



HAL
open science

Atelier "Systèmes d'Information et de Décision pour l'Environnement" (SIDE 2008)

François Pinet, André Miralles

► **To cite this version:**

François Pinet, André Miralles. Atelier "Systèmes d'Information et de Décision pour l'Environnement" (SIDE 2008). Association Inforsid. pp.84, 2008. lirmm-00352389v1

HAL Id: lirmm-00352389

<https://hal-lirmm.ccsd.cnrs.fr/lirmm-00352389v1>

Submitted on 13 Jan 2009 (v1), last revised 22 Jul 2009 (v2)

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.

Congrès Inforsid
27 mai 2008

Atelier « Systèmes d'Information et de Décision pour l'Environnement »

Actes de l'atelier

Edités par :

François Pinet, TSCF/UMR TETIS, Cemagref, Clermont Ferrand
André Miralles, UMR TETIS, Cemagref, Montpellier

Atelier « Systèmes d'Information et de Décision pour l'Environnement » Inforsid - 27 mai 2008

La recherche en informatique et en systèmes d'information offre depuis des années des solutions de plus en plus performantes pour relever les récents challenges environnementaux. Les données environnementales acquises sont de plus en plus nombreuses et sont aujourd'hui structurées et analysées au sein de Systèmes d'Information et d'Aide à la Décision.

L'objectif de l'atelier est de présenter comment les toutes dernières avancées de la recherche en systèmes d'information ou de décision s'appliquent au domaine environnemental. L'atelier était ouvert aussi bien à la présentation de travaux de recherche déjà appliqués au contexte de l'environnement, qu'à des réflexions plus prospectives sur les possibilités d'utilisation d'un produit de la recherche en informatique pour une application environnementale.

La journée d'atelier a été découpée par thème informatique en trois sessions. La première porte sur les ontologies, la seconde s'intéresse aux systèmes intégrés et d'aide à la décision, et la troisième concerne les systèmes communicants.

Les huit articles de l'atelier ont été écrits par des auteurs de cinq pays différents (Allemagne, Autriche, Canada, Chine, France). On retrouve plusieurs domaines environnementaux. Il est par exemple question de géologie pour la capture et le stockage de dioxyde de carbone, de géotechnique avec un système d'aide pour la représentation des connaissances. Des travaux sur la cartographie des risques d'inondation sont aussi présents. Un autre article traite de la modélisation d'ontologies en agriculture. Un autre concerne la conception d'un SAD reposant sur une base de connaissances territoriale, pour l'Agence de gestion du Bassin de Souss-Massa. Différents travaux relatifs à l'acquisition ainsi qu'à l'intégration de données environnementales sont aussi décrits.

La qualité des travaux laisse présager une journée d'atelier particulièrement enrichissante. Nous remercions par ailleurs tous les membres du comité de programme pour leur excellent travail.

Les organisateurs de l'atelier
François Pinet, TSCF/UMR TETIS, Cemagref, Clermont Ferrand
André Miralles, UMR TETIS, Cemagref, Montpellier

Atelier « Systèmes d'Information et de Décision pour l'Environnement » Inforsid - 27 mai 2008

Organisateurs :

François Pinet, laboratoire TSCF/UMR TETIS, Cemagref, Clermont Ferrand
André Miralles, UMR TETIS, Cemagref, Montpellier

Comité de programme :

Ahmed Lbath, LIG, Grenoble	Michel Schneider, LIMOS, Clermont Fd
Catherine Roussey, LIRIS, Lyon	Myoung-Ah Kang LIMOS, Clermont Fd
Éric Jallas, Cirad et Société ITK, Montpellier	Petra Papajorgji, IFAS, Université de Floride - US
Jean-Christophe Desconnets, IRD, Montpellier	Sylvie Servigne, LIRIS, Lyon
Jean-Pierre Chanet, TSCF/TETIS - Cemagref, Clermont Fd	Thierry Badard, CRG, Université Laval, Canada
Mathieu Roche, LIRMM, Montpellier	Vincent Abt, TSCF/TETIS - Cemagref, Clermont Fd
Michel Passouant, Cirad, Montpellier	

Programme :

8h45-9h15 Accueil des participants

9h15-10h30 Session Ontologies

- *Annotation à Base Ontologique de Modèles : Application aux Modèles en Géologie pour le Stockage de CO2* - L.S. Mastella, Y. Ait-Ameur, M. Perrin, J.F. Rainaud (pp.1-10)
Ecole des Mines de Paris, Ecole Nationale Supérieure de Mécanique et d'Aérotechnique (Futuroscope), Institut Français du Pétrole (Reuil-Malmaison)

- *Les Ontologies et leurs Applications en Agriculture*- C. Roussey, M.A. Kang (pp.11-15)
Laboratoire d'Informatique en Image et Systèmes d'information (Villeurbanne), Laboratoire d'Informatique, de Modélisation et Optimisation des Systèmes (Aubière)

Pause

11h-12h30 Session Systèmes Intégrés et d'Aide à la Décision

- *Système Intégré d'Aide à la Décision en Gestion par Bassin, de la Conception à l'Opérationnalisation : Etude de Cas* - C. Prével (pp.17-30)
Groupe d'Etudes Interdisciplinaires en Géographie et Environnement Régional (Montreal - Canada)

- *Sémiologie Graphique Expérimentale et Cartographie du Risque d'Inondation (SGE)* - K. Serrhini, J. Rochman, S. Fuchs, W. Dorner, K. Spachinger (pp.31-40)
Centre Interdisciplinaire Cités, TERritoires, Environnement et Sociétés (Tours), Institute of Mountain Risk Engineering (Vienne - Autriche), University of Applied Sciences (Deggendorf - Allemagne)

Repas

14h-15h30 Session Systèmes Intégrés et d'Aide à la Décision (suite)

- *Intégration de Données hétérogènes pour un SAD Environnemental* - R. Kissi, J.C. Desconnets, T. Libourel (pp.41-55)
Institut de Recherche pour le Développement (Montpellier), Laboratoire d'Informatique, de Robotique et Microélectronique de Montpellier

- *Un Système d'Aide pour les Sciences de la Terre : un Exemple Géotechnique* - N. Faure (pp.57-66)
Centre de Recherche de l'Institut d'Administration des Entreprises de l'Université Lyon 3

Pause

16h-17h30 Session Systèmes Communicants

- *Une Architecture orientée Services pour la Gestion Sémantique des Données Géologiques pour le Stockage de CO2* - N. Belaid, I. Ait-Sadoune, C. Fankam, S. Jean, Y. Ait-Ameur, G. Pierra, J.F. Rainaud (pp.67-76)
Ecole Nationale Supérieure de Mécanique et d'Aérotechnique (Futuroscope), Institut Français du Pétrole (Reuil-Malmaison)

- *Système pour l'Acquisition et la Gestion de Données Environnementales* - G. De Sousa, H.Y. Zhou, K.M. Hou, C. De Vault, J.P. Chanet (pp.77-84)
Laboratoire Technologies et Systèmes d'information pour les agrosystèmes / UMR Territoire, Environnement, Télédétection, Information Spatiale (Aubière), Harbin Institute of Technology (Chine), Laboratoire d'Informatique, de Modélisation et Optimisation des Systèmes (Aubière)

- *Clôture et bilan de la journée*

Annotation à base ontologique de modèles: application aux modèles en géologie pour le stockage de CO2

Laura Silveira Mastella* — **Yamine Ait-Ameur**** — **Michel Perrin***
— **Jean-François Rainaud*****

* *Ecole des Mines de Paris, 60, boulevard Saint-Michel, 75272, Paris*
{laura.mastella,michel.perrin}@ensmp.fr

** *LISI - ENSMA and University of Poitiers, BP40109, 86961 Futuroscope Cedex*
yamine@ensma.fr

*** *Institut Français du Pétrole - DTIMA, 1-4 Av. de Bois-Préau, 92852 Rueil-Malmaison, j-francois.rainaud@ifp.fr*

RÉSUMÉ. Dans le cadre des projets de stockage de CO2, le problème de l'intégration de données hétérogènes reste un grand défi, compte tenu de la nature multidisciplinaire de cette activité. Il est difficile pour les professionnels d'intégrer les résultats de leurs interprétations, ou de poser une requête qui demande des informations issues de plusieurs disciplines. Pour traiter ce problème, nous proposons une approche de l'annotation sémantique de modèles métiers avec des concepts d'ontologies locales (OL). Les concepts des OL sont subsumés par une ontologie globale (OG), qui contient les concepts géologiques partagés entre différents disciplines.

ABSTRACT. The problem of the integration of heterogeneous data within projects on CO2 storage is a great challenge, considering the multidisciplinary nature of this activity. Professionals have difficulty for integrating the results of their interpretations or for querying information originated from different domains. To deal with this problem, we propose to semantic annotate the technical models with concepts of local ontologies (LO). The concepts of the LO are subsumed by a global ontology (GO), which contains the concepts of the basic geology, that are shared between all the disciplines.

MOTS-CLÉS : Annotation de modèles, Intégration d'information, Ontologie, Stockage de CO2.

KEYWORDS: Model annotation, Information integration, Ontology, CO2 Storage.

1. Introduction

Dans le cadre des projets sur la capture et le stockage de CO₂, le problème de l'intégration de données hétérogènes reste toujours un grand défi. L'objectif de ces projets est de construire des *modèles géologiques décrivant des lieux de stockage possibles*. Les processus permettant de construire ces modèles constituent des chaînes de travail au sein desquelles des professionnels originaires de différentes disciplines des sciences de la terre interprètent éventuellement de manière contradictoire les données relatives à un prospect de stockage, en vue de construire différents modèles interprétatifs.

Les différents vocabulaires et terminologies employés par les professionnels de chaque discipline génèrent une importante hétérogénéité de données. En conséquence, il n'existe pas de solution pour communiquer les résultats des interprétations des modèles et il n'est pas possible de poser une requête qui requiert des informations issues de plusieurs modèles différents, ni de fusionner des informations issues d'interprétations différentes.

Pour traiter ce problème, nous proposons ici une approche basée sur *l'annotation sémantique de modèles*. L'objectif est de réussir à extraire des informations à partir de modèles hétérogènes. Nous utilisons des ontologies pour représenter les concepts des domaines, selon la définition classique de Gruber (1995). Nous définissons des *ontologies locales* pour annoter les modèles spécifiques et une *ontologie globale partagée* pour rassembler le vocabulaire commun à tous ces domaines. De plus, en vue de représenter les modèles, les ontologies et les annotations des modèles par des instances de l'ontologie, nous proposons *d'étendre une architecture de base de données ontologique* (BDBO), à savoir OntoDB (Dehainsala *et al.*, 2007).

La suite de cet article est organisée de la façon suivante. Dans la section 2 nous passons en revue les travaux existants en matière d'annotation sémantique. Nous introduisons dans la section 3 les activités de modélisation géologique que nous allons cibler dans ce travail et les problèmes concernant ce domaine. Ensuite, nous exposons dans la section 4 notre proposition d'architecture puis, en section 5 l'implémentation de cette approche sur une extension de l'architecture de la BDBO. La section 6 décrit une application de cette proposition à un cas d'étude réel appartenant à la chaîne de géo-modélisation. Nous concluons enfin et présentons les perspectives de recherche issues de ce travail.

