

Application de méthodes et outils du Web sémantique pour la gouvernance d'un système d'information industriel

Henry Boccon-Gibod

Electricité de France R&D, Département SINETICS,
1, av du général de Gaulle 92147 Clamart Cedex, <http://www.edf.fr>, <mailto:henry.boccon-gibod@edf.fr>

Résumé

A partir de corpus de textes décrivant l'ensemble des éléments d'urbanisme d'un système hétérogène d'information, des technologies linguistiques de modélisation terminologiques sont utilisées pour décrire ce système selon le formalisme d'ontologies OWL du W3C. L'utilisation de ce formalisme permet d'intégrer formellement les différents points de vue d'une démarche d'urbanisme, et de combler les hiatus de ses représentations selon les formalismes UML. Il répond à moindre coût au premier objectif d'une démarche d'urbanisme qui est de restituer une description cohérente du système aux utilisateurs en y intégrant les connaissances qui ont conduit à sa construction. Il répond aussi à son second objectif qu'est la prescription des contraintes d'intégrité de l'organisation du système et de la validation de leur respect pour que ses applications soient interopérables.

Mots clés : Ingénierie des connaissances, Ontologies, Urbanisme de système d'information, Web sémantique.

1 Introduction

Il est de grands objets issus de l'art des hommes dont la durée de vie dépasse celles des représentations qu'ils s'en font, de leur nature comme de leur organisation. Les villes, les ouvrages d'art, et nombre de monuments appartiennent à cette sorte d'artefact, comme le sont plus modestement ceux que l'industrie produit pour de grandes durées de vie.

La plupart de nos objets usuels sont devenus éphémères, qui échappent donc à cette catégorie. Pour réaliser ces produits l'industrie leur associe des systèmes d'information qui en traitent des représentations. Ces systèmes ne vivent que ce que vivent ces produits.

Apparu à la fin des années 90, le concept désigné par l'acronyme « PLM » [1] (pour « Product Lifecycle Management ») que l'on peut traduire en français par « Maîtrise du Cycle de Vie des Produits » a succédé à celui de « PDM » (« Product Data Management », ou « Gestion de

données de produits ») pour tracer les informations des produits de l'industrie.

Ces technologies, et leurs infrastructures aidant, c'est à une modification profonde des productions industrielles à laquelle nous avons assisté - sinon participé - ces dernières années. Des réductions fortes de coûts et de délais de maturation comme de production, des fonctionnalités accrues et la traçabilité en sont les conséquences positives. D'autres conséquences sont moins favorables, qui sont d'ordre social et culturel : la concentration sur un personnel réduit des connaissances nécessaires pour mettre un produit sur le marché, et à l'opposé, pour la plupart des utilisateurs un degré de connaissances restreint voire nul des principes et modalités de fonctionnement de ces nouveaux objets complexes.

Les grands ouvrages de production d'électricité, les chemins de fer, les avions, objets de l'industrie humaine dont la grande durée de vie, dépasse celles de leurs systèmes d'information sont tributaires de contraintes particulières pour prétendre aux mêmes services.

A l'instar des villes dont les habitations, les circulations et la gestion évoluent, ces objets sont représentés sous diverses formes, dont les modalités diffèrent de celles des objets manufacturés : elles sont hétérogènes, souvent indépendantes, et successivement ajoutées et/ou remplacées pour répondre à des problématiques initialement imprévues. Appliquer à ces sujets des principes d'intégration des systèmes d'information utilisés dans les industries manufacturières permet d'en retirer les mêmes bénéfices. En revanche leur mise en œuvre relève d'une démarche d'une autre nature. Les modèles d'information des produits manufacturés courants sont définis dès leur élaboration, avec un nombre limité de points de vue. De leur côté les installations industrielles tendent à être « modélisées » régulièrement, chaque fois qu'un sujet de préoccupation nouveau le requiert.

La cartographie d'ensemble des applications qui servent les multiples activités concourant avec efficacité à chaque étape de la vie d'une centrale nucléaire évoque moins le plan de Brasilia que celui des hauteurs de Rio de Janeiro.

