HAL
open science

Méta modélisation et systèmes d'informations stratégiques

Odile Thiery, Frédérique Peguiron

To cite this version:

Odile Thiery, Frédérique Peguiron. Méta modélisation et systèmes d'informations stratégiques. Veille Stratégique Scientifique et Économique - VSST 2009, Mar 2009, Nancy, France. 20 p. hal-00419984

HAL Id: hal-00419984

https://hal.science/hal-00419984v1

Submitted on 27 Sep 2009

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.

Méta modélisation et systèmes d'informations stratégiques

Odile Thiéry, Frédérique Peguiron
Odile.Thiery@loria.fr, Frederique Peguiron@loria.fr
SITE-LORIA Nancy Université, B.P. 239 54506 Vandoeuvre, France

Mots clefs :
Métas données, systèmes d'information stratégiques, entrepôt documentaire, Métamodélisation.
Keywords:
Meta data, information systems strategic, data warehouse, Meta modelling.
Palabras clave :
Meta datos, sistemas de información estratégica, almacén de datos, modelado de Meta.

Résumé
Les sources documentaires sous forme d'informations primaires, d'informations secondaires, d'informations tertiaires et d'informations à valeur ajoutée sont désormais disponibles par les nouvelles technologies d'information. Pour caractériser correctement les informations issues de base de données et permettre de produire de la connaissance, une première étape, communément admise, est de caractériser les données par des métas données c'est-à-dire des données sur les données. Une vision intéressante dans l'approche du décisionnel consiste à mettre l'acteur du système d'informations (SI) au centre du problème. La méta modélisation permet de décrire les objectifs de l'utilisateur, ses différentes activités et ses besoins dans la modélisation d'un système – d'informations stratégiques (S-IS). Les métas connaissances pour la représentation des connaissances du domaine d'application relèvent de la méta base chargée de contenir les structures des bases métiers. Ces bases métiers sont utilisées pour l'exploration des contenus des bases et pour des analyses décisionnelles. Les connaissances ou les informations contenues dans la méta base portent sur les attributs nécessaires pour les explorations et les analyses multidimensionnelles.
Dans cet article nous explicitons les trois niveaux se modélisation d'un SIS et des acteurs impliqués dans le SIS.

1 Introduction

Le rôle de plus en plus prédominant de l'information dans le secteur économique et social et l'organisation en général n'est plus à démontrer. Les technologies de l'information et de la communication, en particulier l'informatique et l'Internet, permettent d'intégrer des informations de sources et de natures différentes. Par exemple, on peut employer les informations provenant des bases de données d'entreprise ou des sources documentaires sur Internet. S'agissant des sources documentaires, des informations primaires, des informations secondaires, des informations tertiaires et des informations à valeur ajoutée sont désormais disponibles par ces nouvelles technologies. Alors que les informations primaires sont les œuvres directes des producteurs (ou auteurs), les informations secondaires et tertiaires sont des transformations en « modèles réduits » des informations primaires, pour alimenter des bases documentaires[1]. Les informations à valeur ajoutée sont les produits d'analyse et de synthèse de ces différents types d'information. L'information est de plus en plus utilisée comme objet de référence et comme outil d'aide à la décision d'ordre stratégique. Le concept d'intelligence économique (IE) s'affirme, où il s'agit d'étudier les processus impliqués dans la production des indicateurs interprétables pour la prise de décision en se basant sur des informations internes et externes à l'organisme.

Le processus d'IE repose en particulier sur l'utilisation de systèmes d'informations stratégiques (SIS). Les systèmes d'informations (SI) existent depuis fort longtemps. Ils ont subi des évolutions profondes soit par le but final de l'utilisation du système soit par le type d'information géré soit par la combinaison des deux. C'est ainsi que nous voyons apparaître la déclinaison du terme SIS en « système d'information » « stratégique » (SI-S) et « système » d' « information stratégique » (S-IS).

Enfin pour caractériser correctement les informations issues des données de base et permettre de produire de la connaissance, une première étape, maintenant communément admise, est de caractériser les données par des métas données c'est-à-dire des données sur les données. Le paragraphe 2 introduira ce concept.

Dans la littérature on rencontre essentiellement des écrits sur les métas données de type « documentaires » qui représentent de fait le schéma conceptuel de la base d'informations. Cela fera l'objet du troisième paragraphe.

Mais le type de métas données qui fait l'objet de notre recherche sont les métas données des entrepôts de données, noyaux de tout système d'informations stratégiques.

Ce point fera l'objet du quatrième paragraphe qui sera complété par la modélisation et la méta modélisation dans le contexte d'un SIS universitaire. Enfin nous conclurons sur les perspectives de ce travail.

2 Qu'est ce qu'une méta donnée ?

Le concept de méta donnée a été abordé dans le Dublin Core en 1985. Ce travail consistait à décrire des ressources numériques ou physiques et d'établir des relations avec d'autres ressources. Il comprend officiellement 15 éléments de description formels (titre, créateur, éditeur), intellectuels (sujet, description, langue, …) et relatifs à la propriété intellectuelle.

Ces travaux ont été repris par Youcef Amerouali [6], associé au principe de Diffusion Sélective de l'Information. Le but du travail de Youcef Amerouali est de proposer à l'utilisateur, lors des recherches documentaires, des informations en fonction de son profil. Il se propose d'associer à une description de ressources par un standard de métas données, en l'occurrence le Dublin Core, des éléments de profil utilisateur. Dans le processus de recherche de

1 Notons q'une base documentaire est une base de données particulières où les objets gérés se ramènent aux documents. Nous employons le terme document pour designer un ensemble d'informations considéré comme une unité.

l'information, il introduit la notion d'interrogation à relances, notion présente déjà dans certains projets basés sur le Dublin Core, en Australie et en Grande-Bretagne, mais non encore définie dans des contextes analogues à la recherche d'information.

3 Les méta données dans un contexte documentaire

3.1 La nécessité des métas données

[http://www.sciencepresse.qc.ca/lbq/lbq21.6.html] Lorsqu'on voit que les algorithmes de recherche sont de plus en plus raffinés, les ordinateurs sont de plus en plus puissants, on se pose la question : les métas données sont-elles nécessaires ?
Prenons l'exemple d'un usager qui chercherait un livre dans une bibliothèque en lisant tous les documents, à partir de la cote A jusqu'à la cote Z, en se disant qu'il finirait bien par tomber sur l'information qu'il recherche ? C'est un comportement pour le moins inefficace. Pourtant c'est à peu près ce qu'on fait actuellement sur Internet en utilisant la force brute de l'informatique et les index inversés. C'est puissant, c'est rapide, mais ce n'est certainement pas la solution la plus rationnelle. De plus, aucun moteur de recherche ne couvre le Web dans son entier et, compte tenu du rythme d'accroissement, il n'est pas plausible qu'un seul moteur réussisse éventuellement cet exploit.

3.1.1 Définition

Selon Catherine Dhérent, [17] les métas données sont constituées de contenu structuré ou non qui peut être des mots, des formules, des analyses spectrales... Elles sont créées de façon automatique ou manuelle et peuvent avoir plusieurs niveaux de complexité. Les métas données sont faites pour être lues soit par l'humain, soit par la machine et peuvent être partout.