2. État de l'art : l'annotation sémantique

Les travaux les plus significatifs concernant l'annotation sémantique sont relatifs à l'*annotation de documents de textes*. Les connaissances d'une compagnie se trouvent en effet majoritairement consignées dans des documents de projets, rapports et d'autres sources textuelles (Uren *et al.*, 2006). En plus de l'annotation des documents, on trouve quelques travaux traitant de l'*annotation de modèles*. Le travail de (Lin *et al.*, 2006) décrit un modèle d'annotation de patrons de processus, en vue d'améliorer la réutilisation des processus dans chaque nouvelle activité. (Schroeter *et al.*, 2003) décrivent un prototype qui permet l'annotation de vidéos. Ils ont étendu le modèle d'annotation de W3C Annotea (<http://www.w3.org/2001/Annotea/>), pour supporter l'annotation des documents audiovisuels en proposant de changer le contexte du *XPointer* employé par Annotea pour lier une annotation à un document.

Plusieurs travaux sur l'annotation d'images (Maillot *et al.*, 2003) et, plus récemment, sur l'annotation des services web ont été proposés (Farell *et al.*, 2006). Mais nous ne connaissons pas de proposition d'annotation des *modèles métiers*, tels que, dans notre cas, des *modèles géologiques*. Plus spécifiquement, on ne dispose pas de technique permettant de compléter les modèles par des commentaires ou des explications formels, ou bien d'attacher plus de sémantique aux données métiers utilisés dans des outils spécifiques. Cela est un besoin essentiel puisque les interprétations réalisées par les géologues lors du traitement d'un modèle géologique, correspond à la connaissance la plus importante introduite dans les modèles. Ce problème sera détaillé dans la section 3. Nous proposons de combler cette lacune en définissant une *architecture d'annotation de modèles basée sur des ontologies*.

3. Étude de cas : la modélisation géologique

La partie du *processus de géo-modélisation* prise en considération dans ce travail comporte les activités suivantes :

- L'**Interprétation Sismique** vise à identifier les premiers objets géologiques. Une image sismique peut être interprétée en identifiant les traces horizontales comme des *réflecteurs* ($r1$, $r2$, $r3$), et les verticales comme des *interruptions* ($int1$ et $int2$) (figure 1(b)). Ces objets sont décrits dans des fichiers de points 3D (fichier XYZ).

- **Modélisation Structurale** : à partir des objets sismiques le modéleur identifie des *structures*, comme un réseau de failles ($f1$ et $f2$) qui interrompent un contact ($c1$) (figure 1(c)). Ce modèle est construit avec un outil propriétaire de géo-modélisation qui représente les objets interprétés dans sa structure de classes interne.

- **Corrélation de puits** : les objets issus du modèle structural (p. ex., le contact $c1$) sont corrélés avec les objets du modèle de puits (les marqueurs $m1$ et $m2$ sur la figure 1(d)) en vue d'ajuster la position spatiale des premiers. Les données de puits sont représentées à l'aide d'un format de description XML nommé *WITSML*.

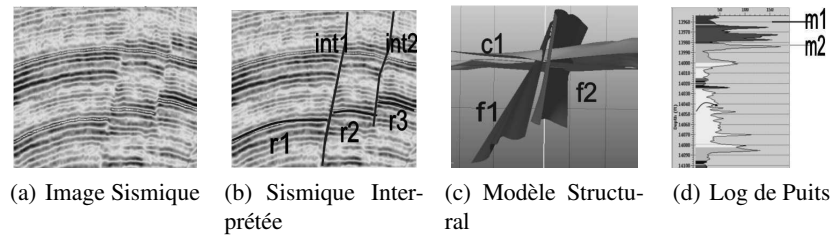


Figure 1. Différents activités de géo-modélisation

La question principale est de déterminer comment une correspondance peut être créée entre des concepts de chacun de ces domaines spécifiques ci-dessus. Cela n'est pas possible actuellement puisque nous ne pouvons pas récupérer la relation entre les objets identifiés dans différentes phases de la chaîne de géo-modélisation. Or une des attentes des utilisateurs est de pouvoir, à n'importe quelle étape du processus, répondre à diverses questions relatives aux objets géologiques pris en considération. Une question typique serait par exemple de déterminer : *quels sont les réflecteurs recoupés par un puits X ?* Cette requête requiert le croisement de deux types d'informations issues respectivement de l'activité *interprétation sismique* et de l'activité *corrélation de puits*. Il n'est pas possible de répondre actuellement à une question de ce type puisque les géologues ne sont pas en mesure de récupérer les relations qui existent entre les objets identifiés dans différentes phases du processus. Une autre difficulté est qu'il n'existe pas de base de connaissances partagée entre les différents outils, ce qui permettrait en effet de stocker les interprétations gérées lors d'un processus de géo-modélisation en vue de pouvoir les récupérer éventuellement.

4. Notre approche

Nous allons maintenant décrire l'approche envisagée pour résoudre ces problèmes. Celle-ci se décline sur trois niveaux conceptuels : (i) les *ontologies*, (ii) les *méta-données* et (iii) le *modèle d'annotation*. Les ontologies, présentées dans la section 4.1, sont utilisées pour formaliser la sémantique des concepts de chaque discipline ou domaine. A l'autre bout de la chaîne de modélisation, on dispose de données métiers, utilisées dans les outils de géo-modélisation, lesquelles exigent aussi d'être formalisées. Nous définissons ces informations en terme de méta-données (données sur les données). Ces modèles sont décrits dans la section 4.2. Ainsi, pour répondre au besoin d'annoter les modèles géologiques par des ontologies, il faut définir un *modèle d'annotation* permettant d'associer les objets du domaine (données issues des modèles) à leurs définitions ontologiques (instances ontologiques) (*cf.* section 4.3).

4.1. Les ontologies en sciences de la terre

Nous disposons déjà de quelques ontologies de domaine formalisées en OWL (<http://www.w3.org/TR/owl-features/>) qui représentent les concepts utilisés par les professionnels dans leurs *domaines spécifiques d'expertise*. Pour cette raison ces ontologies sont appelées *Ontologies Locales (OL)*. Nous présentons ici quelques extraits des ontologies locales suivantes : *Ontologie de la Sismique* (figure 2(a)), *Ontologie de Puits* (figure 2(b)), *Ontologie de la Géologie Structurale* (figure 2(c)).

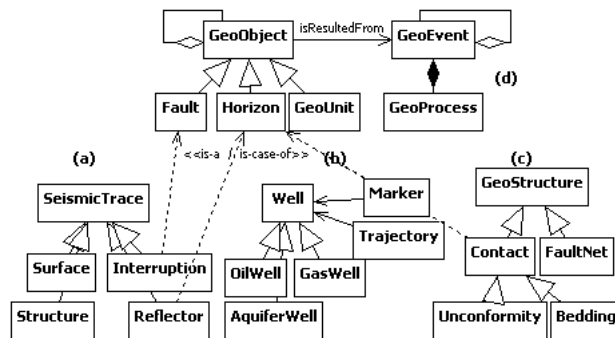


Figure 2. Extraits des ontologies de la géologie

En plus des ontologies locales, nous pouvons identifier des concepts géologiques utilisés tout au long de la chaîne de modélisation. Ces concepts sont employés pour faire référence aux objets qui sont identifiés au début de la chaîne et qui sont utilisés dans les différents modèles géologiques. Cependant, chaque modèle géologique possède sa propre caractérisation de ces concepts. L'annotation consiste à lier ces caractérisations aux concepts des ontologies. Pour cette raison, nous suggérons que les entités dans les divers géo-modèles soient caractérisées en tant qu'*objets géologiques*, ayant une identification unique. Ces objets ont été formalisés comme une *Ontologie Globale (OG)*, qui représente les concepts de la géologie de base auxquels peuvent être rattachés les concepts relatifs aux ontologies locales. Une version de cette *Ontologie de la Géologie de Base* est présentée dans (Mastella *et al.*, 2007). Nous en montrons un extrait sur la figure 2(d).

4.2. Les méta-données des modèles géologiques métier

Ainsi que nous l'avons exposé dans la section 3 chaque activité de géo-modélisation correspond à l'utilisation d'un outil spécifique, qui représente les objets géologiques d'une manière différente : points, surfaces, mailles. Afin de formaliser l'identification unique de ces objets, nous avons décrit les modèles des données utilisés dans chacun des outils de géo-modélisation pour caractériser les objets qu'il mani-

pule. En conséquence, les données produites par ces différents outils sont considérés comme des instances des méta-données. Dans notre étude de cas nous avons identifié les modèles des données correspondant aux tâches citées dans la section 3. Des extraits de ces modèles sont présentés sur la figure 3 : (a) le **modèle de données pour l'interprétation sismique** (modèle de points en 3D, dit *XYZ*), (b) le **modèle de données pour l'analyse de puits** (modèle *WITSML*), et (c) le **modèle de données pour la modélisation structurale** (défini interactivement par observation des exécutions de l'outil de géo-modélisation).

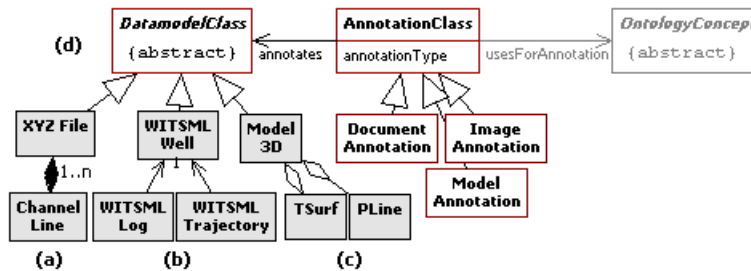


Figure 3. Extraits des modèles de données et le modèle d'annotation

4.3. Le modèle d'annotation

Le modèle d'annotation permet de lier les données des outils de géo-modélisation aux instances ontologiques, en créant une liaison entre la représentation de ces données et le concept de l'ontologie locale qui lui fournit une sémantique à base ontologique. Le modèle d'annotation est constitué de 3 entités (figure 3(d)) : (i) la super-classe de toutes les classe des méta-données (*DatamodelClass*), (ii) la classe que récupère l'URI d'un concept de **OL** (*OntologyConcept*) et (iii) la classe qui concrétise l'association entre les identificateurs des deux (*AnnotationClass*). Cette classe d'association peut inclure aussi des informations supplémentaires, comme l'auteur, la date, et le type d'annotation (*interprétation, géo-localisation, etc.*).

