisateurs, modèle de l'utilisateur, modèle de document électronique, méta données, méta modélisation.

1. Introduction

Une université abrite différents acteurs qui ont recours à des systèmes de ressources documentaires, des systèmes de production d'information, des systèmes de recherche d'information. Au sein de cette même université, cohabitent de nombreux Systèmes d'Information (SI) spécifiques aux besoins des composantes qui la constituent. Ces (SI) éparses abritent des informations qui peuvent être utiles aux composantes voisines. Le recours à l'entrepôt de données (ED) permet de proposer des solutions pour faire évoluer un Système d'Information en un Système d'Information Stratégique (SIS) voire en un Système d'Information Décisionnel (SID). Le transfert d'un (SI) en (SID) repose sur les bases métiers orientées vers les acteurs. Un Système d'Information Décisionnel (SID) est un système qui permet aux décideurs d'une institution de disposer d'informations pertinentes et d'outils d'analyse puissants pour les aider à prendre les bonnes décisions au bon moment. Le suivi d'une politique d'établissement peut se faire par le partage de tableaux de bord et autres indicateurs. D'après Inmon-William H.[20] «Le Data Warehouse est une collection de données orientées sujet, intégrées, non volatiles et historisées, organisées pour le support d'un processus d'aide à la décision». L'(ED) est entièrement construit selon une approche dimensionnelle, c'est-à-dire qui fait appel aux techniques qui favorisent l'analyse multidimensionnelle des données. Un entrepôt de données donne naissance, par filtrage thématique et fonctionnel, et en association avec des profils utilisateurs, à des bases métiers. Une base métier ou data mart est définie par R. Kimball [21] comme un sous-ensemble logique du data warehouse. Pour mieux répondre aux besoins des utilisateurs, on essaye de personnaliser les réponses du système, d'où la nécessité de représenter l'utilisateur et ses comportements dans les bases métiers afin de faciliter le processus de recherche d'information.
La modélisation est une étape préalable pour l'élaboration des méta-données. Le modèle issu de la modélisation reflète une image réelle du système. Si l'on prend le domaine des SI classiques, il existe plusieurs types de modélisations : Merise , OMT , UML etc. Nous utiliserons UML[23]] qui est le standard actuel dans le monde de la conception des (SI). Rappelons les enjeux : favoriser les apprentissages et améliorer le service aux usagers, rationaliser l'offre de services, valoriser les infrastructures, valoriser les productions, réutiliser les ressources, renforcer l'autonomie des étudiants, améliorer la visibilité des auteurs, faciliter la création, corréler la production des enseignants et les besoins des étudiants, inscrire le parcours et les compétences de l'étudiant afin de lui permettre de trouver un stage, un emploi ou de poursuivre ses études dans un autre établissement en France ou à l'étranger. Notre papier s'articule autour de la modélisation des acteurs, puis de la modélisation des documents. Les méta données et la méta modélisation de l'entrepôt de données favorisent la mise en relation du modèle des acteurs et du modèle des ressources documentaires tout en tenant compte du contexte des utilisateurs

1.1. L'existant

En analysant "l'Existant", des éléments importants ressortent et constituent le cadre de base pour élaborer une stratégie. L'agence de mutualisation des universités (AMUE)[2] mène une étude pour produire des cahiers des charges autour de ses futurs produits pour l'année 2006, en vue de l'élaboration fonctionnelle, organisationnelle et technique du système d'information de gestion d'un établissement. L'AMUE élabore un entrepôt de données à partir d'Harpège, Apogée et Nabuco afin d'aider au pilotage des universités. Sur un plan national, plusieurs consortiums d'universités se sont constitués pour proposer des espaces numériques de travail (EPPUN[1], Esup[2] …). Ces consortiums tiennent comptent des recommandations de l'AMUE de la constitution d'un SI sous forme

[1] Espaces pédagogiques pour les universités numériques
2 Environnement numérique de travail d'accès intégré aux services pour les étudiants et le personnel de l'enseignement supérieur

de briques. Quatorze universités adhèrent d'ores et déjà à Esup portail, dont l'Université Nancy 2. L'établissement d'enseignement est le périmètre de référence de l'espace numérique[3] de travail du point de vue de l'utilisateur. L'espace numérique de travail s'adresse ainsi à l'ensemble des usagers, étudiants, enseignants, personnels administratifs et techniques. Esup portail contient les briques assurant l'accès aux applications métiers mises en place dans les universités. La figure 1 permet de situer la n^{ème} brique au SI, emplacement où nous pouvons nous intégrer.

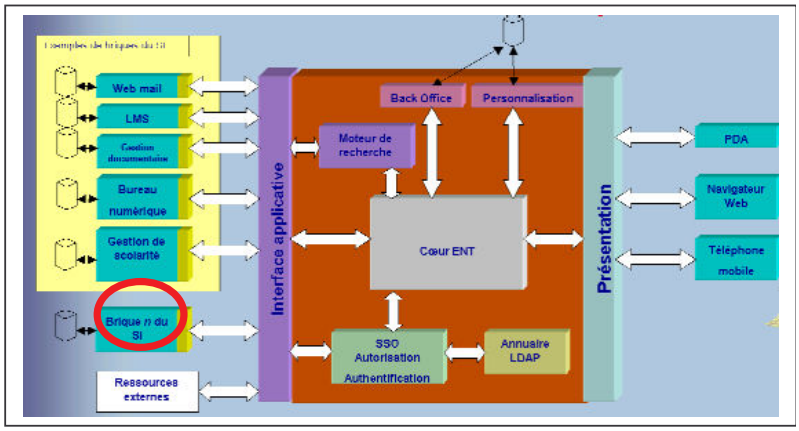

Figure 1 - Place de notre brique au sein du SI de l'université

1.2. Entrepôt de données

La modélisation de l'utilisateur intervient dans deux contextes de S-IS : pour la constitution des informations pertinentes et pour l'exploitation du S-IS. Dans le premier contexte, il s'agit d'utiliser le modèle de l'utilisateur comme un profil pour le filtrage d'information à intégrer dans le S-IS. Dans le deuxième contexte, il s'agit d'utiliser le modèle comme outil d'adaptation du comportement du système aux comportements de l'utilisateur. Pour illustrer notre brique au (SIS), la figure suivante permet de corréler les besoins des utilisateurs et les bases métiers.

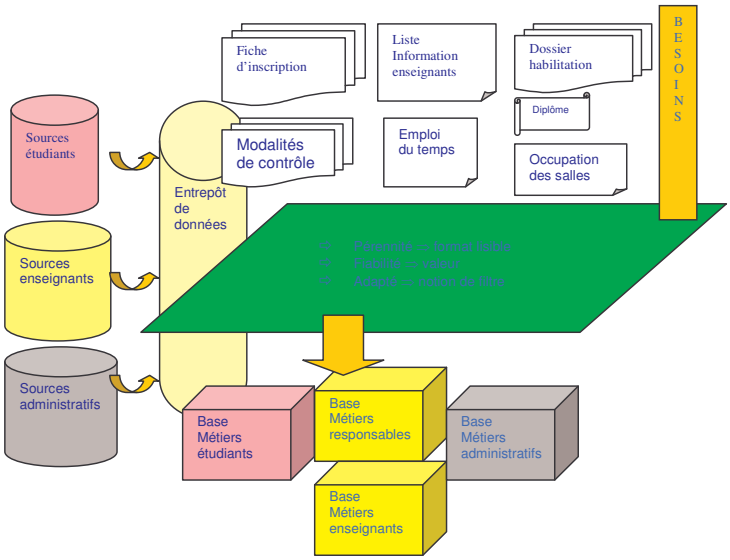

Figure 2 - Bases métiers adaptées aux utilisateurs finals

Notre approche de filtrage de l'information par le profil de l'utilisateur[14] est donc fondée sur la modélisation de l'utilisateur en intégrant des attributs d'identification de l'utilisateur, de son comportement et du contexte