L'urbanisme, au sens de l'organisation de systèmes d'information est donc une préoccupation cyclique des responsables informatiques des divisions ingénierie et production d'Electricité de France.

2 Problématique

Pour nombre de nos objets courants, (téléphone portable, carte bleue, véhicule, titre de transport...), la cohérence constante de l'ensemble de leurs représentations au sein d'un système est un service désormais normalement attendu, d'autant que cette prouesse est indispensable à leur fonctionnement.

Cette performance qui n'étonne plus guère n'a en fait été réalisée qu'assez récemment et ce dans des contextes relativement restreints, de complexité réduite.

Pour des installations industrielles dont les risques doivent être impérativement maîtrisés, la cohérence des représentations a toujours été une obligation réglementaire, autrement plus complexe à établir. Cela est du à la multiplicité des métiers qui interviennent, et l'enchevêtrement de leurs activités en constante évolution.

Ces exigences et les procédures pour les respecter sont antérieures à l'informatisation de la plupart des activités. Elles se sont traduites par des processus documentaires rigoureux. Tant que l'on est resté dans une logique documentaire matérielle, l'informatisation des activités n'a guère changé ces pratiques, exigées qu'elles étaient par les autorités réglementaires, telles que l'Autorité de Sécurité Nucléaire.

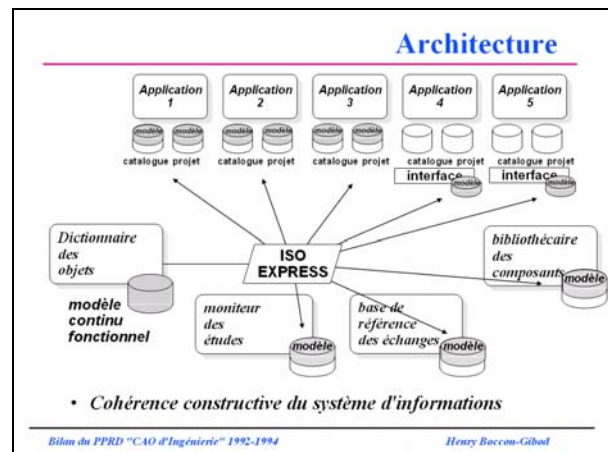
Lorsque tous les objets métiers intermédiaires entre activités qui règlent la vie d'une installation sont matérialisés par des documents imprimés, la question de l'interopérabilité entre les applications informatiques qui servent ces activités paraît secondaire.

Ces applications ont pu se développer indépendamment, et la question de leurs modèles de données et modèles d'information sous-jacents ont rarement été une contrainte. Ainsi se sont agrégés les faubourgs des villes lorsque la démographie les a poussées à s'étendre.

S'il existe quelques îlots d'applications connexes dont les relations sont dématérialisées, c'est avec des productions documentaires parallèles, ce pour le respect de la réglementation. A ces exceptions près, l'essentiel des applications ne coopèrent pas directement, et la mise en cohérence des informations reste une démarche manuelle, réalisée sur la seule base de documents. Les coûts et délais induits par l'établissement de cette cohérence sont élevés, et la réactivité de l'entreprise s'en ressent.

Les projets de nouvelles installations sont a priori les plus simples à traiter, leur élaboration étant semblable à celle d'un produit manufacturé. Les méthodes et technologies de gestion de données techniques qui s'y appliquent permettent de synchroniser les multiples activités de conception menées en parallèle.

Un projet de R&D réalisé au début des années 90 avait permis de définir et de valider les fonctions d'un système d'information destiné aux études d'ingénierie en architecture industrielle. Une fonction d'organisation, dite à l'époque « bureau des modèles » chargée d'établir un « dictionnaire des objets » avait été identifiée. Héritant de la gestion des connaissances des règles et pratiques d'ingénierie traditionnellement contenues dans des manuels d'ingénierie de référence, il définit les conditions d'interopérabilité entre applications.



(FIG. 1 : vision en 1992 des fonctions d'un système d'information d'études.)