4.4. L'architecture proposée

Cette architecture permet d'exploiter des annotations à base d'ontologies opérées sur les modèles géologiques (*cf.* figure 4). Nous proposons que les *données* (fichiers, tables) utilisés dans les outils de géo-modélisation (1) soient représentés comme des instances des *modèles de données* (2). Ces instances seront annotées par des instances des *ontologies locales* (3) (en employant le modèle d'annotation). Finalement, les instances des OL sont subsumées par des concepts globaux (4). Cette architecture vise 2 objectifs : **articuler les ontologies** et permettre l'**annotation des modèles**.

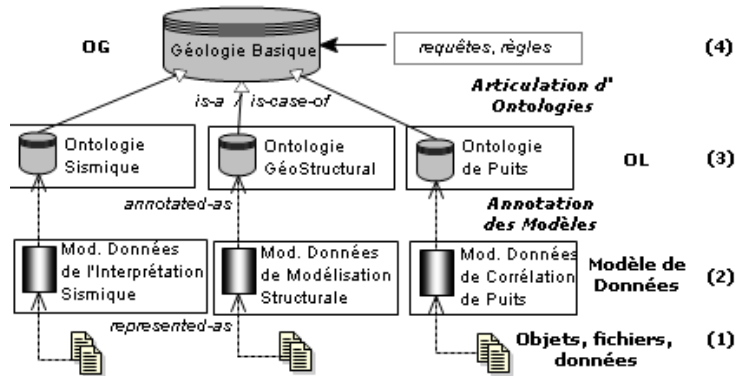


Figure 4. L'architecture d'annotation et intégration de modèles.

Pour l'**articulation des ontologies**, l'objectif n'est pas d'intégrer les concepts des **OL** à l'intérieur de l'**OG**, mais d'établir des *liens de subsumption* entre les concepts des ontologies locales et globale. Ces liens s'appuient sur les relations sémantiques de base *is-a* (est-un) et *is-case-of* (est-cas-de : héritage partiel). Par exemple, les concepts *Reflector* (**OL** Sismique) et *Marker* (**OL** de Puits) sont subsumés par le concept *Horizon* (**OG**) : *Reflector is-a Horizon* et *Marker is-a Horizon* (voir figure 2). Notons que *is-a* n'est pas l'unique relation de subsumption utilisée. D'autres relations de subsumption peuvent être définies.

Une fois le lien entre les ontologies établi, des mappings plus complexes et des règles d'inférence spécifiques de domaines seront appliqués à l'ensemble des ontologies (**OL** + **OG**) en vue de pouvoir faire des requêtes. Un problème demeure : celui d'appliquer les requêtes aux données elles-mêmes. Pour le résoudre, nous faisons appel à l'**annotation de modèles**. Nous distinguons trois types d'annotations selon la tâche et l'application utilisée :

- **Annotation boîte blanche** : le système d'annotation est intégré au logiciel de géo-modélisation. Quand l'utilisateur traite les modèles métier le système crée automatiquement des instances d'annotation. Le système connaît l'ontologie locale associé et produit des instances d'annotation faisant référence aux URI de cette ontologie.

- **L'annotation boîte noire** : lorsque l'utilisateur emploie un outil propriétaire, il n'est pas possible d'y intégrer le système d'annotation. L'annotation doit être réalisée de manière interactive en examinant les fichiers de données produites ou bien lors d'une exécution incrémentale de cet outil. Cela ressemble à un relevé d'expérience.

- **L'annotation intrusive** : le système d'annotation examine les fichiers de données produites et découvre, en utilisant des règles heuristiques, quels objets doivent être annotés et les associe ensuite aux instances de l'ontologie qui correspondent à ces objets. Par exemple, un programme d'annotation peut traiter des fichiers de données (instances de méta-données) et produire des instances d'annotations.

5. Implémentation de l'approche

L'implémentation envisagée pour notre approche doit être en mesure de représenter les éléments suivants : (i) les modèles de données et leurs objets ; (ii) les ontologies et leurs instances et (iii) les annotations d'objets par des instances de l'ontologie. Récemment, les propositions de (Broekstra *et al.*, 2002) et de (Dehainsala *et al.*, 2007) pour stocker et exporter à la fois des ontologies et leurs instances au sein d'une même base de données ont été réalisées. Ces architectures, appelées Base de Données à Base Ontologique (BDBO), traitent le problème de représentation des ontologies en tirant avantage des caractéristiques des bases de données (scalabilité, sécurité, etc.). Nous proposons d'utiliser la BDBO OntoDB pour notre implémentation et de l'étendre en vue de permettre de représenter les objets et leurs annotations.

La figure 5 (a) présente l'architecture de OntoDB. Elle comporte 4 éléments : *méta-base* (1) (constituée de les descriptions des tables système), *instances* (2) (décrit les instances de l'ontologie en associant une table à chaque classe), *ontologie* (3) (contient les ontologies définissant la sémantique des domaines couverts par la base de données), et *méta-schéma* (4) (représente le modèle d'ontologies utilisé). Notre proposition d'extension de OntoDB est illustrée sur la figure 5 (b). Elle consiste à augmenter cette architecture par les parties suivantes.

- *méta-modèle* (1). Le méta-modèle des données, comme un modèle de classes. Il permettra de produire de représentations pour les données issues des modèles métiers.
- *modèle de données* (2). Décrit les méta-données et leurs instances dans des tables relationnelles. Ce modèle est instance de (1).
- *Le modèle d'annotation* (3) et son méta-modèle, la partie *méta-annotation* (4).

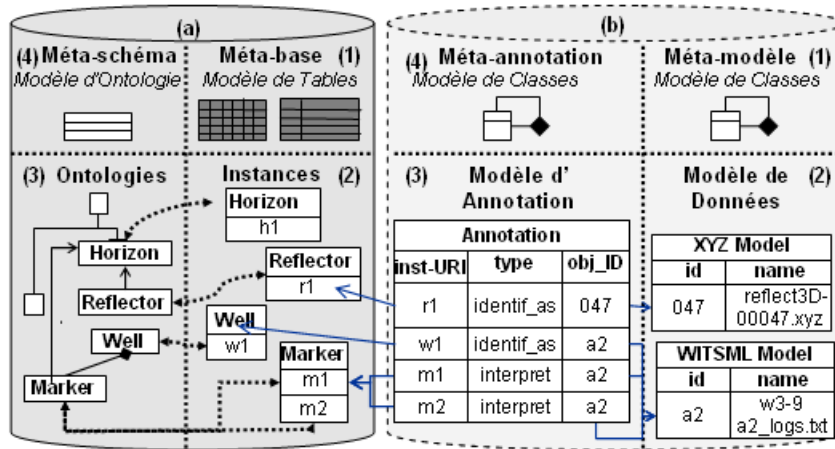


Figure 5. *OntoDB Architecture*

Grâce aux nouvelles parties que nous proposons, OntoDB permettra de stocker au sein d'une unique base de données des modèles et leurs annotations par une ontologie.

Dans la section suivante nous illustrons l'architecture d'annotation en présentant un cas d'étude réel concernant les processus de géo-modélisation.

6. Étude de Cas

Pour illustrer un cas réel du processus de géo-modélisation, nous présentons un exemple concernant l'extension de la base OntoDB (figure 5 (b)).

La donnée dans les géo-modèles est représentée comme des instances des modèles de données. Le fichier du type XYZ, qui stocke les objets sismiques identifiés par le géologue (fichier `reflect3D00047.xyz`), est représenté dans la BD comme une instance du XYZ Model dans la partie **Modèle de Données** (figure 5 (b) (2)). De la même manière, les données du puits nommé A2 (fichier `w3-9a2-logs.txt`) est représenté comme une instance du modèle WITSML.

Les interprétations géologiques sur la donnée sont représentés comme des instances des ontologies locales. Les objets sismiques peuvent être interprétés par le géologue comme des *Reflectors*. Ce concept local est défini dans la partie **Ontologie** (figure 5 (a) (3)), tandis que son instance *r1* est représentée dans la partie **Instances** (figure 5 (a) (2)). Le concept *Reflector* est reconnu par le géologue comme étant un type d'*Horizon* (concept de l'ontologie globale). Ceci est représenté par la relation de subsumption *is-a* entre *r1* et une instance d'*Horizon* (*h1*). Par ailleurs, en plus de reconnaître et d'annoter un fichier de puits (avec l'instance *w1* du concept *Well*), le géologue identifie les marqueurs qui font partie d'un puits (le concept *Marker* et ses instances *m1* et *m2*) et les interprète comme des *Horizons*, si c'est le cas.

L'annotation crée un lien entre les géo-modèles et leurs interprétations. La table *Annotation* (figure 5 (b) (3)) associe l'URI de l'instance *r1* à l'ID de l'instance du XYZ Model (id = 047). Dans cette table on ajoute aussi l'annotation d'un fichier de puits par une instance de *Well* et l'annotation de ce même fichier par des instances de *Marqueur* (puisque les marqueurs sont à l'intérieur du fichier de puits).

La mise en oeuvre de cette approche nous permet de répondre à la question formulée dans la section 3 : *quels sont les réflecteurs recoupés par un puits X ?*. Cette question est convertie en requête de la forme :

```
SELECT R FROM Reflector WHERE exists
(SELECT W FROM Well
WHERE id = 'A2' AND R intersects W)
```

L'algorithme opère en 2 étapes : (i) *parcourir les annotations pour trouver les liens créés entre la représentation des objets réels et les ontologies locales.* Dans cette étape on va chercher les annotations faites sur le puits A2 et trouver les instances *w1*, et *m1* et *m2*. (ii) *Remonter au niveau de l'ontologie globale à travers des relations de subsumption, ce qui permet de passer d'un domaine spécialisé à l'autre.* Dans le niveau ontologique, le lien de subsumption permet de trouver que les marqueurs *m1*

et $m2$ correspondent à l'horizon $h1$, de la même façon que le réflecteur $r1$. Comme conséquence on peut récupérer la relation d'intersection qu'existe entre le puits qui contient le marqueurs (puits $A2$), et ce réflecteur. Le résultat de la requête sera, donc, le réflecteur $r1$, avec lequel on peut retrouver, au moyen de l'annotation, les données métier correspondantes.

7. Conclusions

Nous avons présenté ici une approche à base ontologique pour annoter des modèles métiers. Le domaine d'application est la géo-modélisation dans le cadre des projets sur la capture et le stockage de CO₂. Nous proposons de créer des annotations basées sur des concepts d'ontologies locales subsumés par des concepts d'ontologies globales. Cette approche est intégrée dans une architecture de base de données ontologique (OntoDB) étendue avec un modèle d'annotation. La prochaine étape de ce travail sera la définition de requêtes multidisciplinaires dans le langage OntoQL utilisé dans OntoDB. Nous allons aussi définir des règles qui permettent de traiter des cas complexes d'interprétation, tels que les cas d'objets annotés par des multiples instances.