3.1.2 Quelques exemples de typologie

- *Externes aux ressources et sous forme papi*er **:** une notice papier pour remplir un formulaire dont les données sont saisies par informatique ;
- *Externes aux ressources et stockées dans une base de don*nées : un catalogue bibliographique décrivant des ouvrages numérisés ;
- *Internes et encapsulées dans la ressource* : en-tête d'un document dans une DTD (Document Type Définition, document permettant de décrire un modèle de document *SGML* ou *XML*) ;
- *Métas données de gestion administrative de gestion de droits et métas données techniques :* elles permettent de donner du sens à des codes et de créer des ontologies, elles définissent les droits d'utilisation des ressources et permettent la préservation à long terme des ressources ;
- *Métas données de description* : elles permettent la recherche des ressources les plus pertinentes sur le web et expliquent comment les ressources sont exploitables.

En résumé ces métas données permettent la gestion et la manipulation de ressources documentaires.
Les métas données sont toujours créées dans un but précis. Dans un même organisme, certains documents pourront ainsi être créés avec des jeux de métas données minimaux ou maximaux selon l'objectif (durée de conservation, usages…), il faut noter qu'une partie doit être créée en même temps que le document lui-même et que d'autres métas données peuvent être créées tout au long de la vie du document. Selon [Catherine Dhérent, 2005] de nombreuses solutions sont possibles :

- A l'aide d'un texte ou tableau bureautique décrivant la ressource ;
- En remplissant les champs Propriétés d'un document Office ;
- En désignant les destinataires et l'objet d'un message électronique ;
- En complétant les éléments d'un document HTML *HTML: Hypertext Markup Language ;* pour le référencer sur le web ;
- En remplissant les champs d'un catalogue en MARC *MARC : Machine Readable Cataloguing* ;
- En utilisant un éditeur XML *XML: Extensible Markup Language* pour créer des balises dans un document XML.

4 Les métas données pour les Systèmes d'information stratégiques

4.1 Qu'est ce qu'un SIS ?

L'évolution des SIS par rapport au but final de l'utilisation du système a eu pour résultat la proposition des entrepôts de données. Les entrepôts de données sont devenus maintenant non pas un phénomène de mode mais un instrument indispensable à la bonne marche de l'organisation. Ils sont en effet à la base de toute stratégie et prise de décision de l'entreprise[2]. Ainsi 95% du top 500 des entreprises aux USA ont mis en place un entrepôt de données, à l'origine essentiellement destiné au marketing.

Donnons quelques définitions. Un entrepôt de données[3] est une base de données organisée pour répondre aux besoins spécifiques de la prise de décision. Cette base contient des informations historiques sur l'entreprise, son fonctionnement et son environnement. Elle est alimentée à partir des bases de production et d'informations externes à l'entreprise. Elle recouvre donc les mêmes types d'informations utilisées dans un contexte d'IE. Elle est thématique, relative à un domaine intéressant le décideur, possédant une référence temporelle, sûre, c'est à dire dont la qualité a été vérifiée, facile d'accès, non volatile et régulièrement complétée. En fait, l'entrepôt de données est une vue intégrée de l'organisation. Il est le noyau du SIS.

Les SI peuvent être stratégiques sous deux angles. D'une part tous les SI actuels des organisations comportent des informations stratégiques et permettent l'automatisation de l'organisation pour satisfaire au mieux les objectifs stratégiques de la direction (exemple : un SI améliorant la gestion des stocks, élaboration à partir de résultats comptables de tableaux récapitulatifs), c'est ce que l'on appelle des SI-S (« système d'informations » « stratégique »). D'autre part de plus en plus de SI sont dédiés uniquement à la prise de décision (exemple : un SI d'aide au choix marketing), c'est ce que l'on appelle des S-IS (« système » d' « informations stratégiques »). Là c'est le SI dans son entier qui est consacré aux décisions stratégiques et ne comporte que des informations de type stratégiques. Par exemple, un SI peut permettre d'observer des résultats de chiffres d'affaire par pays sur plusieurs années. Nous nous intéressons ici aux S-IS c'est à dire aux SI de deuxième type, ceux qui sont directement dans les préoccupations des chercheurs en IE.

2 Nous utilisons le terme « entreprise » pour représenter tout type d'organisme socio-économique.

3 Ou Data Warehouse

Un entrepôt de données donne naissance, par filtrage non plus par rapport aux dimensions c'est à dire les axes d'analyse mais par rapport à des profils utilisateurs, à des bases métiers[4]. Ce sont des sous bases de l'entrepôt de données destinées à une fonction de l'entreprise : Marketing, Financier... Elles sont alimentées périodiquement, reposent sur une vue multidimensionnelle des données, et elles sont non modifiables par les utilisateurs.
Notre objectif est de faire des propositions permettant de concevoir un SIS de qualité et répondant aux besoins des différents acteurs de l'organisation.

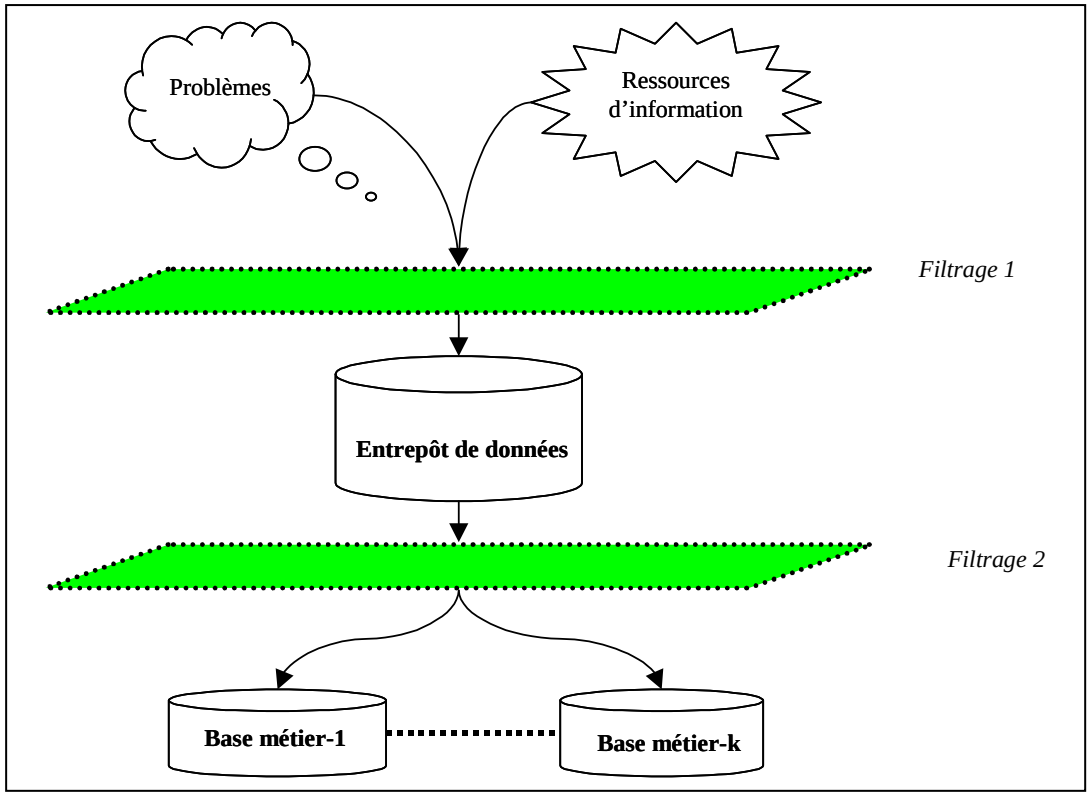

Figure 1 : Modélisation de l'utilisateur et S-IS

La modélisation de l'utilisateur intervient alors dans deux contextes en S-IS : pour la constitution des informations pertinentes (filtrage 1) et pour l'exploitation (filtrage 2) 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