3 Un espace numérique de travail désigne un dispositif global fournissant à un usager un point d'accès à travers les réseaux à l'ensemble des ressources et des services numériques en rapport avec son activité

d'utilisation des informations qu'il cherche. Le processus du filtrage de l'information lors de la constitution de l'entrepôt de données s'opère par un filtrage thématique et dans la constitution des bases métiers par un filtrage fonctionnel. Forte de ces constations variées, l'équipe SITE, qui réfléchit à l'amélioration des SI afin de satisfaire au mieux les utilisateurs finals (en situation de consultation et/ou de décision), dispose de données hétérogènes. Le recours à l'entrepôt de données permet de proposer des solutions pour faire évoluer un Système d'Information en un Système d'Information Stratégique (SIS) voire en un Système d'Information Décisionnel (SID). Le processus d'Intelligence Economique repose en particulier sur l'utilisation de Systèmes d'Informations Stratégiques (SIS). Voyons ce que peut être le SIS d'une université. Dans notre cas, il s'agit de mettre à disposition des décideurs de l'université des informations synthétiques autour d'indicateurs choisis par eux, pour leur permettre de réaliser des tableaux de bord, afin de procéder à des constats, des suivis d'opérations et de prévisions ou pour mettre en évidence des causes de certains faits.

1.3. Analyse multidimensionnelle

Quelles informations contiennent un tableau de bord ? Des indicateurs qui sont des nombres caractérisés par des entités présentées en ligne ou en colonne. Les indicateurs sont généralement placés au croisement d'une ligne et d'une colonne, représentant le croisement de deux entités. Un indicateur est un nombre, une somme, un minimum, ou un maximum, donc toujours un nombre. Cela correspond bien à nos cases de tableau de bord.

UFR	enseignant du second degré	professeur	non enseignant
Mathématiques et Informatique	Nbre(X)	Nbre(Y)	Nbre(Z)

Figure 3 - Exemple d'un tableau de bord

Une solution consiste à utiliser la modélisation dimensionnelle, qui dérive des concepts des bases de données multidimensionnelles (dites OLAP). L'avantage de ce type de modélisation est avant tout d'être indépendant du type de technologie utilisé. Le but de cette méthode est de faire ressortir des besoins en indicateurs de performance sur les données et ainsi de permettre de définir des dimensions (exemple le temps ou le découpage géographique). On voit apparaître plusieurs schémas pour illustrer cette méthode :

Schéma en étoile

Le schéma en étoile représente une table de faits connectée à un ensemble de tables de dimensions. **Exemple :** Cette déclinaison nous permet de dessiner la granularité des dimensions : temps, acteur (étudiant), géographie.

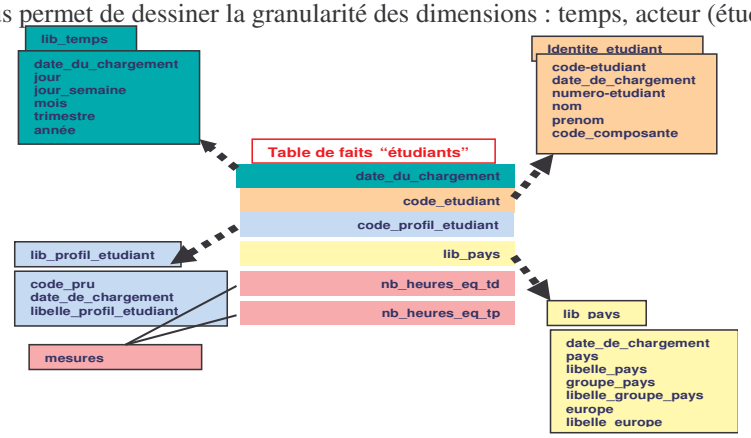

Figure 4 - Schéma en étoile

Schéma en flocon de neige

Le schéma en flocon de neige est un raffinement du schéma en étoile où certaines tables de dimensions sont normalisées (donc décomposées). **Exemple :** L'un des éléments (composante) de la table acteur (étudiant) est décomposé.

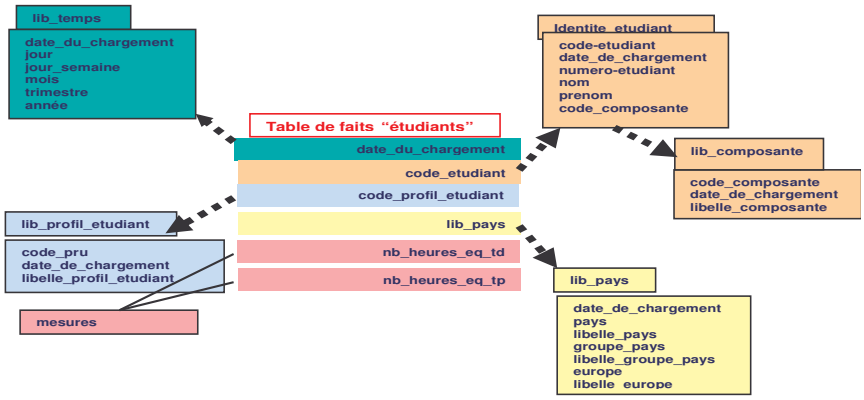

Figure 5 - Schéma en flocon de neige

Constellation de faits

La constellation de faits permet de représenter plusieurs tables de faits partageant quelques tables de dimension. **Exemples :** Les tables de faits (étudiants) et (enseignants) peuvent être mises en relation par la dimension temps et géographie.

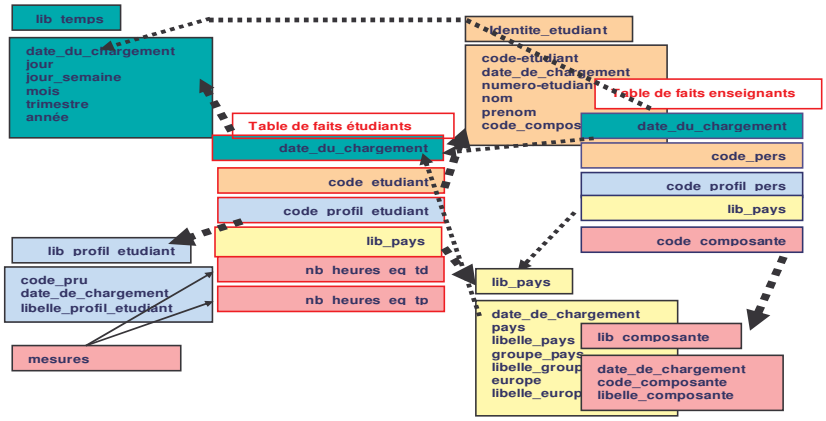

Figure 6 - Constellation de faits

L'OLAP (Online Analytical Processing) est une technique informatique d'analyse multidimensionnelle, qui permet aux décideurs, d'avoir accès rapidement et de manière interactive à une information pertinente présentée sous des angles divers et multiples, selon leurs besoins particuliers. Ce terme englobe à la fois les architectures MOLAP[4] et ROLAP[5]. A titre d'exemple on peut représenter de façon graphique des informations contenues dans une base de données, sous la forme d'un cube à plusieurs dimensions, lequel cube permet d'analyser ces données sous différents angles, grâce à l'organisation de celles-ci en axes d'analyses et en variables à analyser.