Si cette démarche fonctionne bien dans les phases initiales d'un projet, son caractère autoritaire ne peut être conservé dans les phases suivantes, qui sont tributaires de l'insertion du nouveau projet dans un ensemble existant. Il est vain de prétendre y imposer un modèle commun.

Il n'y a encore que fort peu de communications de données informatiques entre les activités d'architecture industrielle et les activités d'exploitation et de maintenance, même si Electricité de France a la particularité de les intégrer. Ces activités d'exploitation et de maintenance ont été servies par leurs propres applications informatiques, qui ont dû modéliser les installations à leurs façons, en fonction de leurs objectifs, qui diffèrent en particulier de ceux de l'ingénierie.

Sortir des contingences documentaires, tirer parti du dynamisme que permettent les technologies actuelles pour assurer la cohérence des représentations suppose de concilier les points de vue, faute de pouvoir raisonnablement les unifier.

On se trouve ainsi face à un problème qui n'est plus de définir un modèle unique imposé pour la définition des applications, mais de fédérer la coexistence des multiples façons existantes de représenter leurs objets communs.

3 Des formalismes concurrents de modélisation

Les formalismes de modélisation Schlaer & Mellor, OMT, Booch qui ont convergé vers le formalisme UML sont d'abord destinés à la spécification de nouvelles applications.

Indépendamment, le langage EXPRESS a été défini pour rendre interoperables des descriptions de produits.

Nées entre-temps des objectifs du Web définis par le W3C, les technologies XML ont introduit la notion de modèle de validation structurelle des documents, qu'ils soient classiques ou simplement supports d'échanges d'information entre applications. Les DTD, les schémas XSD, et leurs concurrents Schematron ou Relax NG ont ainsi introduit d'autres modes de modélisation, partiellement concurrents de la modélisation UML.

Lorsque le W3C a anticipé les limites du Web syntaxique et de ses moteurs de recherche, et en se préoccupant de bruit et de silence a établi les concepts du Web sémantique, d'autres formalismes de modélisation sont apparus. Langages d'ontologie formelle, RDFSchema pour RDF, puis OWL font resurgir les logiques de prédicats de l'Intelligence Artificielle. (Celle-ci y perd son caractère de « programmation en vrac », qui tend à réapparaître avec XSL...).

De leur côté les spécialistes du traitement du langage ont défini plusieurs générations de modèles terminologiques, (Olif, Geneter, Martif, dialectes XML qui convergent vers TMF) et se préoccupent aussi d'ontologie, en séparant concepts et termes pour répondre notamment à des problématiques multilingues.

Issues aussi du traitement du langage naturel, d'autres approches se sont développées qui ont donné d'autres formalismes de représentation de réseaux sémantiques que sont les Topic Maps, venus de la démarche HyTime.

4 Approches méthodologiques et démarches pratiques

4.1 Des instances gouvernementales

Des Etats, dont l'Etat Français, ont adopté des démarches de rationalisation des données traitées dans leurs documents. L'initiative de l'Atica, transformée par la suite en Adae et désormais en DGME (Direction générale pour la modernisation de l'Etat) s'est concentrée sur le partage de ressources entre modèles documentaires. Ceci permet d'unifier l'expression des informations potentiellement redondantes traitées dans les différentes administrations.

4.2 Des organismes de normalisation

Des organismes de normalisation qui traitent de sujets qui concernent notre problématique d'urbanisme ont utilisé des formalismes de modélisation différents, qui sont des alternatives possibles.

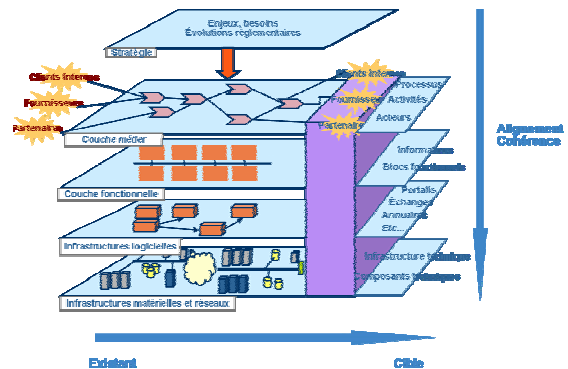
Pour les échanges de « données de produit », Le TC4 de l'ISO a défini le langage EXPRESS, pour exprimer les modèles d'information des différents protocoles de la norme ISO 10303 STEP et aussi la norme ISO 13584 P-LIB. Il en existe maintenant des outils de conversion en schémas XML...