8. Bibliographie

- Broekstra J., Kampman A., van Harmelen F., « Sesame : A Generic Architecture for Storing and Querying RDF and RDF Schema », *The Semantic Web-ISWC*, vol. 2342, p. 54-68, 2002.
- Dehainsala H., Pierra G., Bellatreche L., « OntoDB : An Ontology-Based Database for Data Intensive Applications », *Proceedings of Database Systems for Advanced Applications, 12th International Conference (DASFAA'07)*, Springer, 2007.
- Farrell J., Lausen H., « Semantic Annotations for WSDL - W3C Working Draft », 2006.
- Gruber T., « Toward Principles for the Design of Ontologies Used for Knowledge Sharing », *International Journal of Human and Computer Studies*, vol. 43, n° 5/6, p. 907-928, 1995.
- Lin Y., Strasunskas D., Hakkarainen S., Krogstie J., Solvberg A., « Semantic Annotation Framework to Manage Semantic Heterogeneity of Process Models », *Proc. of the 18th Conference on Advanced Information Systems Engineering*, Springer, p. 433-446, 2006.
- Maillot N., Thonnat M., Boucher A., « Towards Ontology Based Cognitive Vision », *Computer Vision Systems : Third International Conference, ICVS 2003*, Springer, Graz, Austria, 2003.
- Mastella L., Perrin M., Abel M., Rainaud J.-F., Touari W., « Knowledge Management for Shared Earth Modelling », *EAGE Conference & Exhibition*, London, 2007.
- Schroeter R., Hunter J., Kosovic D., « Vannotea-A Collaborative Video Indexing, Annotation and Discussion System For Broadband Networks », *Knowledge Markup and Semantic Annotation Workshop, K-CAP 2003*, 2003.
- Uren V., Cimiano P., Iria J., Handschuh S., Vargas-Vera M., Motta E., Ciravegna F., « Semantic annotation for knowledge management : Requirements and a survey of the state of the art », *Web Semantics : Science, Services and Agents on the World Wide Web*, vol. 4, 2006.

Les Ontologies et leurs Applications en Agriculture

Catherine Roussey*, **Myoung-Ah Kang****

* *LIRIS CNRS, Villeurbanne*

catherine.roussey@liris.cnrs.fr

** *LIMOS, ISIMA, Aubière*

kang@isima.fr

RÉSUMÉ. Pendant les 20 dernières années, les ontologies sont devenues de plus en plus populaires dans divers champs de recherche tels que les technologies Web, les bases de données et les systèmes documentaires, etc. Cet article présente quelques applications générales des ontologies avant de donner des exemples dans le milieu agricole. Deux applications des ontologies en agriculture sont résumées : une au niveau international et une au niveau national.

1. Qu'est-ce qu'une ontologie ?

Cette section décrit brièvement l'histoire du mot « ontologie » depuis la définition d'Aristote jusqu'à celles des chercheurs en Intelligence Artificielle. Ainsi, les lecteurs auront une vue synthétique de l'utilisation des ontologies.

La définition d'Aristote du mot Ontologie (avec un grand O) est « the science of being *qua* being » (la métaphysique d'Aristote). Cette science fait partie de la philosophie, et traite de la description des entités du monde réel. Son but est de définir les catégories générales employées pour classer toutes entités du monde (les êtres humains, les animaux, les objets, etc.) Au début des années 80, les chercheurs en intelligence artificielle ont empruntés le terme « Ontologie » au champ de la philosophie. Les ontologies sont devenues la définition des connaissances d'un domaine. Elles fournissent la possibilité de séparer les connaissances du domaine des connaissances opérationnelles (Studer *et al.*, 1998).

Gruber en 1993 a proposé une définition intéressante des ontologies dans le domaine de l'intelligence artificielle : une ontologie est « *une spécification d'une*

conceptualisation pour aider des programmes et des humains à partager des connaissances” (Gruber, 1993a).

Plus précisément : “...*les ontologies sont des spécifications explicites formelles d'une conceptualisation partagée*”. D'après Gruber, « *une conceptualisation* » est le résultat de la modélisation de phénomènes du domaine d'intérêt. Cette modélisation identifie les concepts et leurs relations décrivant ces phénomènes. « *Explicite* » signifie que les concepts et leurs relations sont typés, et les contraintes sur l'utilisation de ces types sont clairement expliquées. « *Formel* » se rapporte au fait que l'ontologie doit être compréhensible par la machine. « *Partagé* » reflète l'idée que l'ontologie doit capturer les connaissances consensuelles communément admises par l'ensemble de la communauté des acteurs du domaine (Gruber, 1993b).

Dans le domaine des systèmes de bases de données, les ontologies sont employées pour faciliter l'interopérabilité des sources d'information hétérogènes. L'ontologie est le schéma général organisant toutes les propriétés des entités décrites dans les schémas de base de données ou sources d'information. Chaque source d'information possède un médiateur qui traduit son schéma de base de données en éléments de l'ontologie. Ainsi, les utilisateurs peuvent questionner l'ontologie et obtenir un résultat intégrant toutes les données des sources d'information requises pour répondre à la requête (Wiederhold, 1992).

Les techniques de recherche documentaire emploient des ontologies linguistiques telles que les thesauri ou les langages d'indexation afin d'éviter les ambiguïtés sémantiques liées à l'usage des termes polysémiques ou synonymiques. Ainsi, le document et le contenu des requêtes sont représentés par des concepts (c.-à-d. le sens des termes) et non par les termes eux-mêmes (c.-à-d. l'ensemble des caractères). Avec cette technique il est possible d'améliorer la description du contenu des documents (et des requêtes), et également d'améliorer les performances du système de recherche d'information. Un des premiers systèmes opérationnels est le système Ontoseek, qui a utilisé la base terminologique de Wordnet pour décrire les Pages Jaunes (Miller *et al.*, 1993) (Guarino *et al.*, 1999).

Le Web sémantique est une adaptation des technologies précédentes. Tim Berners-Lee, l'inventeur du Web, définit le Web sémantique comme « *un Web des données qui peuvent être traitées directement et indirectement par des machines.* » En effet, le Web sémantique concerne : l'échange et le traitement des données, des documents et des services. Le Web sémantique intègre une communauté d'agents capables d'échanger des données et des services des sources différentes afin d'atteindre un but spécifique. Plus précisément, les services Web sont des composants logiciels réutilisables qui implémentent une fonctionnalité accessible sur le Web (par exemple, le service de réservation d'un hôtel). De plus, la description et le contenu des documents sont définis précisément grâce à une série de méta-données. Le W3C a défini un ensemble de normes, nécessaires pour le développement des technologies associées au Web sémantique (W3C, 2001). Toutes ces normes sont basées sur le langage XML.

2. Deux exemples d'ontologies en agriculture : une de niveau international et une de niveau national.

Dans le domaine agricole, le thesaurus bien connu AGROVOC est employé pour développer des projets de Services d'Ontologies Agricoles (AOS) (AGROVOC, 2008). AGROVOC est un thesaurus multilingue concernant la sylviculture, la pêche, l'agroalimentaire, l'environnement et les domaines connexes. Comme présenté dans (AGROVOC, 2008), « *il se compose de mots ou d'expressions (termes) dans différentes langues, reliés entre eux par des relations assez générales tels que « plus précis», « plus général», « est relatif à»,.... L'objectif est d'identifier ou rechercher des ressources* ». AGROVOC a été développé par la FAO et la Communauté Européenne au début des années 80.

En utilisant les connaissances contenues dans AGROVOC, « *les AOS pourront développer des terminologies spécialisées et des concepts spécifiques qui amélioreront la gestion de l'information sur le Web. Un objectif majeur est d'ajouter plus de sémantique au thesaurus, en indiquant plus précisément et plus systématiquement les relations entre concepts.* » (AGROVOC, 2008). Par exemple, dans le thesaurus AGROVOC le terme « pollution » est associé au terme « polluants » par la relation « termes relatifs ». Ceci signifie que le terme « pollution » est lié au terme « polluants ». Pour améliorer la sémantique de cette terminologie, il faudrait indiquer que la « pollution » est provoquée par des « polluants » en employant une relation plus précise telle que « est causé par » (Wikipedia, 2008).

En France, un projet appelé GIEA a concerné l'étude d'une ontologie du domaine agricole (Dufy *et al.*, 2006). Cette ontologie n'est pas seulement limitée à une chaîne de production particulière ou à un type d'activités ; elle peut être employée dans une majorité de contextes agricoles. En effet, cette ontologie couvre les concepts communs utilisés dans les principales chaînes de production et elle est conçue pour être employée dans différents domaines d'activités agricoles en France. Un objectif majeur de l'ontologie est de fournir des formats d'échange de données complets pour faciliter l'interopérabilité des systèmes d'information agricoles.

La première tâche des équipes thématiques du projet GIEA a été de faire l'inventaire des initiatives traitant de leur thème. Ainsi, différents termes, concepts et relations ont été identifiés pour chaque thème. Ensuite, la majeure partie du travail a consisté à intégrer et compléter les définitions des concepts déjà identifiés dans des projets ultérieurs orientés mono-activité.

Tous les membres du projet pouvaient proposer de nouveaux concepts à intégrer à l'ontologie en cours. Des formats d'échange de données issus des vocabulaires et des concepts de l'ontologie ont été proposés. L'ontologie a été représentée par des diagrammes de classe d'UML. UML a été choisi pour modéliser l'ontologie car les participants du projet GIEA utilisaient ce formalisme couramment.

L'ontologie développée se compose de plusieurs parties. Chacune concerne un thème spécifique, traité d'une façon adaptée aux spécificités françaises (GEIA, 2008) :

- Acteurs (modélisant les caractéristiques principales des acteurs impliqués dans la production agricole - exploitations, agriculteurs, etc.),
- Alimentation,
- Analyse et surveillance de la production,
- Analyse du sol,
- Bâtiments agricoles,
- Animaux,
- Activités reliées à la production agricole (épandage, pulvérisation, irrigation, moisson, etc.),
- Identification et traçabilité des animaux,
- Intrants agricoles (engrais, eau, etc.)
- Règlementations,
- Occupation du sol,
- Récoltes,
- Reproduction du bétail,
- Surveillance sanitaire.

3. En résumé

Pour conclure, les ontologies peuvent prendre diverses formes, et leur utilisation peuvent également variées, mais nous pouvons définir qu'une ontologie devrait principalement contenir :

- un vocabulaire de termes,
- un ensemble de définitions des termes qui identifient les concepts et fixent l'interprétation des termes,
- la représentation des relations entre les concepts,
- un consensus d'une communauté d'utilisateurs de l'ontologie.

Au niveau international, le thesaurus AGROVOC est employé pour développer des projets de Services d'Ontologies Agricoles (AOS). Il intègre des termes généraux en agriculture et leur traduction dans plusieurs langues.

Au niveau national, le projet de GEIA a modélisé différents concepts communs en agriculture ; ces concepts ont été considérés avec les spécificités propres à la France. L'ontologie proposée capture les connaissances consensuelles admises par des experts en France. L'utilisation d'UML a permis aux participants du projet GEIA de discuter des concepts facilement. Ainsi, ces concepts seront employés pour proposer un format standard d'échange de données pour faciliter l'interopérabilité entre les logiciels et les systèmes d'information agricoles.