[4] MOLAP (Multidimensional On Line Analytical Processing) : Traitement analytique multidimensionnel en ligne. Dans l'univers du datawarehousing, technique de modélisation et de stockage des données consistant notamment à distinguer les faits à analyser (indicateurs, métriques, ...) et les différents axes d'analyse encore appelés "dimensions". Ces structures multidimensionnelles sont appelées également hyper cubes ("cubes" à plusieurs dimensions)

[5] ROLAP (Relational On Line Analytical Processing) : Traitement analytique relationnel en ligne. Dans l'univers du datawarehousing, technique de modélisation et de stockage des données fondée sur une structure relationnelle.

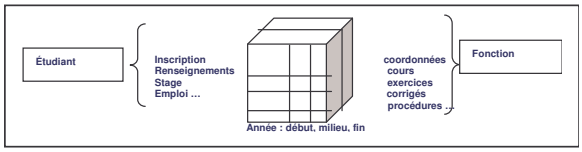

Figure 7 - Cube de données

Le but est bien de répondre au mieux aux utilisateurs. Avant de pouvoir personnaliser les réponses du système, il est nécessaire de connaître les besoins des utilisateurs, leurs rôles et leurs comportements. Dans notre optique, le passage d'un SI à un SIS ou SID se fait par l'intégration des comportements des utilisateurs.

Rappelons que l'équipe SITE [11] travaille à la modélisation du veilleur ou infomédiaire et du décideur. Le veilleur, spécialiste de l'information joue un rôle d'intermédiaire entre le décideur et l'information demandée. Il va exploiter les connaissances dont il dispose sur le décideur en traduisant le problème du décideur en indicateurs. Il aide aussi le décideur à découvrir les paramètres et à les vérifier. Le décideur est défini comme celui qui est apte à identifier et à poser le problème à résoudre en terme d'enjeu, de risque ou de menace qui pèse sur l'institution. Il utilise les indicateurs délivrés par l'infomédiaire pour atteindre son objectif. Il extrait également des indicateurs, les informations nécessaires à la prise de décision. Ces modèles aboutissent à l'élaboration de méta-données.

2. Modélisation de l'acteur

Nous avons pu voir au cours de nos recherches que les logiciels administratifs au sein des universités, ont des limites au niveau du pilotage. Apogée[6] permet d'avoir la répartition des étudiants par diplôme en croisant les instances de l'objet «individu» et les instances de l'objet «diplôme». Toutefois, des requêtes de pilotage pertinentes ne peuvent pas être définies dynamiquement au fur et à mesure des besoins dans Apogée. Lors de la mise en route du processus de modélisation des acteurs : étudiants, enseignants, administratifs nous avons mis en évidence que nous pouvons récupérer des données au niveau de ce logiciel afin de faire le lien entre enseignant, enseigné et enseignement. Nous évoluons dans un contexte universitaire où il s'agit de mettre en relation information et acteurs du SI–SIS. Nous allons voir comment la modélisation des acteurs permet d'améliorer les réponses.

2.1. Modélisation des acteurs

Nous dressons une classification[33] des acteurs où nous disons qu'un utilisateur (**U**) est représenté par un type (**T**), besoins (**B**), fonctions (**F**), et activités (**A**) selon ce modèle :

RU = (T, B, F, A). Le schéma directeur des espaces numériques de travail[40] précise les usagers amenés à intervenir dans cet environnement. Ceci permet de mentionner les utilisateurs concernés par notre réflexion. Dans les paragraphes suivants, nous développons les items T : type, F : fonctions, B : besoins et A : activités des acteurs par rapport au SI.

<u>Type</u>

L'item **T** représente le type d'acteur pour l'enseignement supérieur que nous résumons de la façon suivante : **T** {Etudiants, Chercheurs, Enseignants, Responsables, Personnels, Partenaires, Administrateurs}. Nous utilisons le formalisme UML pour modéliser les types d'acteurs. Empruntons à UML les diagrammes de classe et d'objets pour nous permettre de recenser des objets et des classes dans les données fournies sur les maquettes de cours

[6] Apogée, Application pour l'organisation et la gestion des étudiants, apporte des réponses précises en matière de clarification de l'offre de formation, d'amélioration de l'accueil des étudiants, de gestion de la scolarité et de pilotage de l'établissement.

tirées d'Apogée. L'étape de modélisation de classes d'objets permet de faire apparaître des attributs et des valeurs selon le schéma suivant :

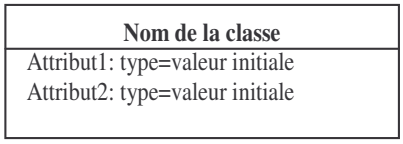

Figure 8 - Formalisme de représentation

La modélisation des types d'acteurs sur la figure 9, permet de dresser des catégories d'acteurs et des sous-catégories d'acteurs.

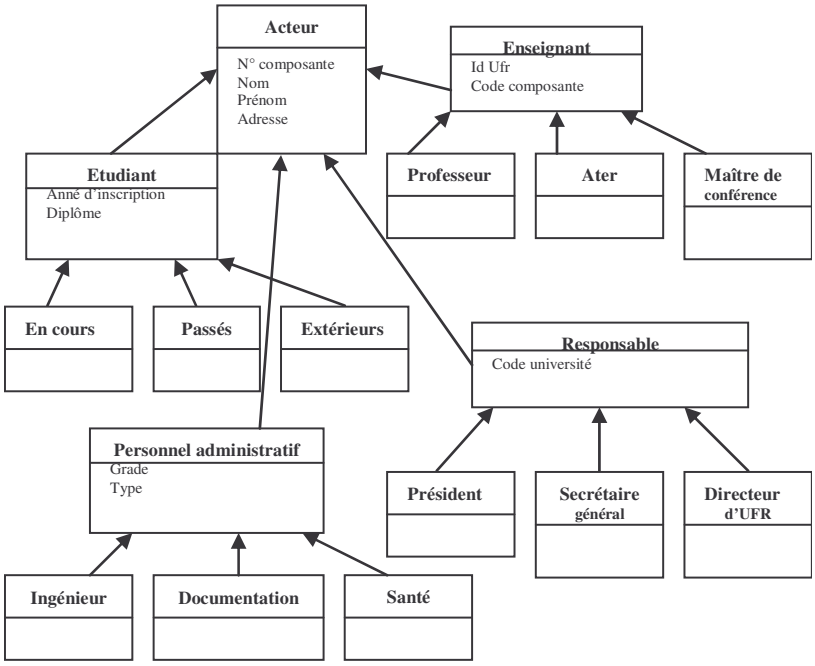

Figure 9 - Modélisation des types d'acteurs

Concentrons nous sur l'acteur «étudiant» de la figure 9. D'après[40] l'univers étudiant est représenté par les classes suivantes : **individu**, inscription pedagog. etape, inscription admin. annuelle, inscription admin. etape resultats aux elements, resultats aux etapes, **resultats aux diplomes**, stages, troisieme cycle, couverture sociale, cursus pre-universitaire, **diplome**, diplome autre cursus, **cursus**, blocage, indicateurs. Chaque classe comporte des objets. Notre but est d'améliorer l'information proposée à l'utilisateur. Nos trois classes d'objets sélectionnées : **diplomes, individu**, **cursus** doivent nous permettre, par exemple d'aider l'acteur étudiant à améliorer ses connaissances autour d'une discipline.