4 ou data mart

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.
Notre approche de filtrage de l'information par le profil de l'utilisateur est donc fondée sur la modélisation de l'utilisateur en intégrant des attributs d'identification de l'utilisateur, de son comportement et du contexte d'utilisation des informations qu'il cherche.
La technique qui consiste en l'évaluation des propositions du système pour indiquer leur degré de pertinence est intégrée dans certains SRI. Le système dispose ainsi des connaissances sur l'adéquation partielle des réponses du système au besoin de l'utilisateur car le système ne connaît réellement ce besoin que par une estimation basée sur les requêtes. Au lieu de calculer ce besoin, nous proposons d'en intégrer la représentation dans le modèle de l'utilisateur, ce qui constitue l'originalité de notre proposition. Cela revient dans un S-IS à stocker parmi les **métas données** du système, une représentation explicite de la structure des différentes bases métiers. Notre préoccupation principale est donc la prise en compte de l'acteur dans la construction puis l'exploitation d'entrepôts de données.
Les métas données sont dans ce contexte des informations sur les données indispensables à une exploitation efficace d'un data warehouse (DW). Le terme de métas données est, nous l'avons dit, couramment défini comme "des données concernant les données". Les métas données décrivent le schéma d'un DW et les données individuelles. Dans l'entreposage de données, les métas données sont classées d'après leur objet et le public auquel elles s'adressent. La donnée est forcément liée à d'autres objets du système d'information ; il est donc également nécessaire de représenter, de décrire et de stocker ces interactions avec d'autres données. Notre préoccupation principale est donc la prise en compte de l'acteur dans la construction puis l'exploitation d'entrepôts de données. 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'entrepôt de données. Les métas données permettent de décrire l'ensemble des règles, des définitions, des transformations et des processus associés à une donnée. On peut donc, par ce moyen, suivre l'évolution des données depuis le système source jusqu'à leur restitution. On est capable, dès lors, d'avoir une vision unifiée des données, transverse à tous les métiers de l'organisation.

4.2 Typologie des métas données dans le contexte SIS

La description des métas données permet d'éviter la perte de données essentielles. Le processus de description des métas données permet d'inscrire les méthodes de collecte et de traitement des données. Cela favorise l'utilisation ultérieure des données.
Les principales informations sont destinées :

- A l'utilisateur : informations sur la sémantique des données utilisées et leur localisation dans l'entrepôt de données ;
- 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.

Les métas données qui permettent de constituer l'entrepôt à partir des sources de données hétérogènes, puis de les exploiter, peuvent alors être de différents types :

- Les métas données des systèmes sources donnent les informations sur les structures de données initiales des BD de production,
- Les métas données du SGBD (ou des SGBD) qui sont en fait les tables systèmes décrites ci dessus,

- Les métas données des outils frontaux, c'est à dire en fait des interfaces utilisateurs,
- Enfin les ***métas données des données de l'entrepôt***.

Nous nous focaliserons ici sur ces dernières, en leur rajoutant **les métas données sur les acteurs du SIS**.

En fait il y a deux types principaux de métas données pour un DW :

⇒ des métas données structurelles qui décrivent la structure et le contenu de l'entrepôt (son schéma conceptuel) qui sont en fait fort proches des métas données décrivant des ressources documentaires ;

⇒ et des métas données d'accessibilité qui représentent le lien dynamique entre l'entrepôt et les utilisateurs.

Dans les référentiels actuels d'entrepôt ce sont surtout les MD d'accessibilité qui sont détaillées : elles concernent les règles de calcul des données, les règles d'agrégation des résultats selon les hiérarchies de dimension enfin elles permettent d'indiquer de quelle source de donnes elles sont issues et donc qui les administre.

La figure suivante représente assez bien ces métas données :

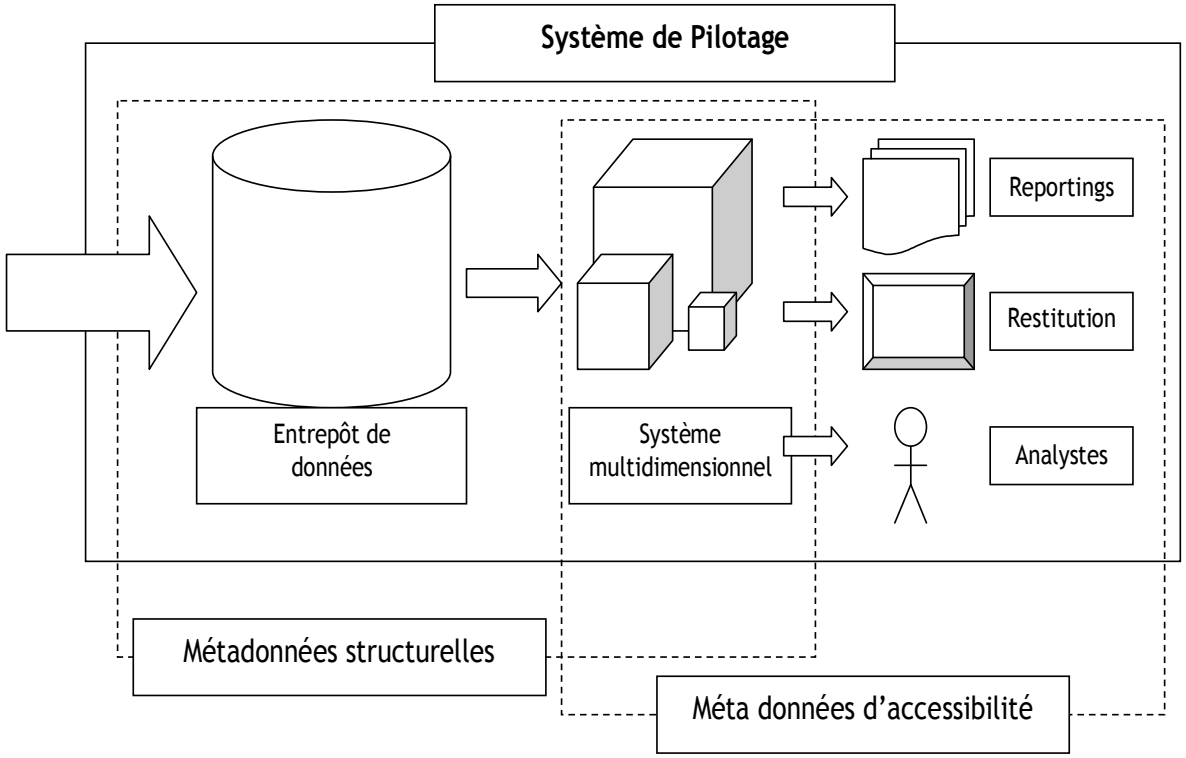

Figure 2: Typologie des métas données d'un entrepôt

Nous rajoutons, comme dit précédemment, les **métas données « utilisateur »** que nous préférons considérer comme des MD structurelles car elles sont là pour spécifier la structure de la base métiers.