Références

- AGROVOC – A multilingual Agricultural Thesaurus -
ftp://ftp.fao.org/gi/gil/gilws/aims/references/flyers/agrovoc_en.pdf
- Dufy L., Abt V., Poyet P. GIEA : Gestion des informations de l'exploitation agricole. Un projet au service de l'interopérabilité sémantique de la profession agricole. Ingénieries – Eau, Agriculture, Territoires, n° 48, p. 27-36, 2006
- GIEA: GIEA Web site: <http://www.projetgiea.fr>
- Gruber, R.: Towards Principles for the Design of Ontologies Used for Knowledge Sharing. In International Workshop on Formal Ontology, Padova, Italy, 1993. Available as technical report KSL-93-04
- Gruber, R.: A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition vol. 5(2), pp. 199-220, 1993
- Guarino, N., Masolo, C., Vetere, G.: OntoSeek: Content-Based Access to the Web, IEEE Intelligent Systems vol. 14(3), 70-80, 1999
- Miller, G., Beckwith, R., Fellbaum, C., Gross D., Miller, K.: Introduction to WordNet: an on-line lexical database'. International Journal of Lexicography vol. 3 (4), 1990, Revised August 1993 -
<ftp://ftp.cogsci.princeton.edu/pub/wordnet/5papers.ps>
- Studer, R., Benjamins, V., and Fensel, D.: Knowledge engineering: Principles and methods. IEEE Transactions on Data and Knowledge Engineering vol. 25: p. 161-197, 1998
- Wiederhold G.: Mediators in the Architecture of Future Information Systems. IEEE Computer 25(3): 38-49 (1992) 8
- Wikipedia: AGROVOC - Web Site: <http://en.wikipedia.org/wiki/AGROVOC>
- W3C: Semantic Web, 2001 - <http://www.w3.org/2001/sw/>

Système intégré d'aide à la décision en gestion par bassin, de la conception à l'opérationnalisation

Étude de cas

Carlo Prével

Groupe d'études interdisciplinaires en géographie et environnement régional (GEIGER), Département de Géographie, Université du Québec à Montréal (UQAM), C.P. 8888, Succ. Centre-Ville. Montréal (Qc) H3C 3P8 Canada. (previl.carlo@uqam.ca)

RÉSUMÉ. Cet article présente le contexte de la mise en œuvre d'un système intégré d'aide à la décision (SIAD_GBV) réalisé spécifiquement pour l'Agence (de gestion) du Bassin hydraulique du Souss-Massa (ABHSM) au Maroc. Le développement du SIAD doit aider l'Agence à se doter d'une base de connaissances territoriales, à jour, et d'un savoir-faire approprié, pour une meilleure prise en charge de l'arbitrage des changements dans l'occupation du sol et la gestion des ressources précaires disponibles en eau (région aride du Maghreb). La conception du SIAD a considéré l'ensemble de la mission de l'Agence et les applications ont visé différentes problématiques de décision territoriale identifiées avec les responsables de l'Agence.

Au-delà des composantes techniques et méthodologiques, la réalisation du SIAD_GBV a mis en évidence l'importance, dans un projet de la sorte, d'un partenariat solide avec les intervenants locaux et de la formation du personnel.

ABSTRACT. This paper relates the implementation of an integrated decision support system (IDSS) designed specifically for the Souss-Massa Hydraulic Basin Agency (AHBSM). This project has been developed to provide AHBSM with an updated spatial database, the appropriate tools and the building capacity for a better support in the management and arbitration of precarious water resources. The conceptual model of the database has been designed to take into account the entire role and mission of the water agency. The applications developed present integrated solution for different territorial decision problems identified with the staff of the Agency.

Beyond the technical and methodological components, the realization of IDSS highlighted the importance, in a project of this kind, of a strong partnership with local stakeholders and staff training.

MOTS-CLE : Système d'information géographique (SIG), Système intégré d'aide à la décision, gestion par bassin versant, modélisation territoriale, analyse spatiale.

KEYWORDS: Geographic information system (GIS); Integrated Decision Support System (IDSS); Watershed management; land use; spatial analysis.

1. Mise en contexte

Les pays de la région méditerranéenne méridionale, comme le Maroc sont exposés à de nombreux défis par rapport à la gestion des ressources hydriques. Les gestionnaires de l'eau doivent trouver des solutions originales pour répondre à la hausse de la demande en eau autant pour la consommation de la population que pour l'irrigation. Depuis 1995, la gestion de l'eau au Maroc est régie par la Loi 10-95 (1995). La mise en œuvre de la loi est confiée à des agences de bassin qui doivent s'appuyer sur un ensemble de principes dont a) le caractère public des ressources hydriques; b) la planification et la répartition des ressources fondée sur la concertation des usagers; c) la réglementation des activités pour pallier à la précarité, la vulnérabilité et les risques inhérents à l'exploitation des ressources hydriques.

Les interrelations nombreuses entre les activités ayant lieu sur le territoire du bassin font que tout changement notable d'affectation du sol peut avoir des impacts en cascade sur la gestion et la disponibilité de l'eau dans le bassin; la production de biens, de services et le bien-être de la population en général (Lasalmil et Boulal, 2003). L'Agence de bassin, dans ce contexte a hérité d'un problème complexe (organisation élevée; incertitude de l'environnement; impossibilité de connaître et comprendre toutes les relations en jeu). Dans ce cas, la solution nécessite une approche intégrée, interdisciplinaire, fondée sur la disponibilité d'une meilleure information territoriale pour l'élaboration de scénarios, pour l'animation des débats avec le public et pour l'éducation des parties prenantes, bref, pour l'aide au processus décisionnel territorial. Cette solution implique une démarche d'apprentissage, fondée sur une axiologie socio-constructiviste pour une gestion adaptative de la ressource (Prévil et *al.* 2004).

1.2. Accompagnement du processus décisionnel

La mission de l'agence par rapport à la ressource en eau couvre tout ce qui touche en quantité, en qualité (pollution), en ouvrages, en répartition, en risques (crues), en redevances (utilisateur-payeur), et en éducation (communication, vulgarisation, sensibilisation et participation). Pour embrasser une telle mission, l'agence doit tenir compte des dimensions sociales, spatiales, de la dynamique espace-temps ainsi que du cadre légal-institutionnel (Keith et Said, 2004). On comprend ainsi que l'Agence se trouve placée au centre d'un ensemble de problèmes affectant à différents niveaux les mécanismes décisionnels. L'obligation de la concertation entre les groupes d'usagers exige la disponibilité de connaissances territoriales de synthèse pour aider à discuter, négocier ou convaincre les partenaires, de la gestion de l'eau, du bien-fondé d'une *façon de faire* par rapport à une autre (Hanna, 2000). Dans cet esprit, on peut citer : l'information actualisée sur l'état de la ressource, des outils non seulement pour suivre et modéliser l'évolution des ressources, mais également pour simuler l'effet probable des scénarios de gestion.

C'est à cet ensemble de défis que nous avons proposé de répondre, avec la participation de l'agence, par un système de solutions (Lévêque, 2001) basé sur les méthodes d'aide à la décision, les représentations graphiques et la modélisation systémique intégrée grâce à la géomatique (Longley et al., 2005). Ce système de solutions s'est concrétisé à travers la réalisation d'un SIAD_GBV (appelé pour le projet SIAD_TIGERS). Dans le SIAD, nous avons cherché à dégager une compréhension globale (conceptuelle) du système de problèmes affectant la gestion de la ressource en eau dans le bassin de Souss-Massa. L'organisation de ces solutions au sein du SIAD_GBV garantit le fait que ces solutions ne soient pas isolées, mais systémiques avec les fonctions d'interaction, de globalité, d'organisation et une structure capable de composer avec la complexité des questions énoncées précédemment. (Prévil et al. 2004 et Simonovic, 2007)

Ce système devra aider à répondre aux aspects liés à : a) la planification territoriale ; b) l'information territoriale (spatio-temporelle) ; c) la gestion des réservoirs ; d) la protection contre les dangers d'inondation ; e) la meilleure exploitation des ressources en eau souterraines et f) l'organisation des connaissances et du savoir du territoire.

1.2.1. Définition, et portée du SIAD

La réalisation d'un SIAD se situe dans l'évolution d'ensemble des méthodes et outils de la science de l'information géographique. Il n'existe pas de méthode formelle de production de SIAD pour les bassins versants. Prévil et al. (2004) rappellent qu'un véritable SIAD adapté à la gestion par bassin ne devrait pas se ramener à un intégrateur de modèles fourre-tout dans lequel la compréhension de la problématique décisionnelle est négligée. Le SIAD ne doit pas être appréhendé comme un outil de superposition de couches de données territoriales et encore moins comme une interface presse-bouton pour exciter un ensemble d'algorithmes ou de logiciels de simulation plus ou moins sophistiqués. Le SIAD doit répondre fondamentalement à des problématiques décisionnelles. De nombreux systèmes dits intégrés négligent tout simplement la problématique décisionnelle sous-jacente et ne constituent plus souvent que des plates-forme où le seul effort d'intégration s'applique au format des données de manière à ce que l'utilisateur n'ait pas à trop se casser la tête pour les faire passer d'un interface à un autre (Cesur et al., 2004; Andreu et al., 1998; Dieulin et Boyer, 2003).

La mise en œuvre du SIAD s'accompagne d'implications technologiques, organisationnelles et sociales importantes pour l'AHBSM chargée de sa mise en œuvre. Le SIAD_GBV s'inscrit dans le courant de recherche pour le couplage des dimensions humaines et environnementales des systèmes (*Coupling Human and Environmental Systems*, CHES, Biosphere Data project, 2004) avec des spécificités opérationnelles découlant a) des aspects socioconstructivistes et contextualistes de l'apprentissage, b) des approches multicritères de la Théorie de la décision, et c) des méthodes de la science de l'information géographique (géomatique). Dans cette mise en œuvre, le SIAD est appréhendé comme une démarche d'apprentissage et de

réalisation transversale, à long terme, visant à doter les décideurs et les parties prenantes de la gestion d'un bassin, comme le Souss, d'un système complexe fondé sur la gestion de l'information territoriale (Longley et *al.*, 2005). Ce système intègre des modules qui ont été rendus opérationnels rapidement pour s'ajuster au modèle d'affaires des gestionnaires du bassin versant.

Une démarche d'implémentation d'un SIAD n'est pas destinée à remplacer le débat politique ou la décision humaine. La finalité vise d'abord à tenir compte de la rationalité limitée de la décision (Rosenhead, 2001), à informer les étapes d'un processus territorial (Carver et *al.*, 1998) et à organiser les actions du processus décisionnel (Heywood et *al.*, 2006). De cette façon, on peut justifier la légalité et améliorer la légitimité des débats, des négociations voire la décision, quand c'est possible (Roy et Boyssou, 1985 ; Longépé, 2006).