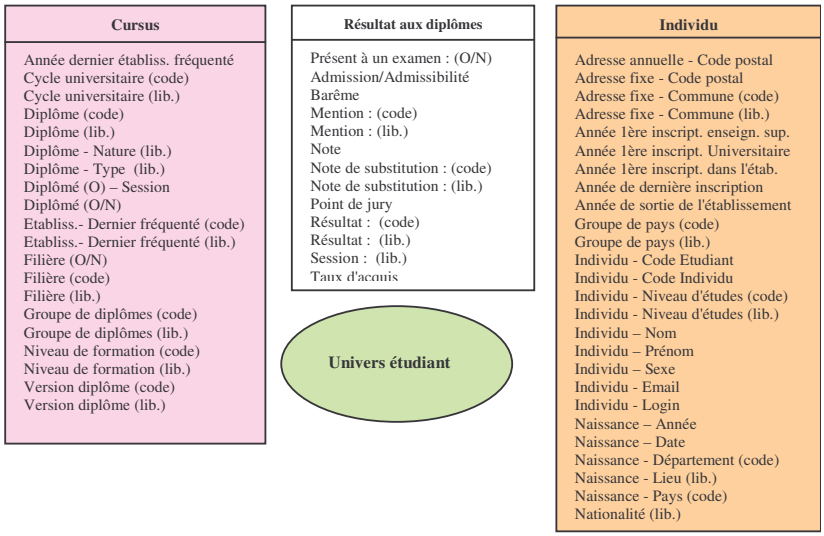

Figure 10 - Exemple de 3 classes d'objet modélisant l'acteur étudiant

Pour l'acteur étudiant, choisissons la classe **individu**, qui permet de tirer parti des données existant dans les champs ou dans les vues partielles, en vue de les réutiliser pour lui permettre d'améliorer ses recherches.

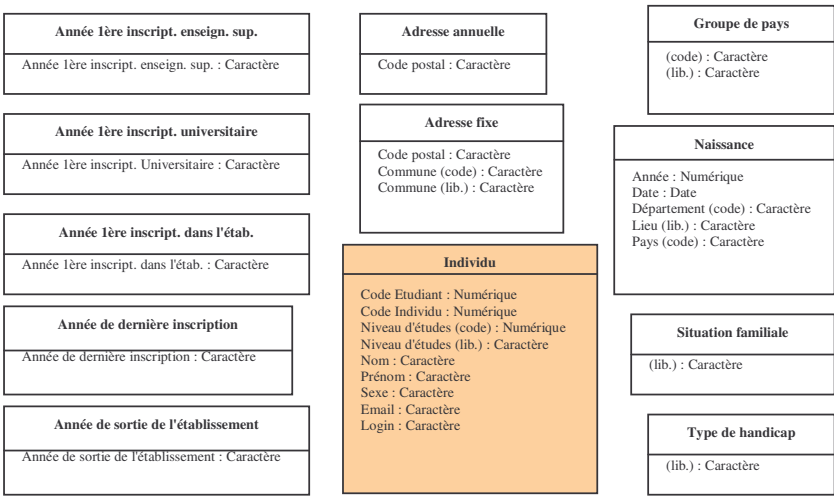

Figure 11 - Développement de la classe individu pour l'acteur étudiant

La figure 11 fait apparaître des données propres à l'identification de l'étudiant, sa localisation géographique, son niveau d'étude, son login, son historisation d'inscription dans l'établissement. Avant de poursuivre sur la modélisation d'une composante, présentons et développons la classe **diplôme** formée d'U.E. (unités d'enseignement) :

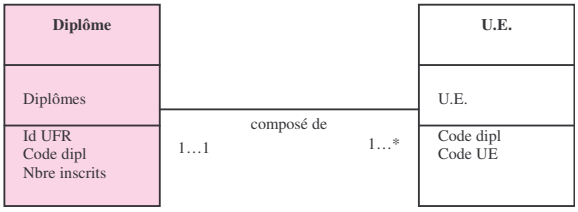

Figure 12 - Classe diplôme

En développant la classe cursus de la figure 10, les objets autour du diplôme nous renseignent sur les modules. Auparavant, analysons une composante pour observer les diplômes proposés. Voici les diplômes proposés par la composante UFR Mathématiques Informatique de Nancy 2 pour l'année 2002.

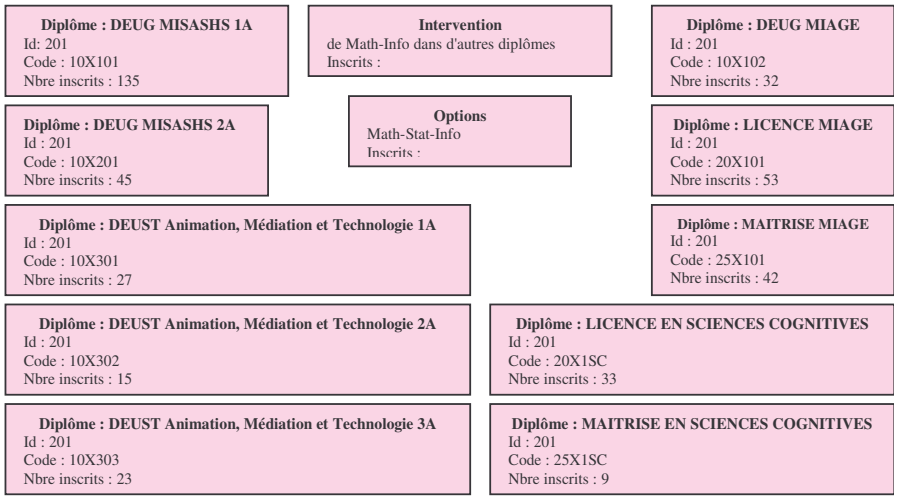

Figure 13 - Modélisation des interventions et diplômes de la composante UFR Math de Nancy 2

Pour développer le contenu des U.E, choisissons le diplôme DEUG MISASHS Info 1A de la composante UFR Math de Nancy 2,.

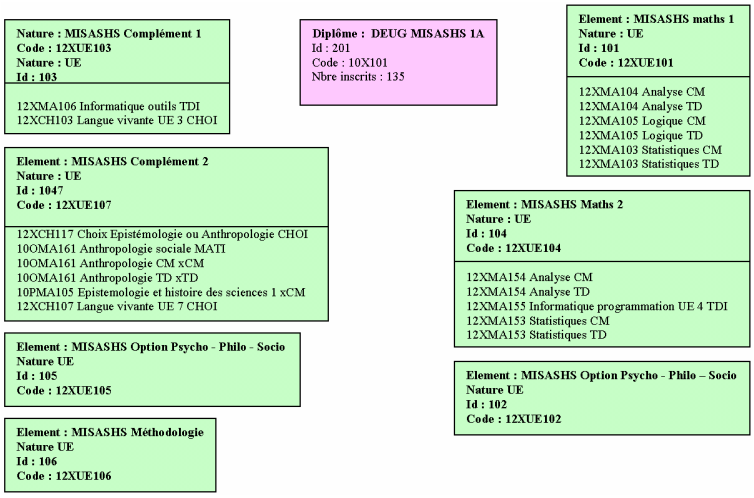

Figure 14 - Modélisation d'un diplôme DEUG MISASHS 1A

La modélisation des types d'acteurs a permis de modéliser l'acteur étudiant, de développer les classes individu et diplôme. Les attributs et les valeurs donnent des données chiffrées comme, le nombre d'étudiants pour une discipline. Ils permettent de mettre en valeur des notions en procédant par des calculs par association : comme cerner les compétences d'un étudiant autour d'un domaine ou de calculer le nombre de redoublants.