5 Modélisation et méta modélisation

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[5], WISP[6].et MORPRIE[7] qui reprend le modèle MEDP en l'étendant à une meilleure prise en compte de l'utilisateur puisqu'en fait il est support d'un système d'information adaptatif à l'évolution des besoins des acteurs[2]. Parallèlement à ces deux modèles, nous trouvons un panel de normes autour des documents électroniques et de ressources en ligne.
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'informations les plus adaptées.

5.1 Modélisation des ressources dans le cadre d'un SIS universitaire

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[8]. Dans un contexte mondial, on favorise les échanges internationaux pour les formations, les stages et les emplois.

5.1.1 Typologie des ressources documentaires

Le SI de l'université propose différents types de documents à destination de types d'acteurs différents. 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. Par exemple, les classes individu, réussite aux diplômes, stages et ressources d'information permettent : de proposer des ressources par 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.

---

5 MEDP : Modèle pour l'explicitation d'un problème décisionnel

6 WISP : Watcher-Information-Search-Problem

7 MORPRIE : Modèle de résolution d'un problème de recherche d'information en IE

8 LMD : Licence-Master-Doctorat

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.

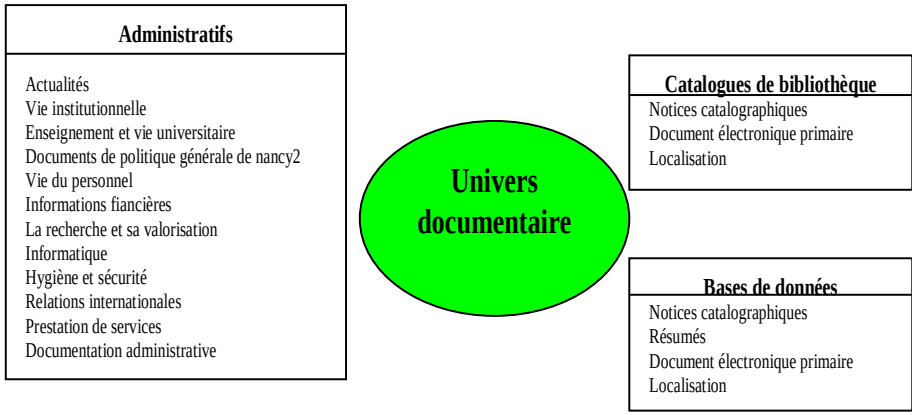

Figure 3 Modélisation du type des documents

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 ?

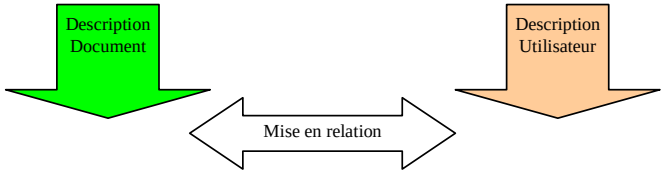

Figure 4 - Filtrage d'information pour les ressources documentaires par mise en correspondance du profil document et du profil utilisateur

5.1.2 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.

2ème 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.

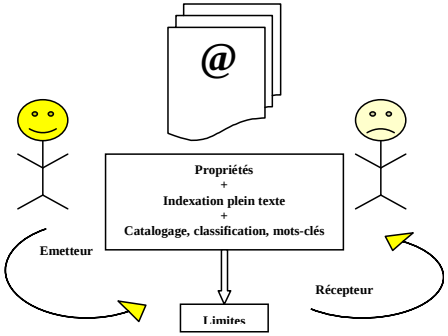

Figure 5 - Limites des descriptions des ressources documentaires

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

Documents de tout type

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[9] ou XMP[10] sont des réponses pour ajouter des métas données[11], 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[12], EXIF[13], DIG35[14], JPEG[15] ont été plus particulièrement développées pour les images. MPEG-

9 RDF : Resources Description Framework

10 XMP : Extensible Metadata Platform

11 Une métadonnée est une "donnée sur des données"

12 IPTC : International Press Telecommunications Council

13 EXIF : EXchangeable Image File

14 DIG : Digital Imaging Group

15 JPEG : Join Photographic Experts Group

7[16] concerne la description des objets multimédia. RKMS[17] aide à décrire les ressources audio. PRISM[18], NewsML[19], NITF[20] 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[21], EML[22] et Scorm[23] 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 métas données de l'entrepôt de données. Nous allons à présent définir et décrire ce processus et tout d'abord comment concevoir le référentiel de l'entrepôt de données, c'est-à-dire le dictionnaire des métas données.

5.1.3 Référentiel de l'entrepôt de données [60]

Il existe plusieurs manières de représenter et de stocker ces métas données. Pour en faciliter l'accès et la manipulation, on pourrait utiliser une fiche contenant une description de la donnée et de ses liens. Mais généralement, on utilise un symbolisme courant en informatique : les modèles de données au sens Merise du terme, ou les modèles de classes d'objets au sens UML du terme. Les métas données sont exprimées sous forme de modèles conceptuels et de modèles logiques de données. Ainsi, on obtient une vision synthétique permettant de naviguer entre les différentes données. Néanmoins, pour affiner cette vue, on peut utiliser, à l'instar de l'entrepôt de données, des modèles multidimensionnels pour représenter le référentiel (selon axe d'origine, de sémantique, etc.). C'est souvent le lien entre le modèle opérationnel et le modèle décisionnel qui est décrit par les métas données. Or les modèles opérationnels sont dits locaux : dans un entrepôt nous avons besoin d'élargir leur vision afin de consolider une vue au niveau institution. L'idée est d'obtenir ainsi une représentation unique d'une donnée quel que soit le secteur ou l'acteur de l'université. L'ambition avouée : fédérer l'ensemble.

Pour cette raison, le besoin d'une fonction de gestion des méta données apparaît clairement lors de la mise en place d'un projet décisionnel, fonction que l'on va appeler souvent «administration des données». L'ensemble des outils nécessaires à la mise en oeuvre d'une fonction est le référentiel de l'entrepôt de données. Cet outil doit être apte à gérer et à construire un modèle d'institution, enrichi de données de suivi propres à l'entrepôt.