Du côté du DMTF le modèle CIM (adopté par les Electriciens) a été défini en langage UML, selon la démarche MDA promue par l'OMG... Alors, en suivant les travaux de l'UN-CEFACT, il s'est avéré nécessaire de définir des « profils » pour convertir des modèles UML en schémas de validation XSD, ce qui s'est traduit par 192 règles non formelles, qui ne permettent donc pas encore de validation.

En revanche, le modèle RM/ODP, (Reference Model/Open Distributed Processing) concrétisée à Electricité de France par la méthodologie DASIBAO [2] qui a l'intérêt de séparer proprement les niveaux de description, n'impose pas de formalisme pour la supporter.

4.3 Des groupements professionnels

Le CIGREF propose une démarche cartographique d'urbanisme en couches. Son intérêt est didactique pour des managers : elle illustre par une arborescence d'emboîtement l'analyse fonctionnelle des activités considérées, et d'y insérer des informations sur les applications, les flux de données et les objets métier traités.



(FIG. 2 : démarche CIGREF)

4.4 Des mésusages de logiciels de présentation

En raison de leur grande diffusion et de leur commodité, les acteurs métiers tendent à exprimer leurs concepts au moyen de logiciels de présentation, qui sont ainsi devenus un moyen privilégié de partage de connaissances ; la pensée s'y projette en un enchevêtrement de listes, de boîtes et

de flèches de toutes formes. Paradoxalement seuls des humains peuvent tenter de comprendre ces expressions informatiques de connaissances, qu'aucun ordinateur ne peut effectivement exploiter !

4.5 Objectifs recherchés

Trois motivations de modélisation de la connaissance sous-tendent une démarche d'urbanisme :

1. obtenir des descriptions cohérentes compréhensibles de l'ensemble du système d'information, depuis ses principes d'organisation jusqu'aux détails jugés nécessaires de sa réalisation,
2. en prescrire les contraintes (règles de validité, conditions d'interopérabilité, etc.),
3. simuler le fonctionnement de l'ensemble pour en prouver l'intégrité.

On vise plus précisément ici à offrir un jeu de services quotidiens aux utilisateurs métiers; situer leurs activités et informer leurs données dans l'ensemble et en valider la qualité structurelle. Dans ce sens notre démarche d'urbanisme cherche à associer :

1. la validation des connaissances des métiers associées aux applications servant leurs activités,
2. la fédération des modèles qui spécifient ces applications et des modèles qui en valident les données,
3. l'identification des objets métiers qui transcendent le système d'information,
4. l'identification des transformations nécessaires des informations entre activités métiers.

5 Exploration méthodologique et technologique

L'idée que nous proposons d'explorer est d'utiliser le formalisme des ontologies pour décrire dans un référentiel d'urbanisme l'organisation d'un système d'information d'entreprise.

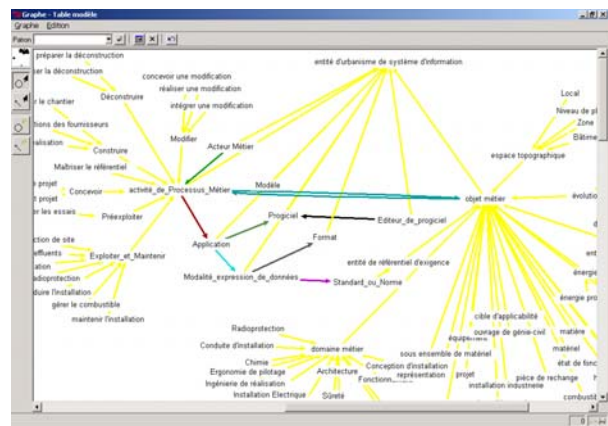
Les méthodes classiques de démarche d'urbanisme mettent en œuvre les différents diagrammes UML définis par l'OMG, en commençant par des Use Case, des diagrammes de classe, d'activité, et maintenant de coopération. De fait, cela ne répond pas réellement à notre problématique, qui n'est pas de spécifier un nouveau système d'information homogène, mais de décrire et valider la cohérence d'un système hétérogène existant et de fournir, à partir de cette description, des services de qualité à ses utilisateurs.