La mise en œuvre d'un SIAD nécessite un travail patient et rigoureux qui doit s'initier avec la construction de la problématique décisionnelle suivie d'une structuration des objets des processus décisionnels. Dans le cas d'une problématique territoriale (spatiale) les outils de la géomatique et de la modélisation peuvent aider à la mise en relation des principaux actes, actions ou phases d'activités des processus territoriaux (Heywood et Carver, 1994; Hill et *al.*, 2005). L'objectif d'appropriation par les parties prenantes de la décision territoriale doit être toujours présent à toutes les étapes de la démarche (Prévil et *al.*, 2004).

1.2.2. Démarche de réalisation et de structuration du SIAD

Le savoir-faire dans l'élaboration d'un SIAD inclut trois dimensions intrinsèques (générique, fonctionnel et structurel). Différentes problématiques décisionnelles de proximité (de simple à complexe) auxquelles font face l'Agence, ont été analysées, de concert avec l'Agence. Par exemple, l'ensemble des opérations soumises à autorisation a été passé en revue. Ces opérations concernent des activités comme le creusement de puits ou forages, les prises d'eau établies sur les cours d'eau ou les prélèvements dans la nappe souterraine au-delà de certains seuils préétablis, etc.

À partir de là, il était permis de travailler de manière approfondie en vue de la concrétisation de l'outil géo-informatique. Cette implémentation a été réalisée selon trois axes :

- un axe de formalisation qui intègre la constitution des différentes phases de modélisation des données et l'alimentation du SIAD
- un axe d'utilisation qui consiste en la démonstration de l'exploitabilité du SIAD à travers la résolution de différents problèmes territoriaux
- un axe de facilitation qui correspond à la programmation d'une barre d'outils facilitant la mise à jour et le développement des applications de l'axe précédent par les responsables de l'Agence.

Les réalisations ont porté sur différents aspects complémentaires. La mise en œuvre du SIAD a nécessité la réalisation d'un ensemble d'activités. Dans ce papier, nous rappellerons les réalisations principales découlant de ces activités. Des publications spécifiques présentent les principaux jalons de la mise en œuvre et certains résultats obtenus de ces applications (Prévil et *al.*, 2008 ; Dolcine et *al.*, 2008).

2. Mise en œuvre

2.1. Constitution du modèle territorial et alimentation du SIAD (Axe de formalisation)

La réalisation de l'axe de formalisation du SIAD vise à répondre à deux questions : Comment améliorer le processus décisionnel dans le bassin et assurer la bonne gestion des ressources en eau du bassin? Et ensuite, comment améliorer les connaissances sur les tendances (court terme et long terme) de l'effet des transformations de l'utilisation des ressources du territoire?

La démarche de compréhension a abouti à la schématisation d'un arbre décisionnel pour représenter les principales relations entre les services délivrés par la gestion par bassin (la mission de l'agence et les données nécessaires (pour définir les options disponibles pour la dispense de ces services) (Figure1). Cette étape a permis la première conception d'un modèle conceptuel de données (MCD). Il s'agit d'une représentation schématique des données territoriales indépendamment de toute considération technologique. Il a été organisé grâce à un formalisme standard UML orienté objet, bien adapté au SGBD objet relationnel de ArcGIS (Maidment, 2002 ; Bédard et *al.*, 2004) par transformation progressive d'une carte de connaissances, pour que la procédure demeure accessible aux non spécialistes (Piaget, 2002). C'est ainsi qu'à l'occasion d'une activité de terrain, il a été discuté en atelier par les représentants des organisations formant le Conseil d'administration de l'Agence et les tirages révisés ont été régulièrement soumis pour commentaires aux cadres de l'Agence. Ce MCD du bassin de Souss-Massa représentait les entités territoriales pertinentes à la gestion des ressources en eau du bassin ainsi que les relations pouvant être établies entre ces entités, en plus des attributs caractérisant ces entités. ArcHydro a constitué une excellente synthèse de données géospatiales et temporelles pour supporter la modélisation hydrologique (quantité et qualité) et la prise de décision, *in extenso* (Maidment, 2002).

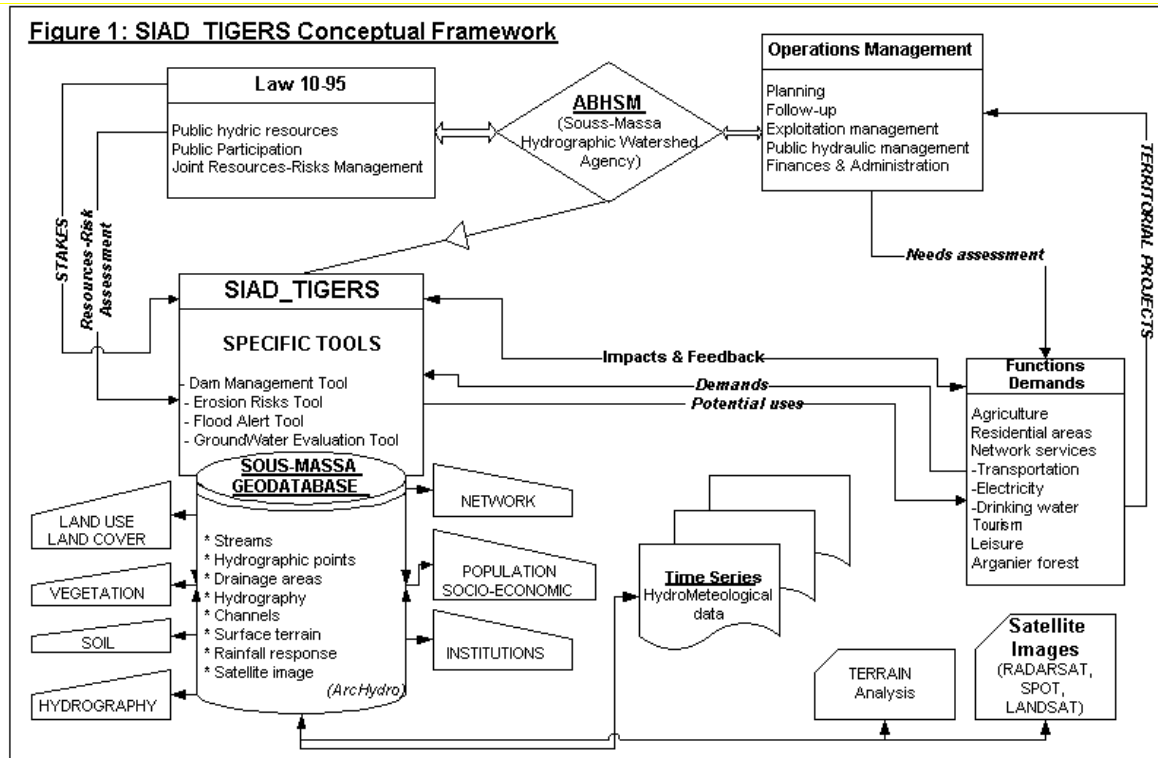
2.2. Alimentation et mise à jour de l'information territoriale dans la GeoDataBase

Les processus décisionnels informés portant sur le territoire et l'environnement sont déficients dans la région Souss-Massa. Pour aider à résoudre ce problème de disponibilité d'information territoriale, l'une des finalités de la réalisation du SIAD visait à améliorer les connaissances territoriales sur les réalités et les tendances (court terme et long terme) de l'effet des transformations de l'occupation et l'utilisation du sol.

La mise en place de la base de données intègre la collecte, la numérisation, la préparation (*geoprocessing*) et la structuration selon le modèle adopté. Dans la GeoDataBase, la structuration des données a été réalisée de manière à assurer que leur accès puisse s'effectuer selon trois approches : a) thématique; b) spatiale et c) temporelle (Zeiler, 1999 ; Maidment, 2002).

Un ensemble d'informations stratégiques portant sur les aspects bio-physiques, administratifs, économiques ou socioculturels de la gestion du bassin ont été recueillis. Des images satellitaires RADARSAT, Landsat et SPOT ont été traitées pour établir l'occupation et l'utilisation du sol (parcellaire et culturale) ainsi que l'analyse de terrain.

L'approche du SIAD_GBV a permis à l'Agence de bassin d'assurer l'enrichissement et la mise à jour des données, particulièrement grâce à l'exploitation des images de télédétection (Ziad et *al.*, 2007). Les images radar (RADARSAT) ont été utilisées pour l'extraction du modèle numérique d'altitude (MNA). De ces images et MNA, les attributs de terrain ont été extraits (pente, linéaments, exposition de la pente, vue, découpage des versants, réseau d'écoulement, etc.). Par la suite les images ont été utilisées également pour l'établissement des cartes de linéaments ainsi que les cartes d'utilisation du sol. Une fois cette première étape de constitution de la GeoDataBase et de mise à jour a été franchie, on a pu passer à une seconde phase d'élaboration d'information nouvelle portant sur la dynamique de l'évolution de l'utilisation du sol dans le bassin entre 1985 et 2006, notamment, grâce à l'analyse des produits d'observation de la terre (Ziad et *al.*, 2007).



2.3. Exploitation du SIAD (Axe d'utilisation)

L'exploitation du SIAD constitue une façon de montrer et de démontrer avec l'Agence la pertinence du SIAD par rapport à des problèmes réels. Il inclut l'utilisation du système d'information, le développement d'outils spécifiques pour la gestion par bassin versant et des outils complémentaires dans le contexte de Souss-Massa. Ces applications, chaque fois que c'était possible, ont été développées avec une perspective globale et une perspective locale (Vieux, 2001). Les applications ont permis la construction, étapes par étapes d'un ensemble de produits fournissant une meilleure connaissance territoriale par rapport à la disponibilité en eau souterraine (échelle globale et échelle locale - Igrhem); la vulnérabilité du milieu aux crues et aux inondations (échelle globale et échelle locale – Souss à Taroudant); et la gestion des barrages (Youssef Ben TachFine et AbdelMoumen) et érosion (Bassin global de Souss-Massa, bassin d'apport des réservoirs). Il importe de présenter succinctement ces différentes applications et de montrer comment elles contribuent à illustrer les potentialités du SIAD pour l'Agence.

La démarche de résolution pour toutes les applications comporte généralement cinq étapes. Les trois premières étapes s'insèrent dans les activités décrites pour

l'alimentation de la GeoDataBase. L'étape 3 correspond partiellement aux activités de mise à jour de la GeoDataBase avec la production de la LULC, du MNT et des produits dérivés. Pour compléter cette troisième étape, des activités spécifiques sont entreprises pour transformer ou générer de nouvelles données comme pour le calcul des linéaments, la densité de drainage ou une distance par rapport aux linéaments, etc. (Vieux, 2001).

2.3.1. Évaluation des potentialités en eaux souterraines

Les eaux de surface sont globalement limitées au Maroc (quantité et permanence). L'eau souterraine constitue une alternative d'appoint. Le bassin Souss-Massa contient l'un des aquifères les plus importants du pays. L'essor de l'agriculture exportatrice dans la région a beaucoup profité de cette potentialité (jusqu'à en abuser).

Les méthodes combinées les plus connues intègrent des modèles empiriques basés sur des combinaisons des paramètres découlant des caractéristiques du milieu bio-physique. En fonction des conditions de sites, on cherche les facteurs permettant de mieux représenter l'infiltration de l'eau, prédire la localisation de la ressource-eau et le comportement des aquifères continus ou confinés (Muthy, 2000; Rao et Jugran, 2003).