---

16 MPEG : Moving Pictures and associated audio information coding Experts Group

17 RKMS : Recordkeeping Metadata Schema

18 PRISM : Publishing Requirements for Industry Standard Metadata

19 NewsML : News Markup Language

20 NITF : News Industry Text Format

21 LOM : Learning Object Metadata

22 EML : Educational Modelling Languages

23 Scorm : Sharable Content Object Reference Metadata

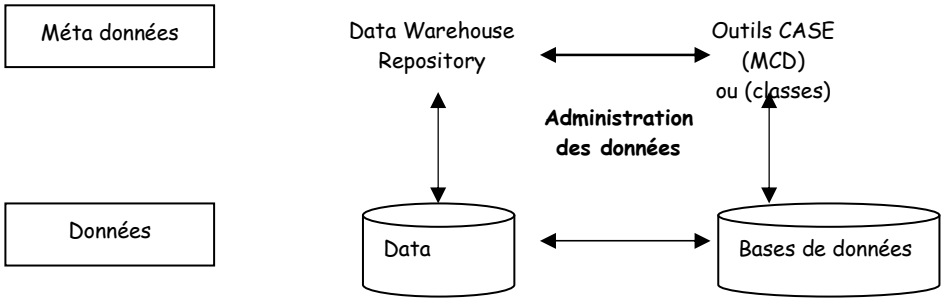

Figure 6 Vision du référentiel de l'entrepôt de données

5.1.4 Méta modélisation [57]

Les métas connaissances dont nous avons besoin pour obtenir un S-IS adaptable sont des connaissances sur 1) la représentation d'un problème à résoudre, 2) la représentation des comportements de l'utilisateur et 3) la représentation des informations du domaine d'application. Les métas connaissances de 1) et 2) sont liées à la modélisation de l'utilisateur et sont indépendantes des domaines d'application. Ces métas connaissances constituent en fait le méta modèle de l'utilisateur.

Une vision intéressante dans l'approche du décisionnel consiste à mettre l'acteur du SI au centre du problème et de lui offrir une possibilité d'analyser les données qui l'intéressent et seulement celles-ci. Proposer une solution globale mais dont chaque acteur pourra n'avoir à sa disposition que les parties qu'il souhaite. Pour répondre au mieux aux besoins des utilisateurs, l'enjeu est 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 rendue nécessaire s'appuie sur la modélisation de la dimension humaine représentée par les différents acteurs pour concevoir un SIS. La modélisation reflète une image réelle du système. Pour être performant, un système d'information doit tenir compte des besoins spécifiques de chaque acteur. Dans [45] nous avons fait émerger des besoins, des fonctions et des activités propres à des types d'acteurs d'un système d'information pour proposer de représenter l'utilisateur ainsi RU = (T, B, F, A), où :

T est le type d'acteur, B sont les besoins, F sont les fonctions et A sont les activités des acteurs.

Si T = {Etudiants} voici comment qualifier leurs besoins, leurs fonctions et leurs activités.

b-étudiant = {s'inscrire, s'exercer, se former, rechercher emploi, rechercher stage}

f-étudiant = {apprendre, créer, intégrer, vérifier}

a-étudiant = {déposer, explorer, interroger, analyser, synthétiser, annoter}

Si T = {Enseignants} nous qualifions de la même façon leurs besoins, leurs fonctions et leurs activités.

B-enseignant = {exercer, former, corriger, recenser, évaluer, budgétiser, déployer, planifier, se conformer aux textes officiels}

F-enseignant = {créer, enseigner, diriger, missionner, organiser, gérer, conseiller, superviser}

A-enseignant = {déposer, indexer, diffuser, explorer, interroger, analyser, synthétiser, annoter}

Besoins, fonctions et activités évoluent en temps réel et cela en relation avec l'évolution des mutations technologiques des SI où par exemple les enseignants écrivent ou co-rédigent des articles. Ainsi apparaissent des environnements numériques où se tissent des relations entre les acteurs de même types ou de types

différents qui sont intéressantes à découvrir. Des niveaux institutionnels ou des niveaux thématiques sont ainsi mis en évidence par la production d'acteurs : Qui écrit avec qui ? Sur quoi ? Ce principe de coopération dynamique favorise un passage de l'information à la connaissance.
Cette représentation de l'utilisateur aboutit au modèle formel RUBI[3] {Représentation des Utilisateurs et de leurs Besoins en Information lors de l'Interrogation après Identification} qui améliore la représentation des utilisateurs pour la fabrication des bases métiers. RUBI[3], permet de donner des vues différentes du SI-S (« Système d'Information » Stratégique) aux différents acteurs. Dans un S-IS (Système d'« Informations Stratégiques »), l'idée est de stocker parmi les méta données du système, une représentation explicite de la structure des différentes bases métiers. Les Meta données de l'utilisateur et la méta-modélisation de l'entrepôt de données sont imbriquées puisque guidées par les besoins des utilisateurs. RUBI[3] est complété par RUBICUBE[24] qui résume le processus global de modélisation d'un système d'information stratégique avec prise en compte du modèle utilisateur selon un niveau modélisation, un niveau application, un niveau méta modélisation. La plate forme RUBI[3] offre aux utilisateurs, pour le moment des personnes de l'université, de créer leurs propres profils avec les actions qu'elles souhaitent et les mesures qui les intéressent. Le système est intelligent et peut proposer des profils par défaut avec certaines options déjà existantes par rapport aux profils similaires. Le point fort de ce programme est de permettre une analyse multidimensionnelle des données en temps réel.
Les trois niveaux de méta modélisation :
Pour répondre à notre objectif de méta modélisation d'un S-IS, nous étudions le S-IS sous forme de trois niveaux de modélisation que nous décrivons dans la figure 7.
Au niveau de la couche 3, on insiste sur le méta modèle de l'utilisateur : en capturant des connaissances sur le décideur pour les mettre dans les métas données de l'entrepôt afin de construire une base métier spécifique à un groupe de décideur ou mieux encore à un décideur particulier.
Au niveau de la couche 2 : en stockant parmi les métas-données du système, une représentation explicite de la structure des différentes bases métiers, ou plutôt celle des utilisateurs à qui elles sont destinées.
Au niveau de la couche 1, on prend en compte l'acteur (utilisateur) dans la construction puis l'exploitation de l'entrepôt de données.

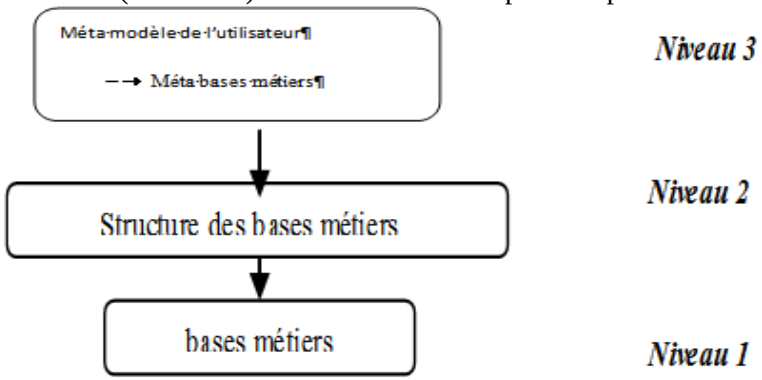

Figure 7 Les trois niveaux de modélisation d'un S-IS

Ce méta modèle fait partie du troisième niveau de la hiérarchie de modélisation que nous proposons pour la modélisation d'un S-IS. Il permet de décrire d'une part l'objectif de l'utilisateur, d'autre part ses différentes activités.
Les métas connaissances pour la représentation des connaissances du domaine d'application relèvent de la méta base chargée de contenir les structures des bases métiers. Comme ces bases sont utilisées pour l'exploration des contenus des bases et pour des analyses décisionnelles, les connaissances ou les informations contenues dans la méta base portent sur les attributs nécessaires pour ces explorations et analyses.
Ainsi au niveau 1 on trouve les instances des deux bases métiers étudiants et enseignants obtenues par extraction de données de l'entrepôt, éventuellement complétée par des données externes :