Besoins

L'enseignant [32] est placé à la fois dans un environnement d'innovation et aussi dans un environnement réglementé. Dans un contexte universitaire, l'enseignant a besoin de répondre aux besoins des étudiants. On renseigne un certain nombre d'éléments de l'item besoins pour les acteurs étudiant et enseignant. Cet item peut s'enrichir au cours du temps : $\mathbf{B}_{\text{étudiant}}$ {s'inscrire, s'exercer, se former, rechercher emploi, rechercher stage} ; $\mathbf{B}_{\text{enseignant}}$ {exercer, former, corriger, recenser, évaluer, budgétiser, déployer, planifier, se conformer aux textes officiels}.

Fonctions

Les enseignants et les étudiants ont des fonctions qui leur confèrent un rôle. L'enseignant est amené à distribuer des devoirs, décrire des tâches à effectuer, évaluer, réceptionner des travaux. Nous spécifions un certain nombre

d'éléments de l'item fonctions pour les étudiants et les enseignants : $\mathbf{F}_{étudiant}$ {apprendre, créer, intégrer, vérifier} ; $\mathbf{F}_{enseignant}$ {créer, enseigner, diriger, missionner, organiser, gérer, conseiller, superviser}. Cette liste n'est pas close.

Activités

Les activités des acteurs concernent leurs activités lors de l'utilisation du système d'information. Ils peuvent être amenés à rechercher, télécharger des dossiers, comprimer des fichiers, annoter des images ou du texte, indexer des documents, consulter des notes. Ils doivent parfois disposer d'un certain degré d'interactivité avec le système. Nous résumons les activités des étudiants et des enseignants dans l'item activités: $\mathbf{A}_{étudiant}$ {déposer, explorer, interroger, analyser, synthétiser, annoter} ; $\mathbf{A}_{enseignant}${déposer, indexer, diffuser, explorer, interroger, analyser, synthétiser, annoter}.

3. Modélisation des documents

Dans notre étude, le SIS universitaire repose sur un Entrepôt de Ressources Documentaires. Nous allons tenter de répondre à la question : comment construire le SID d'un environnement pédagogique ? Cette construction passe par une phase de collecte des connaissances au profit de conceptions pédagogiques. Pour illustrer nos propos, nous poursuivons la modélisation des documents en utilisant toujours le scénario de mise en correspondance, enseignant et enseigné, par l'intermédiaire de média d'enseignement. La conception de ressources documentaires au service de formations peut être de plusieurs types, dont nous décrirons deux aspects.

Transposition de la conception d'un document sur papier

La conception d'une formation peut être une transposition un peu améliorée d'un document papier en un document électronique, dans ce cas on aboutit à un parcours linéaire. Cette solution est assez répandue car elle a l'avantage d'être relativement économique, mais, en contre partie, elle n'apporte pas de réponses aux différents besoins de l'élève au sein d'une même formation.

Processus plus élaboré de gestion de contenu

Il s'agit de séparer des connaissances au sein d'un cours. Le recours au XML, permet de séparer le contenu du contenant. Il existe une base de connaissances constituée de modules. Selon un mécanisme d'assemblage, il est possible de répondre aux différents besoins d'étudiants d'UFR différentes.

3.1. Normes et standards

Le recours à des normes et des standards pour la conception d'un (ED) de ressources documentaires dans un cadre pédagogique, intégrant la modélisation de l'utilisateur assurent l'interopérabilité des composants et la réutilisation de ressources. La normalisation n'est pas systématiquement synonyme de restrictions ou de contraintes fortes. La norme fournit un cadre dans lequel il est possible de trouver de multiples moyens d'expression. Cependant tous les cas de figure ne sont pas exprimables par le recours à une norme. L'équipe SITE a développé deux modèles propres aux acteurs : MEPD[7] et WISP[8]. Parallèlement à ces deux modèles, nous trouvons un panel de normes autour des documents électroniques et de ressources en ligne. Si on s'attarde sur le processus d'apprenant lors du parcours de l'étudiant, on s'aperçoit que la modélisation du système d'information permet de mettre en relation des objets propres à différentes classes. D'après les figures 10 et 11, les classes individu, réussite aux diplômes, stage et ressources d'information permettent : de proposer des ressources par

[7] MEDP : Modèle pour l'explicitation d'un problème décisionnel
[8] WISP : Watcher-Information-Search-Problem

rapport aux diplômes suivis, d'historiser le parcours en mettant en valeur les différentes étapes, de catégoriser les acquis comme compétence, notamment en procédant à des calculs par association des attributs et des valeurs.

Sachant qu'il n'existe pas de normes préétablies pour un Entrepôt de Données Documentaires (EDD) et que l'élaboration d'un standard est un long processus qui s'étale sur plusieurs années, nous analysons les normes existantes. Nous devons identifier les classes des ressources documentaires de notre (EDD). Une fois ces classes d'objets définies, nous pouvons avoir recours aux normes existantes, qu'il faut enrichir selon les besoins spécifiques à notre SI–SIS. Nous sommes à un stade de réflexion où à la fois il faut décrire les utilisateurs de l'(ED) et à la fois décrire les ressources d'information. Le but étant de favoriser la mise à disposition de l'utilisateur final, les ressources d'information les plus adaptées.

3.2. Modélisation des ressources

Cette étape permet d'énoncer les principales caractéristiques d'une ressource selon différentes facettes : technique, pédagogique, utilisation, référencement, droits, relations. Nous disposons de normes de description françaises, européennes et internationales. Ces normes prennent en compte les concepts propres à leurs contours géographiques. Il est donc quasiment impossible de recourir à leurs recommandations et de les calquer à nos critères. Néanmoins, nous évoluons dans un contexte où il y a souci d'harmonisation sur un plan européen, cela se concrétise par la mise en place du LMD[9]. Dans un contexte mondial, on favorise les échanges internationaux pour les formations, les stages et les emplois.

Typologie des ressources documentaires

Le SI de l'université propose différents types de documents à destination de types d'acteurs différents, décrits précédemment en paragraphe § 2.1. On recense entre autres : des documents administratifs, des cours, des plaquettes, des images, des vidéos, du son, des catalogues de bibliothèque, des bases de données, des livres électroniques, des journaux électroniques. Tous ces éléments sont au service des conceptions de formation, pour les enseignants et les étudiants en situation de recherche ou de création d'information.

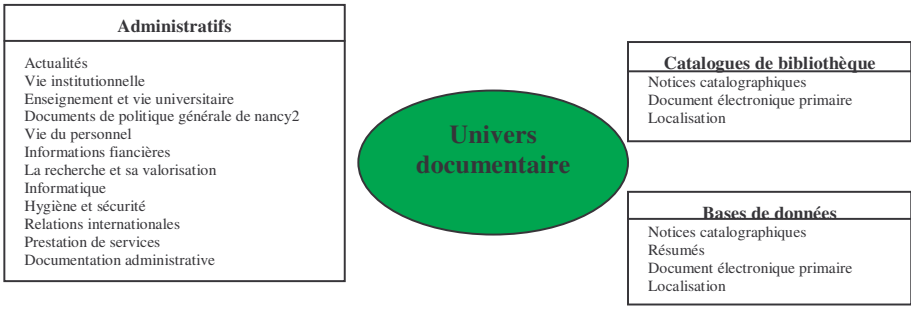

Figure 15 - Modélisation du type des documents

L'objectif de la modélisation, puis de la description est de favoriser la visibilité d'un patrimoine pédagogique tout en préservant une expression simple des informations pouvant répondre aux questions suivantes : quelles sont les caractéristiques de la ressource ? Comment est gérée la propriété intellectuelle ? Comment classer cette ressource ? Comment mettre en relation le profil de la ressource et le profil utilisateur ?