Chaque diagramme UML a son propre mode opératoire ; cependant rien dans le formalisme n'oblige à associer les informations projetées d'un diagramme à l'autre. La formalisation du savoir-faire associé aux relations entre diagrammes n'est pas aujourd'hui traitée. D'autre part les mécanismes d'instanciation qui sont proposés ont été définis avec un objectif de simulation et non de description exhaustive.

5.1 De la terminologie à l'ontologie

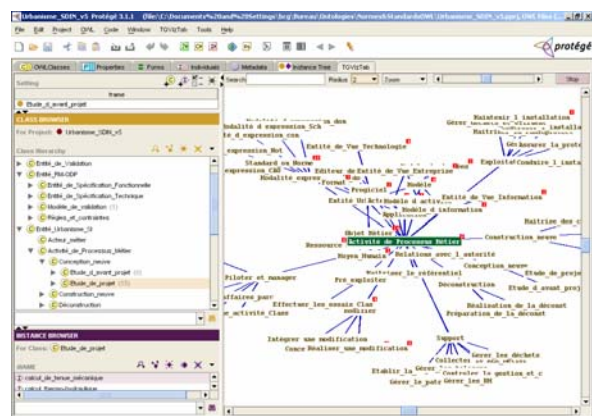
La première étape de notre démarche d'urbanisme commence par la collecte et la cartographie aussi exhaustive que possible de ce qui existe. Cette collecte se matérialise par une collection hétérogène de documentations, sous formes de textes et/ou de tableaux et par des expertises exprimées par des diagrammes dessinés au moyen de logiciels de présentation de diapositives.

A partir de tout ou partie de cet ensemble il est possible de constituer un corpus dont on est capable d'extraire une terminologie structurée. En l'occurrence notre chaîne Lexter® [3] [4] ou désormais Exter® [5] pour l'extraction et notre logiciel Worldtrek [6] pour l'édition permet de construire un réseau terminologique. Il est possible, par une expertise d'en remonter un réseau sémantique de concepts.



(FIG. 3 : Worldtrek, structuration terminologique)

Pour ce faire après une première expérience de publication des résultats sous forme de Topic Maps (XTM), nous avons expérimenté des modalités d'expression du travail de modélisation terminologique compatibles avec le langage d'ontologie formelle OWL.



(FIG. 4 : édition d'ontologie résultante dans Protégé®)

Les résultats des travaux du projet SYNTAX qui ont défini le modèle TMF [7] de terminologie, en séparant formellement « termes » et « concepts » sont favorables à l'intégration entre terminologie structurée et ontologie formelle. Elle ne nécessite a minima que de spécialiser les concepts en « classes » et en « individus », et de formaliser trois liens terminologiques : un lien de spécialisation entre classes, un lien d'individuation entre classes et individus et un lien de désignation entre termes et concepts. Le résultat (encore partiel à ce jour) est exploitable via n'importe quel éditeur d'ontologie tel que Swoop® ou Protégé®.

Dans le cas ici étudié d'une étude d'urbanisme, les activités métier sont les premiers concepts utiles à formaliser, avec ceux d'acteurs, de ressources, et d'objets métier en entrée et en sortie de chaque activité.