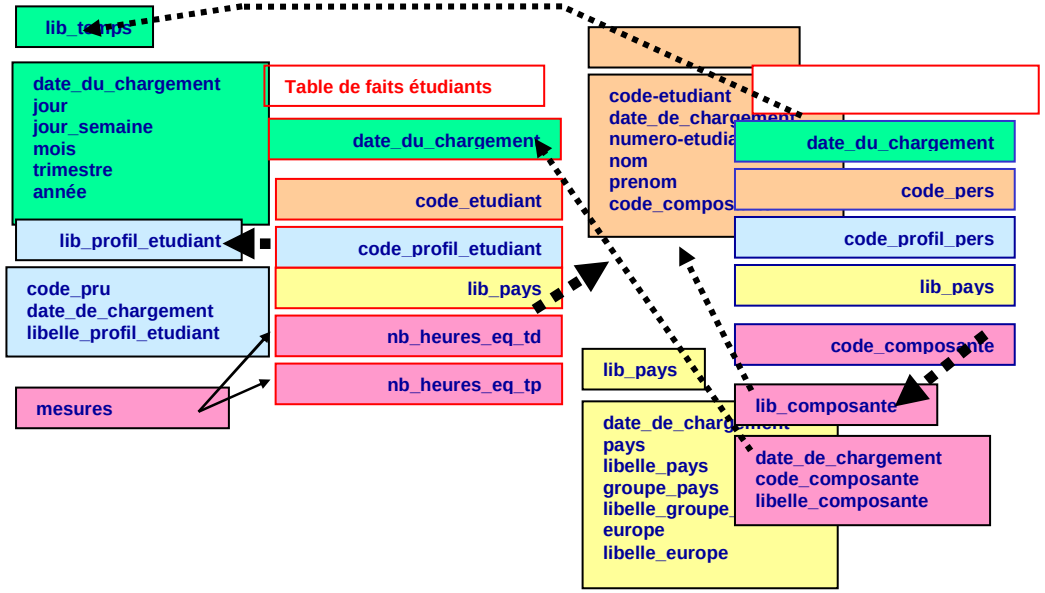

Figure 8. Mise en relation de 2 tables de faits

La constellation de faits (figure 8) permet de représenter plusieurs tables de faits partageant quelques tables de dimension Les tables de faits (étudiants) et (enseignants) peuvent être mises en relation par la dimension temps et géographie,
Le niveau 2 permet de savoir comment extraire les données de la base de données initiale, vers la structure de l'entrepôt de données. A ce niveau :
La figure ci après représentant l'univers étudiant, permet de spécifier d'où vient l'information. Quels sont les liens entre la source d'information et la structure actuelle de l'entrepôt de données.

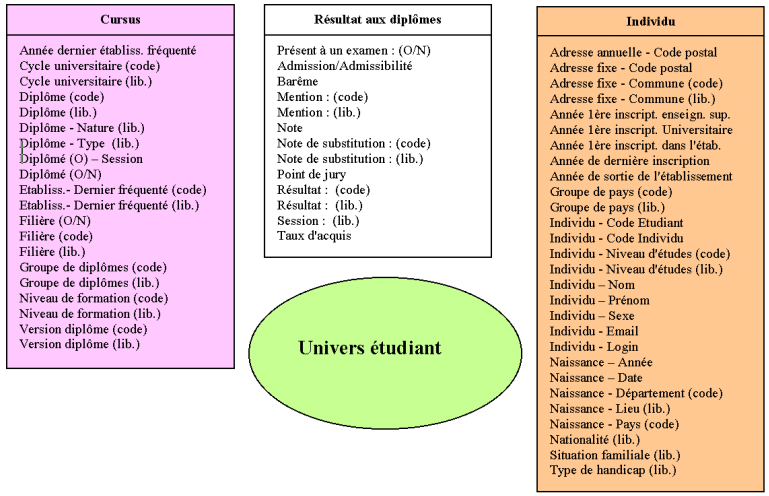

Figure 9. Exemple de 3 classes d'objet modélisant l'acteur étudiant

La figure 10 est un extrait de schéma conceptuel : il donne pour un diplôme la conception d'une U.E. Pour un diplôme, il y a plusieurs unités d'enseignement. Ce niveau nous permet de savoir comment associer chaque unité d'enseignement au diplôme.

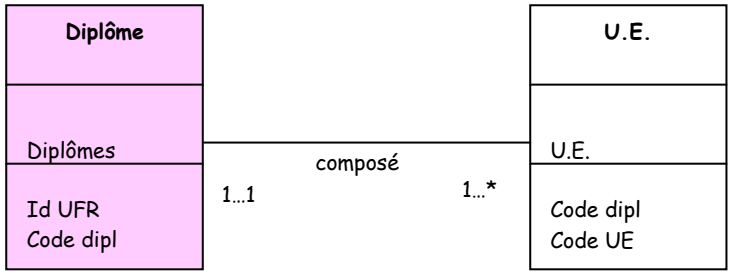

Figure 10. Classe diplôme

Enfin au niveau 3, il faut définir la méta structure de l'entrepôt de données. Ce niveau permet d'élaborer la différence entre la structure pour l'étudiant, l'enseignant, l'administration :

La méta structure de la base de données permettant de façon générique le stockage des schémas conceptuels des différentes bases métiers : étudiant, enseignant, administration (figure 11)

De la même façon il faut définir la méta base permettant de stocker les structures des modèles utilisateurs intégrés aux différentes bases métiers.

La figure suivante donne un méta schéma permettant de stocker non seulement les structures de bases métiers comme tout méta schéma étendu ou UML mais aussi les métas modèles des acteurs. En effet cette méta structure est tout à fait générique.
On peut la représenter ainsi mais à vrai dire on aurait pu s'appuyer sur le méta schéma proposé par la notation UML et concernant le diagramme de classes d'objets.