[9] LMD : Licence-Master-Doctorat

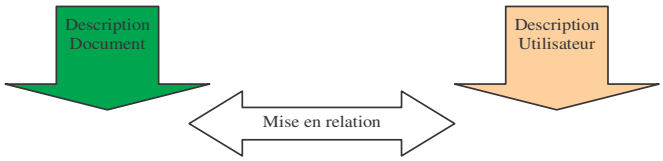

Figure 16 - Filtrage d'information pour les ressources documentaires par mise en correspondance du profil document et du profil utilisateur

Description des ressources documentaires

1er cas de figure :

Ces types de documents sont gérés par des systèmes qui leur sont propres, accessibles au travers de leur propre moteur de recherche ou par des moteurs de recherche fédérés. En ce qui concerne les catalogues de bibliothèque, les bases de données, les livres électroniques, les journaux électroniques le format utilisé pour la description des notices tend vers le format MARC, avec des sous-parents : LC-MARC, UNIMARC, US-MARC. En ce qui concerne la classification et l'indexation, on s'aperçoit qu'il est impossible d'avoir recours à un système unique. Quant à l'indexation, on voit les limites du recours à un thésaurus unique. Ces systèmes de classification et d'indexation ne suffisent pas à tous les impératifs de description pour les ressources documentaires.

2è cas de figure :

Les documents sont déposés sur le SI sans classification, sans indexation, sans aucune information sur leur contenu, leur contenant, leur but, leur impératif technique. Les pratiques mettent en évidence que peu d'utilisateurs remplissent les propriétés du document qu'ils conçoivent et déposent sur un SI. D'où la difficulté de récupérer des zones vides.

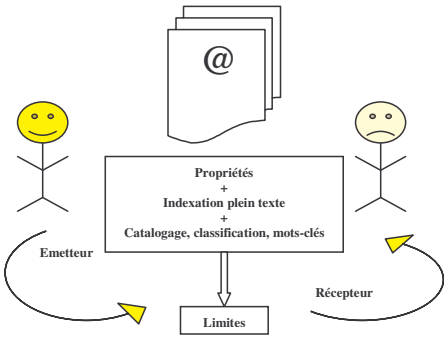

Figure 17 - Limites des descriptions des ressources documentaires

Fort de ces préambules, il s'agit bien pour nous de faire rencontrer utilisateur et documentation. Corréler acteurs et ressources passent par l'analyse des normes et standards en vigueur dans les universités. On peut distinguer les normes propres aux documents de tous types et les normes propres aux documents pédagogiques.

Documents de tous types :

Ce souci d'indexation de documents électroniques a pris toute son ampleur avec l'indexation des pages web par les moteurs de recherche. La Dublin Core, la RDF[10] ou XMP[11] sont des réponses pour ajouter des méta-données[12], prises en compte lors de l'indexation et sensée améliorer les réponses du système par rapport aux mots-clés de la question posée. Les normes IPTC[13], EXIF[14], DIG35[15], JPEG[16] ont été plus particulièrement

[10] RDF : Resources Description Framework
[11] XMP : Extensible Metadata Platform
[12] Une métadonnée est une "donnée sur des données"
[13] IPTC : International Press Telecommunications Council

développées pour les images. MPEG-7[17] concerne la description des objets multimédia. RKMS[18] aide à décrire les ressources audio. PRISM[19], NewsML[20], NITF[21] concernent la presse. Les ressources décrites sont très variées : monographies, publications en série, articles, archives, pièces de musée, images, séquences audio ou vidéo, des textes, graphiques, photos, séquences audio, vidéo et animations.

Ressources pédagogiques :

Les normes LOM[22], EML[23] et Scorm[24] apportent de nombreux éléments de réponses sur plusieurs points : le domaine, le matériel, l'interopérabilité des plates-formes, le type de médias, l'interface homme – machine, la description des contenus, l'architecture du système, les technologies collaboratives, le vocabulaire, les informations sur le participant, la description des compétences, la propriété intellectuelle, la qualité. Nous voyons qu'au même titre que nos acteurs, ces domaines constituent des classes d'objet. Cette mise en relation est réalisable, grâce aux méta données de l'entrepôt de données. Nous allons à présent définir et décrire ce processus.

4. Métadonnées et méta modélisation de l'entrepôt de données

Il est bien convenu maintenant que la structuration systémique [24] d'une organisation peut se voir sous la forme de trois sous systèmes : le système de pilotage, le système d'information et le système opérant. Piloter : c'est définir, déclarer, choisir les informations dont on a besoin pour obtenir une vue de l'état du système modélisé. Nous avons pu mettre en relief que la prise de décision revient à l'utilisateur final qui évolue dans un système de recherche d'information. L'intérêt que nous avons porté à la classification des acteurs de l'université fondée sur leurs activités pour la construction d'un entrepôt de données et des bases métiers associées, nous permet de faire évoluer un Système d'Information (SI) en un Système d'Information Stratégique (SIS).

4.1. Qu'est ce qu'une métadonnée

Les métas données (M.D.) sont des informations sur les données indispensables à une exploitation efficace d'un data warehouse (DW). Les méta données décrivent le schéma d'un (DW) et les données individuelles. Dans l'entreposage de données, les M.D. sont classées d'après leur objet et le public auquel elles s'adressent. Notre préoccupation principale est donc la prise en compte de l'acteur dans la construction puis l'exploitation d'E.D. C'est elle qui permet d'être garant de la cohérence fonctionnelle et technique et de la qualité des informations fournies aux utilisateurs métier par le biais de l'E.D. Les M.D. décrivent l'ensemble des règles, des définitions, des transformations et des processus associés à une donnée. On est capable, dès lors, d'avoir une vision unifiée des données, transverse à tous les métiers de l'organisation. Les principales informations sont destinées :

- A l'utilisateur (informations sur la sémantique des données utilisées et leur localisation dans l'entrepôt de données),

[14] EXIF : EXchangeable Image File
[15] DIG : Digital Imaging Group
[16] JPEG : Join Photographic Experts Group
[17] MPEG : Moving Pictures and associated audio information coding Experts Group
[18] RKMS : Recordkeeping Metadata Schema
[19] PRISM : Publishing Requirements for Industry Standard Metadata
[20] NewsML : News Markup Language
[21] NITF : News Industry Text Format
[22] LOM : Learning Object Metadata
[23] EML : Educational Modelling Languages
[24] Scorm : Sharable Content Object Reference Metadata

- Aux équipes responsables des processus de transformation de données de l'environnement de production vers l'entrepôt de données (informations sur la localisation de la donnée dans les systèmes de production, sur la description des règles et des processus de transformation),
- Aux équipes responsables des processus de création des données agrégées à partir des données détaillées,
- Aux équipes d'administration de la base de données,
- Aux équipes de production (informations sur les procédures de chargement, historiques des mises à jour, etc..).