L'équivalent d'un modèle SADT ou de sa forme normalisée IDEFO s'exprime facilement selon le formalisme OWL avec des classes correspondantes « Activité », « Acteur », « Ressource », « Objet Métier » et des liens sémantiques formellement explicités. Les différents points de vue du modèle RM/ODP et des cartographies du CIGREF se projettent sans difficulté, pour autant qu'il ne s'agit que d'expliciter les connaissances qu'elles requièrent. A ces concepts se raccordent ceux que l'étude terminologique a fait remonter, dont une expertise métier d'un expert généraliste, aidé des fonctionnalités d'accès au corpus d'origine permet de catégoriser et de peupler.

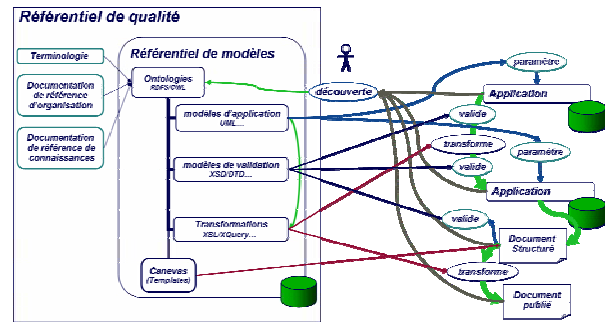
Des applications identifiées y apparaissent comme ressources des activités, supportées par des logiciels, développés par des éditeurs de logiciels... Des modèles d'information leur sont associés avec éventuellement des références aux normes et standards métiers implémentés. Ensuite les applications sont supportées par des serveurs, reliés par des réseaux, etc.

5.2 Services apportés par l'ontologie

L'ontologie résultante est un modèle hybride, peuplé d'objets concrets pour les utilisateurs. On y retrouve formellement un ensemble d'éléments associés connexe, qui sont ordinairement séparés dans les différents diagrammes indépendants du langage UML. Cette description formalise et détaille la superstructure du système d'information, à la fois dans ses concepts d'organisation (l'ontologie proprement dite) et dans son contenu décrit (les individus qui la peuplent). Elle peut s'étendre à volonté et sans discontinuité du modèle d'activité de la partie considérée de l'entreprise, jusqu'au détail des serveurs d'applications, des postes de travail et de leurs utilisateurs.

Le modèle d'expression générique des ontologies permet d'intégrer les cinq différents points de vue (Entreprise, Information, Traitement, Ingénierie et Technologie) du modèle RM/ODP. Il associe les descriptions des points de vue et définit les prescriptions que l'on en attend (telles que les contraintes de transposition entre points de vue, et les contraintes d'intégrité au sein d'un point de vue donné).

Le premier service – décrire – qu'apporte cette démarche est celui d'un répertoire de modèles : il permet de découvrir, pour une action donnée, les services applicatifs à mettre en œuvre et les modèles validants des jeux d'informations qui les traitent. Il les met à disposition avec les droits d'accès correspondants sur le réseau intranet de l'entreprise.



(FIG. 5 : services offerts par un référentiel décrit par une ontologie)

Le second service – prescrire – est de vérifier l'intégrité d'un système d'information aussi hétérogène soit-il, pour en gager le bon fonctionnement.

Ceci suppose que soit formellement intégrée sa description, et de pouvoir y attacher les règles de son intégrité.

Le troisième service – simuler – de preuve du fonctionnement du système reste à réaliser.

Un service supplémentaire – contextualiser – qu'une telle formalisation devrait donner procède de la finalité même du web « sémantique ». La contextualisation du paramétrage des systèmes de consultation, que ne peut apporter le web « syntaxique », réunit l'ensemble des informations définies comme utiles pour chacune des activités des métiers, sans bruit et sans silence.

6 Conclusions et perspectives

6.1 Retour d'expérience

Cette démarche est initialement motivée par un souci d'ergonomie. Suite à la collecte de l'état en cours d'un système d'information, elle vise à faciliter le travail de structuration et de mise en cohérence de cette masse d'informations. Sa formalisation dans une ontologie s'avère efficace pour un coût minimal.

Peu de concepts sont à expliquer pour la collecte, dont les restitutions sémantiques sont compréhensibles par des acteurs non avertis ; le maintien de la connexion avec les corpus de connaissances exprimés en langage naturel est un atout pour la validation des descriptions.