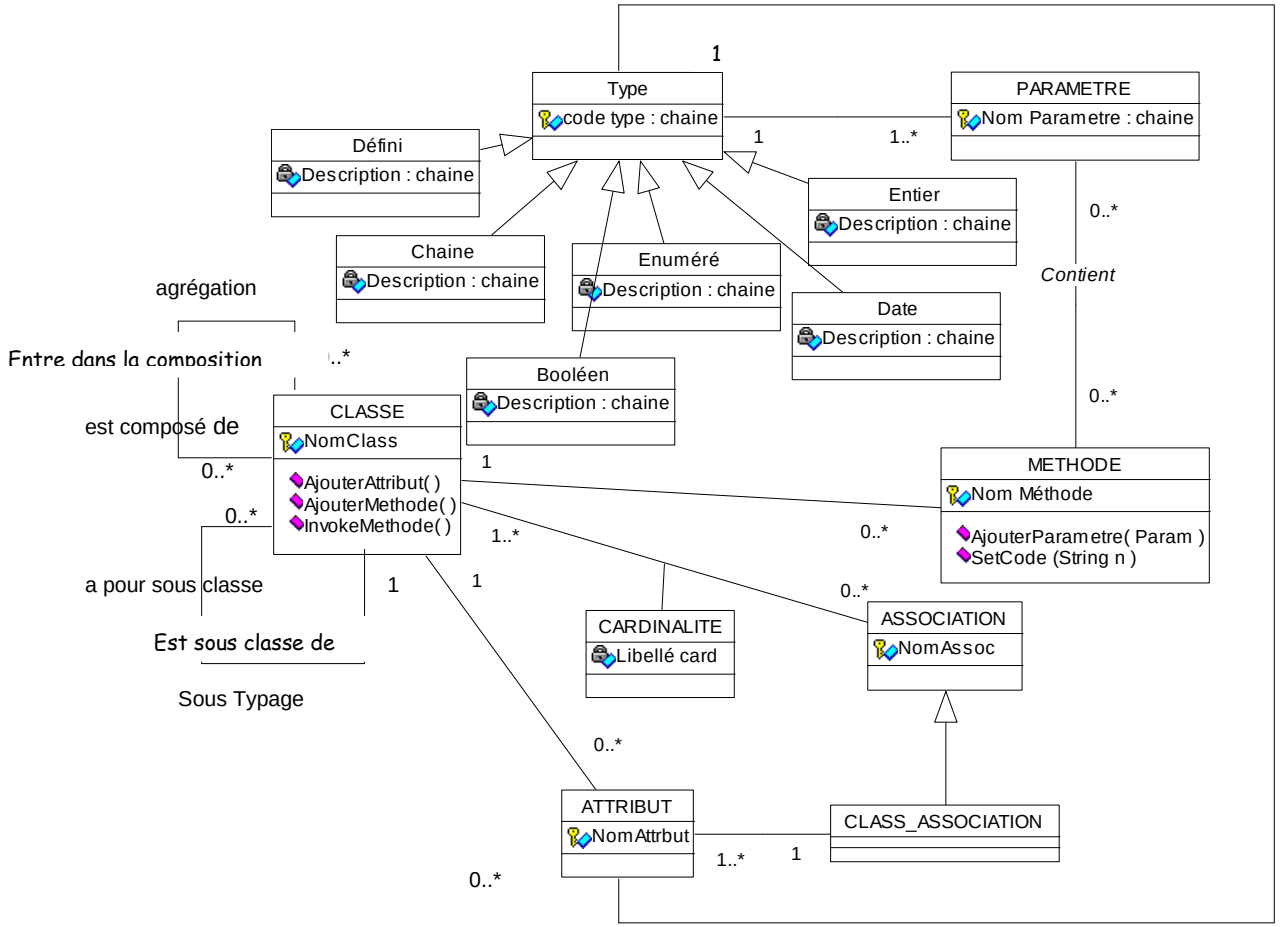

Figure 11. Méta structure générique, bases métiers et modèles d'acteurs

Quelques commentaires :
Le schéma de classes UML s'appuie sur deux concepts principaux : la classe d'objets et la méthode ou opération rattachée à la classe et, ce, par les classes d'objets « classe » et « méthode ».
Les classes sont reliées entre elles pas des associations avec des cardinalités (par exemple 0..* veut dire qu'une classe est reliée à une autre classe par 0 ou plusieurs occurrences) et, ce, par les classes « association » et « cardinalité ».
Certaines associations sont des classes d'associations qui permettent d'exprimer des propriétés de classes d'objets. En effet en UML une association n'a pas de propriété, classe « class_association ».
Ainsi on retrouve une classe « attribut » qui caractérise une classe d'objets ou d'associations.
Un attribut a un certain « Type » qui peut être redéfini en différents sous types. Ceci est valable également pour les éléments de la classe « paramètre ».
Enfin une classe d'objet peut se redécomposer en sous types (association « Sous Typage » rôles « a pour sous classe » et « est sous classe de ». Et aussi entrer dans une agrégation de classes ou être agrégé d'autres classes (association « agrégation », rôles « est composé de », « entre dans la composition ».
En résumé cette méta base permet d'une part de stocker le schéma conceptuel d'un SIS ou des bases métiers ainsi que les différents modèles d'acteurs conçus selon le modèle RU. Ce schéma exprimé en amont de la construction d'un SI coopératif et dynamique permettrait par exemple de générer en temps réel les relations entre les types d'acteurs, leurs institutions et leurs thématiques.

6 Conclusion

Dans cet article nous avons défini dans un premier temps ce que nous entendons par métas données. Puis nous les avons présentées dans un contexte classique de ressources documentaires et dans le cadre plus novateur de SIS.
Enfin le paragraphe 5 nous a permis d'expliciter ce que nous entendons par méta modélisation. Quelques exemples démonstratifs ont été développés dans le cadre d'un SIS universitaire et une méta base UML conclut le propos.
Les mondes de l'indexation ou du référentiel et les mondes du décisionnel sont reliés par les entrepôts de données. Le travail de conception et de construction d'entrepôts favorise un travail collaboratif des acteurs pour mettre en commun des ressources qu'ils ont besoin de partager. Le monde de l'indexation et le monde du décisionnel sont reliés par les entrepôts de données ou plus exactement le monde du référentiel et le monde du décisionnel. C'est dans la nécessité à réfléchir au référencement sous l'angle du choix du référencement et de sa mise en application que seront facilitées les possibilités d'analyses multidimensionnelles. Il ne s'agit plus seulement d'indexer pour répondre à des requêtes par les utilisateurs à l'aide de mots clés, mais de référencer pour favoriser des analyses faites par les acteurs en vue de passer d'un processus de recherche d'information à un processus de recherche de connaissances qui confère une « intelligence » au système d'information. A l'instar de ce qui est fait dans les bases de connaissances chaque service d'une organisation devrait exposer une ressource à l'aide de méta données pour en permettre des analyses multidimensionnelles.
Ce qui nous reste à faire : outre une définition plus approfondie d'un modèle d'acteur générique implantable sous forme de métas données, il nous reste à mettre en place les métas données « acteur » dans la plate forme RUBI[3] (développée autour de Mondrian et Pentaho). C'est notre travail à court et moyen terme avec la préoccupation suivante : comment notre plate forme décisionnelle pourrait s'articuler avec un WIKI sémantique notamment autour de la notion de méta modélisation ? Ceci en réponse aux questionnements sur les répercussions institutionnelles induites par la recherche et la pédagogie au travers d'écrits.

7 Bibliographie

[1] ADAE *: Le répertoire des schémas XML de l'administration, http://www.adae.gouv.fr/article.php3?id_article=167.*

[2] AFOLABI B., *La conception et l'adaptation de la structure d'un système d'intelligence économique par l'observation des comportements de l'utilisateur, Thèse Université Nancy 2, Mars 2007.*

[3] AFOLABI B., THIERY O. Using Users' Expectations to Adapt Business Intelligence Systems, in G. Budin, C. Swertz et K. Mitgutsch (Eds.), Advances in Knowledge Organization (ISKO), vol. 10, Ergon Verlag, Würzburg, Austria. p. 247 - 254. June 2006.

[4] Agence de modernisation des universités et des établissements, Séminaire Harpège, Présentation de l'univers Business Objects d'Harpège, 4 et 5 juin 2002.