4.2. Typologie des métadonnées

La description des M.D.[44] permet d'éviter la perte de données essentielles et d'inscrire les méthodes de collecte et de traitement des données. Cela favorise l'utilisation ultérieure des données. La figure 10 contextualise les métas données dans le système de pilotage. On remarque deux types de métas données :

- structurelles : elles décrivent la structure et le contenu du DW (méta schéma)
- accessibilité : elles représentent le lien dynamique entre l'entrepôt et les utilisateurs (référentiel de l'entrepôt ou DW ou Repository)

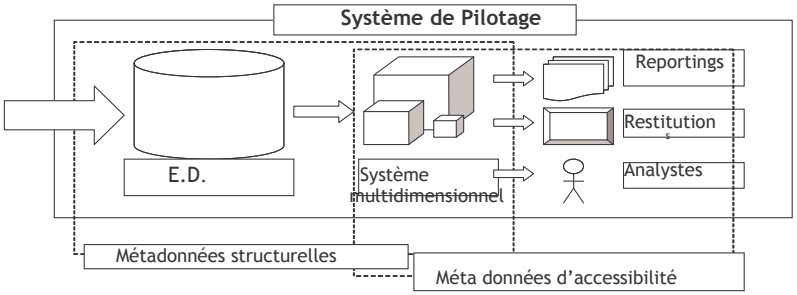

Figure 18 - Métas données structurelles et d'accessibilité

5. Notre domaine d'application : mise en relation du modèle des acteurs et du modèle des ressources documentaires

Notre domaine d'application [34] met en situation un enseignant qui doit concevoir une même formation à deux publics d'étudiants appartenant à des UFR différentes et un étudiant amené à améliorer ses compétences autour de modules. L'utilisateur, dont le profil est décrit dans des méta-données, obtient des informations personnalisées. Ces informations peuvent concerner des ressources documentaires ou des informations économiques et financières. En fonction de son rôle, l'utilisateur pourra visualiser des niveaux d'information. Nous mettons met en évidence que les étudiants et les enseignants ont des besoins, des fonctions et des activités spécifiques[33].

5.1. Prise en compte du contexte des utilisateurs

L'étude des normes en vigueur des ressources documentaires, permet d'élaborer des classes d'objet. Nous répartissons ces classes d'objets autour de quatre pôles qui prennent en compte le contexte de l'utilisateur :

- Propriétés de la ressource : fonction pédagogique, format, forme, phase de conception, identifiant
- Identification des besoins des utilisateurs auxquels répond un document : ressource, objet à définir, public concerné, principes d'apprentissage, type
- Données de description relatives au contexte : scénarios pédagogiques, titre, mots-clés
- Ressources documentaires autour d'un module : module, activités, experts, densité sémantique, couverture

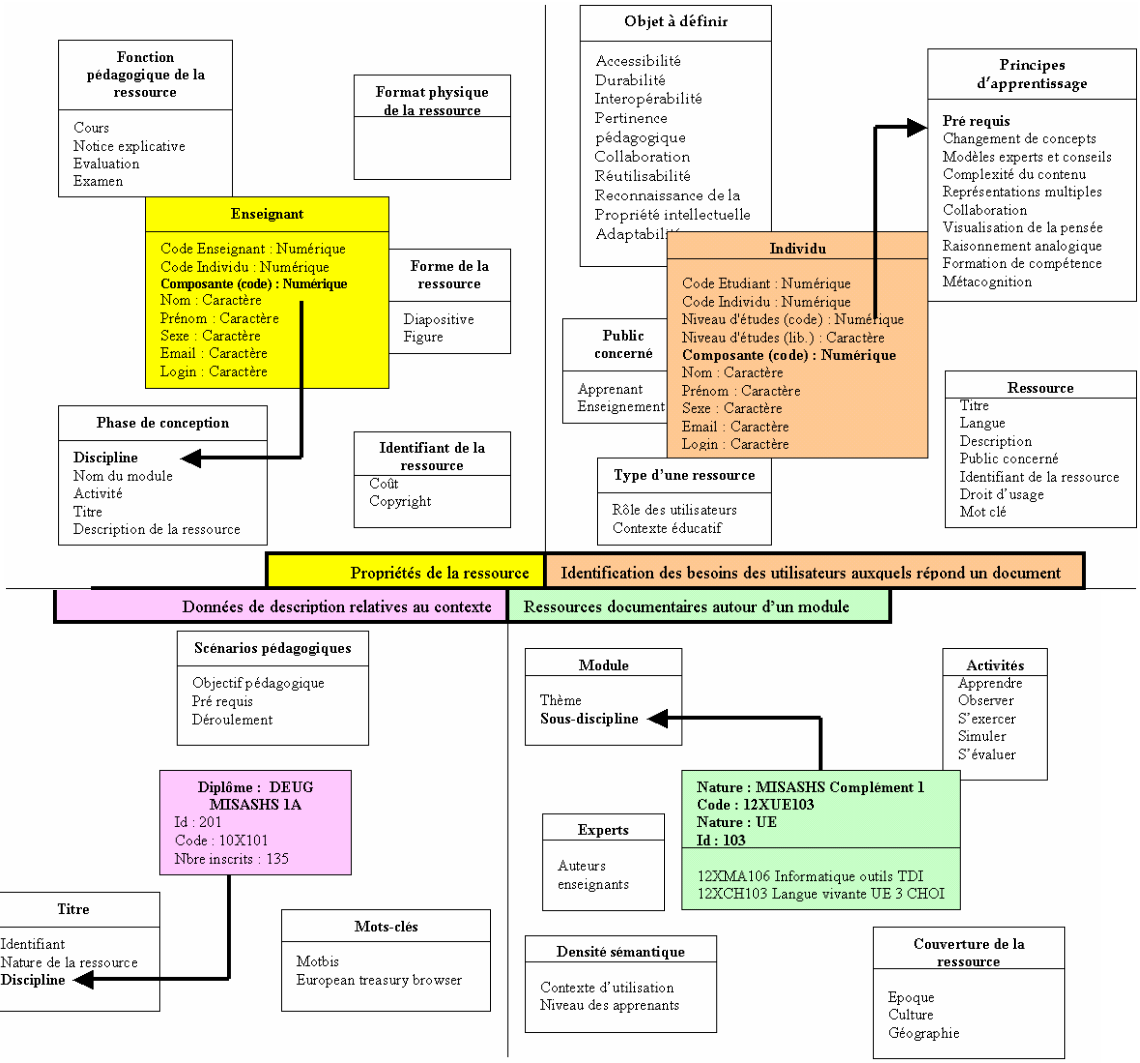

Figure 19 - Prise en compte du contexte des utilisateurs pour la conception de l'ERD

Au niveau du pôle «propriétés de la ressource», nous ajoutons la table de faits «enseignant». Nous ajoutons la table de faits des étudiants «individu» au pôle «identification des besoins des utilisateurs auxquels répond un document». Nous ajoutons la classe «diplôme» au pôle «données de description relatives au contexte». L'élément «unité d'enseignement» est placé dans le pôle «ressources documentaires autour d'un module». La figure 20 illustre la mise en relation des classes d'objets propres aux acteurs et des classes d'objets propres aux documents. Notre exemple permet de mettre en relation l'enseignant créateur d'une ressource à l'attention de l'étudiant. La table de faits «enseignant» renseigne sur la composante de l'enseignant ainsi que sur la discipline

concernée par le document. On retrouve la composante dans la table de faits «étudiant». La table de faits «étudiant» indique le niveau d'étude de l'apprenant. L'élément niveau d'étude de l'étudiant peut être mis en correspondance avec les pré requis nécessaires de l'apprenant, pour aborder une ressource documentaire. La description de la *discipline* du document, permet une mise en relation avec le *diplôme* poursuivi de l'étudiant. La description de l'*unité d'enseignement* favorise une mise en relation avec les ressources documentaires autour de *modules* spécifiques.