La plasticité du modèle facilite la mise au point de la structure des descriptions à partir de l'expertise des acteurs, ce qui est aussi un atout important, de même que la possibilité d'héritage multiple qu'autorise le formalisme.

L'efficacité de cette forme de modélisation d'urbanisme vient de ce qu'elle permet de relier des points de vue et des niveaux d'abstraction différents sans obstacle technique.

A ce jour cette approche de l'urbanisme répond principalement aux premiers services que l'on en attend, soit de décrire précisément l'organisation. Elle commence à répondre au second qui est de prescrire des conditions d'intégrité du système d'information.

6.2 Insuffisances en l'état de l'art

Les principaux inconvénients actuellement identifiés viennent de l'imaturité des technologies de traitement et de restitution des contraintes décrites. Les technologies Web d'exploitation d'ontologies sont encore rares, peu industrialisées et d'emploi complexe.

La demande devrait être forte de restitutions graphiques cohérentes, éditables, qui nécessitent de lier une description sémantique formelle à des représentations graphiques. Ceci nécessiterait par exemple des adaptations du langage SVG d'expression. Il serait nécessaire d'y attacher des possibilités d'édition permettant d'exprimer dans sa richesse le contenu d'une ontologie. Il serait en particulier nécessaire de pouvoir intégrer des bibliothèques de symboles adaptés à la compréhension des acteurs, dotés de ports de connexion typés, pour exprimer la portée des relations ontologiques. Associé à des outils tels que Amaya proposé par le W3C, ceci permettrait de supprimer les inconvénients liés au mésusage habituel de logiciels de présentation.

Simuler pour le prouver le fonctionnement du système nécessite aussi d'autres fonctionnalités que celles offertes par la logique de description ou les langages de règles.

Si ceci est aujourd'hui du ressort des ateliers de génie logiciel utilisant le langage UML, celui-ci ne devrait-il pas évoluer pour intégrer les notions des ontologies ?

6.3 Extension de la démarche

Après la description de l'organisation globale d'un système d'information la méthode peut s'étendre à chacune des applications.

Pour en définir les modalités d'interopérabilité, nous avons commencé à expérimenter la description des modèles d'information à partir des modèles de données via des ontologies formelles. L'objectif est ici de générer automatiquement des interfaces d'échange. La méthode expérimentée fait appel à des ontologies intermédiaires entre description des modèles d'information de chaque application, peuplées via des méthodes d'alignement.

L'association à la précédente description d'ensemble du système d'information de ces ontologies d'application et d'interopérabilité permettrait d'en compléter les services, depuis l'organisation du système jusqu'au détail de chaque sorte de données.

7 Références

- [1] Figay N. (2005), e-Collaborative Concurrent Engineering and PLM services, OMG documents.
- [2] Picault A., Bedu P., Le Delliou J., Perrin J., Traverson B. (2004), DASIBAO, Une démarche d'architecture des SI fondée sur les standards MDA et RM/ODP, NOTERE'2004.
- [3] Bourigault D. (1994), Lexter, un logiciel d'extraction de terminologie. Application à l'acquisition des connaissances à partir de textes. Thèse en informatique linguistique, Ecole des hautes Etudes en Sciences Sociales, Paris.
- [4] Bourigault D. (1994), Extraction et structuration automatiques de terminologie pour l'aide à l'acquisition des connaissances à partir de textes. In Actes du 9ème congrès Reconnaissance des Formes et Intelligence Artificielle (RFIA'94).
- [5] Cyril Chantrier, (2005), EXTER: A Breakthrough Solution for Efficient Terminology Extraction, ASLIB CONFERENCE.
- [6] Abbas Y., Picard M-L (2000), With WORLDTREK Family, Create, Update and Browse your Terminological World, C 2000 2nd International Conference on Language Resources & Evaluation.
- [7] Romary, L. (2001), An abstract model for the representation of multilingual terminological data: TMF - Terminological Markup Framework. In TAMA 2001. (Anvers, Belgium). 2001.