Sur le plan global, la méthode utilisée a confirmé les potentiels de la région de la plaine du Souss. Les observations sur d'autres espaces locaux (zones pilotes) s'avèrent intéressantes. L'étude localisée dans la région d'Ingrhem, par exemple, peut aider à l'agence de mieux cibler les plans de forages annuels réalisés dans cette région.

L'évaluation des potentialités en eaux souterraines permet à l'Agence d'atteindre un double objectif. Sur le plan global, elle dispose d'une meilleure information sur les potentialités d'ensemble et la vulnérabilité des eaux souterraines du bassin. Cette information, combinée aux données des piézomètres et aux activités de terrain, aidera l'Agence à mieux modéliser la dynamique de la nappe et à identifier des zones à richesses potentielles pour des forages exploratoires. Sur le plan local, particulièrement à Ingrhem, l'approvisionnement de la population dépend des eaux de pluie (quand il y en a) et des forages. L'utilisation du SIAD permettra de détecter les sites les plus propices à la détection de nappes confinées pour améliorer le taux de succès des forages

On pourra par la suite raffiner la démarche en y introduisant a) une méthode de somme pondérée (*Weighted Index Overlay*); b) l'intégration de *l'Analytical Hierarchy Process* (AHP) et finalement; c) l'exploration les autres apports possibles en introduisant une approche basée sur le *fuzzy logic*.

2.3.2. Gestion des barrages :

Sur le territoire du bassin Souss-Massa ont été érigés huit barrages à capacité variable. La gestion des barrages constituait, jusqu'à très récemment, l'activité prédominante de l'Agence. Les gestionnaires ont toujours été intéressés pour voir comment les produits d'observation de la Terre et le SIAD peuvent les aider dans la gestion des barrages pour : a) la gestion de la demande pour l'irrigation, la consommation urbaine et rurale, les loisirs; b) le bilan hydrique au niveau des réservoirs des barrages (entrées, sorties, évaporation, pertes latérales) en fonction des scénarios de gestion préétablis; c) la répartition des assiettes entre les catégories d'utilisateurs (eau potable, irrigation gravitaire, recharge de nappe); d) le suivi des autorisations de prélèvement.

Cette application a aidé à montrer comment la gestion des barrages, au lieu d'être une activité isolée d'un service peut intervenir en input et comme output par rapport à l'ensemble des autres missions de l'agence. Les différentes données modélisées dans le SIAD sont mises en interrelation avec les séries hydro-météorologiques particulièrement : hauteur d'eau dans le barrage, pluie, rayonnement solaire, débit, température. On estime l'évaporation ainsi que d'autres paramètres pour déterminer l'infiltration, les prélèvements. Des outils de visualisation, d'exportation des données, de calcul des paramètres dérivés liés à la problématique de l'eau ETP, P-ETP, Demande en eau et interpolation spatiale de ces données ont été également évalués pour une meilleure gestion des barrages au sein de SIAD.

Il s'agit de permettre l'interaction des modèles hydrauliques et hydrologiques avec les autres modules de données à travers le SIAD. Les scénarios de gestion et le suivi en temps réel des barrages s'en trouvent grandement améliorés. Le SIAD permet à l'agence de donner du sens à l'interaction des facteurs localisés (type de versant, type d'occupation du sol, prélèvement, caractéristiques des travaux de conservation du sol, simulation) de plus le SIAD permet d'évaluer en temps réel la pertinence des scénarios de gestion des lâchers en fonction de la pluviométrie, par exemple.

2.3.3. Érosion hydrique

La spécialisation de l'économie de la région de Souss-Massa dans les créneaux de l'agriculture intensive et du tourisme international de masse nécessite le recours à des outils puissants pour le suivi de certaines caractéristiques environnementales comme l'érosion du sol. L'érosion du sol entraîne une réduction de la qualité productive des sols en emportant des éléments favorisant les agrégats et contenant les matières nutritives pour les plantes. D'autre part, les particules arrachées vont accélérer les mécanismes de pollution diffuse, l'eutrophisation et la turbidité des eaux de surface, la déstabilisation des dunes de sable, l'altération des habitats et surtout l'envasement des réservoirs des barrages.

Différentes démarches sont disponibles pour étudier la sensibilité des sols à l'érosion. Les travaux de Wischmeier et Smith (1978) constituent sans conteste une référence incontournable dans le cas des milieux naturels et des sols agricoles depuis la formulation de l'équation universelle des pertes des sols (*Universal Soil Loss Equation*, USLE).

De nos jours, l'utilisation des produits d'observation de la terre et des SIG fournit de nouvelles avenues pour l'opérationnalisation de ces modèles, l'évaluation des principaux facteurs et leur résolution grâce à l'analyse spatiale (Duchemin et al., 2001; Mitasova et al., 1996). Dans cette application, il s'agissait d'abord d'aider l'Agence de bassin à utiliser les capacités du SIAD en vue de mieux évaluer et suivre (monitoring) la vulnérabilité et les risques d'érosion (particulièrement hydrique) des sols dans le bassin au niveau global et aux échelles locales pour mieux décider sur le territoire. L'application a été portée sur la USLE par la suite, les administrateurs du SIAD pourront prendre le temps d'évaluer la pertinence de recourir à d'autres modèles plus sophistiqués pour évaluer la perte des sols.

Les différentes données utilisées pour évaluer les facteurs de l'équation universelle de la perte des sols proviennent directement ou indirectement de la GeoDataBase. Certaines ont été transformées. Cette approche globale a été reprise localement dans les différents sous-bassins en amont des barrages.

L'érosion généralement est présente selon le gradient de pente, c'est -à-dire faible en plaine et forte en montagne. Ceci permet de vérifier la cohérence du modèle utilisé. En plus, On peut réaliser que 35% du territoire est exposé à une érosion forte. Les informations à l'échelle locale permettent d'apprécier le contexte d'insertion de certains habitats et équipements collectifs. À partir de là, on a pu affiner la méthode pour étudier l'érosion au niveau local, particulièrement dans les voisinages des têtes de barrages.

2.3.4. Zones menacées d'inondation

Même si la région de Souss et le Maroc généralement évoquent une idée de précarité de disponibilité en eau et de voisinage désertique, il est toutefois important de retenir que cette région est également exposée à des pluies torrentielles et à des crues (flash flood) qui peuvent occasionner des dégâts considérables. La protection contre les inondations se réalise grâce à des mesures avant, pendant et après les inondations. Dans cette application, il s'agira de partir d'expériences vécues sur le territoire pour étudier les probabilités de danger et finalement d'esquisser des mesures pouvant aider dans la gestion des crises d'inondation en temps réel.

Certains des paramètres affectant la variation du niveau d'eau peuvent être repérés et contextualisés pour aider à gérer les zones d'inondation. Un modèle empirique multicritère combinant les précipitations, l'utilisation du sol, la densité de drainage, la distance au réseau de drainage, le type de sol et la densité de population

a été appliquée pour développer des cartes de risques d'inondation pour la zone. Ces cartes ont été affinées par des zones inondables déterminées par des enquêtes de terrain. Un de ces points sur la rivière Souss à Taroudant a été retenu pour une étude détaillée avec simulation hydraulique. La gestion pluie-débit et les systèmes d'alertes aux crues dans le SIAD intègrent le modèle HEC-HMS pour la prévision hydrologique et le modèle HEC-RAS pour la simulation hydraulique (Hydrologic Engineering Center : *User's manual*, <http://www.hec.usace.army.mil/> , December. 2007).

2.4. Organisation de la barre d'outils (Axe de facilitation)

L'ensemble des manipulations a été implémenté au sein d'une plate-forme géo-informatique. En vue du suivi de l'utilisation intégrée du SIAD au sein de l'Agence et entre l'agence et ses partenaires, il s'agit de se demander comment assurer la cohérence de l'animation et l'utilisation de la connaissance territoriale pour une meilleure concertation des communautés de pratiques du bassin? La barre-outils a été élaborée en ce sens. Elle représente une synthèse des mécanismes de modélisation des produits dérivés intégrés (érosion, inondation, eau souterraine, etc.) à l'échelle globale et à l'échelle locale.

Cette barre d'outils dans l'interface d'ArcGIS permet la prise en charge de l'ensemble du concept de SIAD (incluant les fonctionnalités d'Arc Hydro, de *Model Builder* et de fonctionnalités programmées en Arc Object / Visual Basic). Cette barre a permis la prise en compte de l'ensemble des modules utilisés pour l'élaboration des produits précédemment cités ainsi que des produits intermédiaires comme l'analyse temporelle, la visualisation des données, la cartographie (précipitations, évapotranspiration et bilan hydrique), l'information contextuelle sur la disponibilité en eau, les eaux de surface, etc., comme annoncé dans la section de la gestion des barrages.

3. Conclusions et perspectives

Le projet SIAD_GBV pour la gestion par bassin versant représente une bonne illustration de la pertinence du développement de savoirs et d'un savoir-faire pour l'utilisation des données d'observation de la terre, des SIG, de différentes catégories d'approches de modélisation. Les différentes applications développées ont permis d'atteindre deux grands objectifs par rapport à la mise en œuvre du SIAD, en plus de la dimension formation du personnel. D'une part, vérifier la cohérence globale-locale du système d'information (SIAD). D'autre part, produire de l'information territoriale nouvelle autant à l'échelle globale qu'à l'échelle locale. La dernière étape du travail en a été une de synthèse des mécanismes de modélisation technologique avec la réalisation d'une barre d'outils. Cette barre d'outils permettra

l'automatisation de la plupart des tâches récurrentes dans la préparation et la mise à jour des produits dérivés développés dans les applications précédentes.

Cette application permettra d'améliorer la gestion et la décision dans le bassin Souss-Massa par rapport à : l'érosion et l'envasement des barrages; la gestion des barrages; la vulnérabilité et les risques d'inondation; l'exploration de forages; etc., en plus du suivi des changements dans l'occupation et l'utilisation du sol. Cette démarche projet constitue une bonne illustration pour de nombreuses autres applications basées sur l'intégration structurée des données territoriale multisources, multitudes et multicritères, dans la mise en place de SIAD, pour la gestion des ressources naturelles comme l'agriculture, les forêts, l'eau, le sol, la planification territoriale, les risques majeurs, différentes catégories de problématiques décisionnelles territoriales auxquelles sont de plus en plus confrontées de nombreux pays et pour lesquels ils manquent cruciallement de capacités pour des décisions informées et durables.

REMERCIEMENTS : Ce projet a bénéficié d'un financement de l'Agence spatiale du Canada. L'auteur remercie également Info-Electronics Systems, le Centre Royal de Télédétection spatiale et l'Agence hydraulique du Bassin de Souss-Massa pour leur collaboration.