6. Conclusion

Par ce papier, nous avons montré comment il est possible d'exploiter les bases de données existantes afin de concevoir un entrepôt de données qui intègre dès la conception les besoins des utilisateurs finals. Les bases métiers, orientées acteurs, permettent de suggérer des prises de décision. La modélisation fine des différents acteurs permettent d'élaborer les méta données des acteurs pour intégrer leur type, leurs besoins, leurs fonctions et leurs activités. Nous avons montré comment il est possible d'améliorer la représentation des utilisateurs pour la fabrication des bases métiers et aboutir à un modèle formel, nommé RUBI3, qui signifie {Représentation des Utilisateurs et de leurs Besoins en Information lors de l'Interrogation après Identification}. Nous mettons en évidence que ce modèle RUBI3, permet de donner des vues différentes du SIS aux différents acteurs. Le processus d'élaboration des méta-données propres à l'entrepôt de données, définissent les méta-données structurelles et d'accessibilité propres au système de pilotage. Cela revient dans un S-IS à stocker parmi les méta-données du système, une représentation explicite de la structure des différentes bases métiers. La métamodélisation nous permet de tracer l'évolutivité du schéma de l'entrepôt de données. L'aspect dynamique du schéma est dû au fait que le contenu de l'entrepôt est guidé par les besoins des utilisateurs. Il est donc nécessaire de prévoir l'évolutivité de l'entrepôt de données dès sa conception. Cette étape a favorisé la mise en relation des acteurs et des ressources documentaires. Nous avons montré qu'un entrepôt de données favorise l'accès à des informations stratégiques : ces informations pouvant être de type financier mais également de type administratif, pédagogique ou documentaire. Nos perspectives concernent la problématique du référentiel de méta données en tenant compte des certifications en vigueur.

7. Bibliographie

1. ADAE : Le répertoire des schémas XML de l'administration, http://www.adae.gouv.fr/article.php3?id_article=167.
2. Agence de modernisation des universités et des établissements, Séminaire Harpège, Présentation de l'univers Business Objects d'Harpège, 4 et 5 juin 2002.
3. Agence de mutualisation des universités, http://www.amue.fr/Amue/Default.asp.
4. Akoka, J., Entrepôts de données et bases multidimensionnelles, Paris, 2002
5. Boussaid, O., Entreposage et fouille de données, Toulouse, 2003
6. Chrisment, C., Fouille, transactions, évaluation dans les bases de données, Paris, 2004
7. Bouaka N. et David A., Modèle pour l'Explicitation d'un Problème Décisionnel : Un outil d'aide à la décision dans un contexte d'intelligence économique, IERA2003.
8. Boyer A., Nominé B., Managing new educative technology in a medium size university, In 20th ICDE World Conference on Open Learning and Distance Education. (Düsseldorf, Germany), 2001.
9. Cap Gémini Ernst & Young . Etudes préalables à l'élaboration d'un système d'information de gestion des établissements, 2003.
10. Chartron G., Les chercheurs et la documentation numérique : nouveaux services et usages, Paris, 2002.
11. David A., Modélisation de l'utilisateur et recherche coopérative d'information, 1999.
12. David A., Thiery O., Application of "equa2te" architecture in economic intelligence, 2002, http://ictei2002.loria.fr/papers/equate.htm.
13. David A., Thiery O., Prise compte du profil de l'utilisateur dans un système d'information stratégique, in veille stratégique scientifique et technique - VSST'2001.
14. David A., Thiery O., Prise en compte du profil de l'utilisateur dans un système d'information stratégique, VSST2001.
15. Desnos J.-F., Projet «Entrepôt de données», 2002, http://www.amue.fr/Telecharger/seminaire_pilotage_mars2002/J.F.Desnos.pdf.
16. Ducloy J., Cours IUT Paris 2002, http://dilib.inist.fr/~ducloy/iut.html.
17. Esup portail : Environnement numérique de travail d'accès intégré aux services pour les étudiants et le personnel de l'enseignement supérieur, http://www.esup-portail.org/.
18. Franco J.-M., Le data warehouse : le data mining, Paris, 1997.
19. Goglin J.-F., La construction du datawarehouse : du datamart au dataweb, Paris, 2001.
20. Inmon-William.-H., Building the data warehouse, New York, 2002.
21. Kimball R., The data warehouse toolkit, John Wiley and Sons, 1996.
22. Kislin P., David A., Peguiron F., Caractérisation des éléments de solutions en recherche d'information : conception d'un modèle dynamique dans un contexte décisionnel, ISKO2003, Grenoble.
23. L'Essentiel d'Unified Modeling Language (UML), http://madchat.org/coding/other/CSI_UML_2003.pdf.
24. Le Moigne J.L , Les systèmes de décision dans les organisations, Editions PUF, 1974
25. Manuel d'utilisation de l'infocentre PILOTAGE, http://ftp.amue.fr/documents_publics/apogee/II_doc_fonctionnelle/D_pilotage/MANINFO.doc.
26. Martre, H. Intelligence économique et stratégie des entreprises, Rapport du Commissariat Général au Plan, Paris, La Documentation Française, 1994.
27. Muller P.-A., Modélisation objet avec UML, Eyrolles 1997.
28. Muller P.-A., Représentation des vues d'architecture avec UML, http://magda.elibel.tm.fr/refs/UML/architecture.pdf.
29. Nataf J.-B., Structure de l'entrepôt de données de pilotage, 2001, http://www.cpu.fr/Telecharger/NatafP6StructureEntrepotDeDonnees.pdf.

30. Nominé B., ESUP portail : espace numérique de travail pour tous, Nancy 2004.
31. Peguiron F., Accès à l'information sur Internet, pratiques et tendances des utilisateurs : dans un contexte de documentation électronique, Université de Nancy-Metz, 2001.
32. Peguiron F., David A., Thiery O., Intelligence économique dans un cadre universitaire intégrant la modélisation de l'utilisateur, IERA 2003, Nancy, http://www.sciences.bu.u-nancy.fr/Parc/recherche/IERA2003.doc.
33. Peguiron F., Kislin P., Bouaka N., Activity-based classification of university actors for the construction of a domain-oriented data warehouse, SCI2003, http://www.sciences.bu.u-nancy.fr/Parc/recherche/SCI2003.doc.
34. Peguiron F., Thiery O. Modéliser l'acteur dans le système d'information stratégique d'une université, VSST 2004, http://www.sciences.bu.u-nancy.fr/Parc/recherche/VSST2004.doc.
35. Peguiron F., Thiery O., Modélisation des acteurs et des ressources : application au contexte d'un SIS universitaire, ISKO-France 2005, Nancy, http://www.sciences.bu.u-nancy.fr/Parc/recherche/ISKO2005.doc.
36. Pottier, S., Mise en place de méthodes et d'outils pour le processus d'extraction de données en vue d'analyse décisionnelle : La méthode RADHE, Université de Nancy 2, 2002.
37. Revelli C., Intelligence stratégique sur Internet : comment développer des activités de veille et d'intelligence économique sur le web, 2000.
38. Rochfeld A., Morejon J., La Méthode Merise, Tome 3, Gamme opératoire, Editions d'Organisation, 1989.
39. Rumbaugh J., et al., Modélisation et conception orientées objet, Masson, 1995.
40. Schéma directeur des espaces numériques de travail, Ministère de la jeunesse, de l'éducation nationale, et de la recherche, 2004, http://www.educnet.education.fr/chrgt/SDET-v1.doc.
41. Teste O., Modélisation et manipulation d'entrepôts de données complexes et historisées, http://www.irit.fr/recherches/IRI/SIG/personnes/teste/these/intro.pdf.
42. Thiery O., David A., Modélisation de l'utilisateur : systèmes d'informations stratégiques et intelligence économique, Revue association pour le développement du logiciel (ADELI), 2002.
43. Thiéry O., Ducreau A., Bouaka N., David A., Piloter une organisation : de l'information stratégique à la modélisation de l'utilisateur ; application au domaine de la GRH, GREFIGE 2004.
44. Thiéry O, Support de cours recherches avancées en SIS DESS ACSI et SID de Nancy.
45. Thiéry O, Support de cours SIS IUP Miage de Nancy.