[5] *Agence de mutualisation des universités, http://www.amue.fr/Amue/Default.asp.*

[6] AMEROUALI Y., *Metadata et Profil Utilisateur, 2000* www.cais-acsi.ca/proceedings/2000/amerouali_2000.pdf.

[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, IERA200,3, Nancy Avril 2003.*

[8] BOYER A., NOMINE 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, Cercle de la librairi, 2002.*

[11] DAVID A., *Modélisation de l'utilisateur et recherche coopérative d'information, HDR, Nancy 2, 1999.*

[12] DAVID A., SIDHOM S., *Intégration de la démarche d'Intelligence Économique dans l'architecture fonctionnelle d'un système d'information, CERIST, 2004.*

[13] DAVID A., THIERY O., *Prise en compte du profil de l'utilisateur dans un système d'information stratégique, in veille stratégique scientifique et technique - VSST'2001, Barcelone Octobre 2001.*

[14] DAVID A., THIERY O., *Application of "equa2te" architecture in economic intelligence, 2002, http://ictei2002.loria.fr/papers/equate.htm.*

[15] DAVID A., THIERY O., *Prise en compte du profil de l'utilisateur dans un système d'information stratégique, VSST2001, Barcelone, Octobre 2001.*

[16] DELEFOSSE M., BOUAZIZ F., *Gestion des métas données dans un entrepôt de données : état de l'art et implantation dans les logiciels actuels, projet bibliographique M2 Miage ACSI, décembre 2008.*

[17] DHERENT C., *Journée d'information AFNOR CG 46, 7*, Bibliothèque nationale de France, juin 2005.

[18] DESNOS J.-F., *Projet «Entrepôt de données», 2002,*

http://www.amue.fr/Telecharger/seminaire_pilotage_mars2002/J.F.Desnos.pdf.

[19] DUCLOY J., *Cours IUT Paris 2002, http://dilib.inist.fr/~ducloy/iut.html.*

[20] DUFFING G. , THIERY O., *Gestion et qualité de l'information stratégique : une approche par les risques des systèmes décisionnels, Congrès AFME2008, Grenoble, mars 2008***.**

[21] 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/.

[22] FRANCO J.-M., *Le data warehouse : le data mining, Paris, 1997.*

[23] FRANCO J.-M., *Le Datawarehouse : objectifs, définitions, architectures. Eyrolles, 1997*

[24] GABRIEL, A., OHAYON, E., *Les outils décisionnels : description de l'offre commerciale et Open Source. 2007 http://www.loria.fr/~peguiron/html/profile_fset.html*

[25] GABRIEL A., OHAYON E., *Projet bibliographique : guides d'utilisations des outils décisionnels. 2007 http://www.loria.fr/~peguiron/html/profile_fset.html*

[26] GOGLIN J.-F., *La construction du datawarehouse : du datamart au dataweb, Paris, 2001.*

[27] INMON-WILLIAM.-H., *Building the data warehouse, New York, 2002.*

[28] JARKE M., LENZERINI M., VASSILIOU Y., VASSILIADIS P., *Fundamentals of Datawarehouse, Springer 2002.*

[29] KIMBALL R., *The data warehouse toolkit, John Wiley and Sons, 1996.*

[30] KIMBALL R., CASERTA J., *The data warehouse ETL Toolkit Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data,2004*

[31] 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.*

[32] L'Essentiel d'Unified Modeling Language (UML), http://madchat.org/coding/other/CSI_UML_2003.pdf.

[33] LE HENAFF D., DUCLOY J., DUCASSE J.-P. , *Métadonnées pour une cyberinfrastructure de la recherche, 2006.*

[34] MALLERAY E., *Meta données et analyses multidimensionnelles à travers les hypercubes. 2008 http://www.loria.fr/~peguiron/html/profile_fset.html.*

[35] Manuel d'utilisation de l'infocentre PILOTAGE,

http://ftp.amue.fr/documents_publics/apogee/II_doc_fonctionnelle/D_pilotage/MANINFO.doc.

[36] MARTRE, H. *Intelligence économique et stratégie des entreprises, Rapport du Commissariat Général au Plan, Paris, La Documentation Française, 1994.*

[37] MULLER P.-A., *Modélisation objet avec UML, Eyrolles 1997.*

[38] MULLER P.-A., *Représentation des vues d'architecture avec UML,*

http://magda.elibel.tm.fr/refs/UML/architecture.pdf.

[39] NATAF J.-B., *Structure de l'entrepôt de données de pilotage, 2001, http://www.cpu.fr/Telecharger/NatafP6StructureEntrepotDeDonnees.pdf.*

[40] NOMINE B., *ESUP portail : espace numérique de travail pour tous, Nancy 2004.*

[41] PEGUIRON F., *Accès à l'information sur Internet, pratiques et tendances des utilisateurs : dans un contexte de documentation électronique, Université de Nancy-Metz, 2001.*

[42] PEGUIRON F., *Application de l'Intelligence Economique dans un Système d'Information Stratégique universitaire : les apports de la modélisation des acteurs, 2006 .*

[43] 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.*

[44] 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.*

[45] 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.*

[46] PEGUIRON F., THIERY O. *Recensement des besoins de l'utilisateur d'un SIS universitaire via un entrepôt de données en open sou*rce , *Congrès WSST'07 Marrakech Octobre 2007.*

[47] PEGUIRON F., THIERY O. *Analyses multidimensionnelles de contenus documentaires dans un ENT au service de l'acteur enseignant chercheur, Colloque International sur le Document Electronique (CIDE 10), Nancy INIST, Juillet 2007.*

[48] PEGUIRON F., THIERY O. *Un entrepôt de données vu comme une base de connaissance intégrant la modélisation des acteurs : application au système d'information universitaire ISKO 6ème colloque du chapitre français, Toulouse, Juin 2007.*

[49] 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, DRT SIO, Université de Nancy 2, 2002.*

[50] PARREND P., FRENOT S., HOHN S., *PRIVACY-Aware Service Integration, 2007.*

[51] PERENON P., *ProfilDoc : un Système de Recherche d'Information scientifique, 2006.*

[52] REVELLI C., *Intelligence stratégique sur Internet : comment développer des activités de veille et d'intelligence économique sur le web, 2000.*

[53] ROCHFELD A., MOREJON J., *La Méthode Merise, Tome 3, Gamme opératoire, Editions d'Organisation, 1989.*

[54] RUMBAUGH J.*, et al., Modélisation et conception orientées objet, Masson, 1995.*

[55] 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.

[56] TESTE O.*, Modélisation et manipulation d'entrepôts de données complexes et historisées, http://www.irit.fr/recherches/IRI/SIG/personnes/teste/these/intro.pdf.*

[57] THIERY O., DAVID A.*, Modélisation de l'utilisateur : systèmes d'informations stratégiques et intelligence économique, Revue association pour le développement du logiciel (ADELI), 2002.*

[58] THIERY 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.*

[59] THIERY O, *Support de cours recherches avancées en SIS M2 ACSI et SID Miage de Nancy.*

[60] THIERY O*, Support de cours SIS M1 Miage de Nancy.*

[61] P. VASSILIADIS, C QUIX, Y.VASSILIOU, M. JARKE*, Data warehouse process management, Springer 2001.*