4. Bibliographie

- ABHSM, 2002 : *Stratégie d'intervention de l'Agence pour la période 2004-2007*. Royaume du Maroc, Min. chargé de l'aménagement du territoire, de l'eau et de l'environnement. Agadir, Maroc, 25 p. + Annx.
- Andreu J., Capilla J., et Sanchis, E., 1996 : « AQUATOOL : a Generalized decision support-system for water-resources planning and operational management ». *Journal of Hydrology*. Vol. 177, pp. 269-291.
- Bédard, Y, S. Larrivée, M.J. Proulx & M. Nadeau, 2004: "Modeling Geospatial Databases with Plug-Ins for Visual Languages: A Pragmatic Approach and the Impacts of 16 Years of Research and Experimentations on Perceptory". S. Wang et al. (Eds.): *Conceptual Modeling for Geographic Information Systems (COMOGIS) Workshop ER2004*, LNCS 3289, pp. 17–30, 2004.
- Carver, S., Kingston, R. and Turton, I. 1998: "Accessing GIS over the web: an aid to public participation in environmental decisionmaking". *GISRUK '98*, Edinburgh, Scotland (web only), <http://www.ccg.leeds.ac.uk/vdmisp/publications/paper1.html>
- Cesur D., Beene K. N., Englehart S. T., 2004: "Integrated GIS based Modeling System for watershed Resources". *Proceedings of the 2004 AWRA Spring Conference: GIS and Water Resources III*. American Water Resources Association. Nashville, (Te). Paper in Conference.
- Dieulin Claudine et Boyer Jean-François, 2004: "SIEREM Système d'Informations Environnementales sur les Ressources en Eau et leur Modélisation". UMR, *HydroSciences*, Montpellier. <http://www.hydrosciences.fr/sierem/produits/dieulin/dieulin.htm> [consulté en décembre 2007]

- Dolcine L., Prével C., Ahluwalia H, Bouguenouch B. 2008 : « Modélisation des apports et gestion des barrages du bassin de souss-Massa (Maroc) ». En Révision.
- Duchemin M., Lachance M., Morin G. et Lagacé R., 2001 : « Approche géomatique pour simuler l'érosion hydrique et le transport des sédiments à l'échelle des petits bassins versants ». *Water Qual. Res. J. Canada*, 36 (3), 435-473 et al., 2001
- Hanna K. S., 2000: "The Paradox of Participation and the Hidden Role of Information. A Case Study". *APA Journal, Autumn 2000. Vol. 66, No. 4, pp. 398 – 410.*
- Heywood, I.; Carver, S.J. , 1994: "Decision Support or idea generation: the role for GIS in policy formulation". *Proceedings Symposium fur Angewante Geographischeh.*
- Heywood, I.; Cornelius, S.; Carver, S.J. (2006) *An Introduction to Geographical Information Systems*. 3rd ed. Prentice Hall/Pearson Education
- Hill, Michael J., Braaten, Robert., Lees, Brian and Veitch, Simon M. (2005). "Multi-criteria decision analysis in spatial decision support: the ASSESS analytic hierarchy process and the role of quantitative methods and spatially explicit analysis". *Environmental Modelling and Software* 20, 955-976.
- Keith, J. E. and Said Ouattar, 2004: "Strategic planning, impact assessment, and technical aid: The Souss Massa integrated water management project". *Journal of Environmental Assessment Policy and Management*. 6(2, June 2004):115.
- Lasalmil A. et Boulal R., 2003 : *Schéma d'aménagement territorial de l'aire métropolitaine d'Agadir* (Projet SATAMA). Rapport synthèse. Coopération Franco-Marocaine et Min. délégué de l'habitat et de l'urbanisme. Agence urbaine d'Agadir. 64 p..
- Lévêque Christian, 2001 : *Écologie : de l'écosystème à la biosphère*. Dunod, Paris, 502 p.
- Longépé Christophe, 2006: *Le projet d'urbanisation du S.I. Management des S.I.* DUNOD, Paris, 296 p.
- Longley, P.A., Goodchild, M., Maguire, D.J., and Rhind, D. 2005. *Geographic Information Systems and Science* (2nd Ed.). Wiley & Sons.
- Maidment, D. R., ed., 2002. *Arc Hydro: GIS for Water Resources*, ESRI Press, Redlands, Ca.
- Mitasova, H., J. Hofierka, M. Zlocha, and R. L. Iverson, 1996, Modeling topographic potential for erosion and deposition using GIS, *Int. Journal of Geographical Information Science*, 10(5), 629-641.
- Murthy KSR (2000) « Groundwater potential in a semi-arid region of Andhra Pradesh: a geographical information system approach ». *Int J Remote Sens* 21(9):1867–1884;
- Paquette, G., 2002 : *La modélisation des connaissances et des compétences, pour concevoir et apprendre*. Presses de l'Université du Québec, 2002, 352 pages
- Prével C., St-Onge B. et Waaub J.-P., 2004: "Aide au processus décisionnel pour la gestion par bassin versant au Québec : Étude de cas et principaux enjeux" . *Cahiers de Géographie du Québec.*, Vol. 48, No. 134, septembre, 2004, pp. 209 - 234.
- Prével C.; Dolcine L.; Brahm A.; Erraji A. et Qaimi A., 2008 : « Mise en œuvre d'un système intégré d'aide à la décision pour la gestion du bassin souss-Massa (Maroc) ». En Révision.
- Rao Srinivasa Y, Jugran KD (2003) « Delineation of groundwater potential zones and zones of groundwater quality suitable for domestic purposes using remote sensing and GIS ». *Hydrogeol Sci J* 48:821–833
- Rosenhead J., 2001: *Rational Analysis of a problematic World, Revisted*. J. Wiley & Sons, New York. 384 p.

- Roy B., 1985: Méthodologie multicritère d'aide à la décision. *Economica, Paris*, 423 p.
- Simonovic, S.P., 2007 : "Sustainable Floodplain Management – Participatory Planning in the Red River Basin". Chapter 9 in Castelletti, A. and R. Soncini-Sessa, editors, Topics on System Analysis and Integrated Water Resources Management, Elsevier, Amsterdam, the Netherlands, pp. 173 – 189, 2007.
- VIEUX, Baxter E., 2001 : *Distributed hydrologic modelling using GIS*. Dordrecht, Pays-Bas : Kluwer Academic. Water science and technology library. ; 38; 293 p.
- Wishmeier W.M. et Smith D.D., 1978: "Predicting Rainfall Erosion Losses – a Guide to Conservation Planning". *Agricultural Handbook, No 537, U.S. Dep. of Agriculture, Washington D.C., 58 p.* Zeiller M., 1999
- Zeiler Michael, 1999 : *Modeling Our World: The Esri Guide to Geodatabase Design*. Redlands, CA: ESRI Press, 1999.
- Ziad A., Baudouin Y. et Prével C., 2007 : *Analyse diachronique pour l'étude des changements dans l'occupation globale du sol, Souss-Massa, Maroc, 1987-2006*. Rapport de recherche, UQAM, 61 p. + Annexes + Dossier numérique.

Sémiologie graphique expérimentale et cartographie du risque d'inondation (SGE)

Kamal SERRHINI* - Juliette ROCHMAN - Sven FUCHS*** - Wolfgang DORNER et Karl SPACHINGER******

* *Docteur en aménagement de l'espace, urbanisme (24^e section du CNU), maître de conférences (MC) au département Génie de l'Aménagement, École Polytechnique de l'Université de François-Rabelais de TOURS. Chercheur / membre permanent de l'UMR 6173 CITERES.*

** *Doctorante en aménagement du territoire, UMR 6173 CITERES.*

*** *Docteur à « University of Natural Resources and Applied Life Sciences », Institute of Mountain Risk Engineering at University of Natural Resources and Applied Life Sciences, Wien, Austria.*

**** *Ingénieur diplômé à « University of Applied Sciences », Department of Civil Engineering, University of Applied Sciences, Deggendorf, Germany.*

RESUME : L'article propose une synthèse d'une évaluation de cartes de gestion du risque d'inondation réalisée grâce à étude de la perception visuelle (stratégie du regard) sur différents groupes d'utilisateurs. L'analyse permettra de réaliser un certain nombre de recommandations dans le but d'améliorer la transmission de l'information en matière de risque d'inondation.

ABSTRACT: The article carries out a synthesis of an evaluation of management of the risks of flood maps, thanks to the study of visual perception (strategies of the glance) on various user groups. The analysis will make it possible to carry out a certain number of recommendations aiming at improving the transmission of information as regards risk of flood.

MOTS CLÉS : Sémiologie graphique, sémiologie graphique expérimentale, mouvements oculaires, stratégie visuelle, cartographie du risque.

KEY WORDS: Graphic semiology, experimental graphic semiology, ocular movements, visual strategy, cartography of risk.

1. Introduction

Le but de cette contribution réside dans l'amélioration de l'information des décideurs, des riverains, des associations, des services de secours, des médias... en matière d'information, de prévention (Plan de Prévention des Risques d'Inondations –PPRI) et de gestion d'une crise inondation (Plan de Secours).

Des travaux précédents¹ ont montrés l'insuffisance voire l'absence de la prise en compte de la perception des populations destinataires lors de la conception et de la réalisation des documents graphiques du risque d'inondation.

Dans cette recherche l'idée a été de construire une sémiologie graphique à partir du récepteur et non uniquement de l'émetteur. C'est sur ce principe que repose la sémiologie graphique expérimentale (SGE). Dans une telle perspective, la communication cartographique suivra le processus en boucle suivant : récepteur – émetteur – récepteur. Elle permet donc d'intégrer la perception et les attentes du lecteur dans la conception et la réalisation des documents cartographiques.

Cette méthode a été mise en œuvre pour la cartographie du risque d'inondation pour trois sites² allemands et autrichiens. Les différents scénarios du développement spatio-temporel du risque d'inondation ont abouti à la réalisation de cartes spécifiques qui ont ensuite été analysées en utilisant la méthode dite de SGE. Il s'agira ici d'exposer la démarche et les principaux résultats de cette évaluation.

Après avoir rappelé les bases de la sémiologie graphique expérimentale et présenté le protocole expérimental de celle-ci, nous présenterons de manière synthétique les principaux résultats obtenus. Les conclusions de cette approche ont pour objectif de servir de base à des recommandations sémiologique pour la cartographie du risque d'inondation.

2. Description de l'approche « sémiologie graphique expérimentale » (SGE)

Le processus de production d'un document graphique fait appel à certaines règles et recommandations qu'on appelle Sémiologie Graphique (SG). L'ensemble de ces règles est l'œuvre de spécialistes³ BERTIN (1977) et la communication a généralement lieu depuis le spécialiste (émetteur), vers le récepteur ou destinataire du message (le lecteur : élu, riverain, étudiant...) BEGUIN et PUMAIN (1994). Ce

¹ Notamment le programme « Risque Inondation – RIO » (K. Serrhini, 1999) et le colloque « Évaluation et Prise en compte du Risque » EPR (16-17 juin 2005).

² Le site de « Lower Vils » en Allemagne situé dans le bassin « Donau » à 50 km Sud de Deggendorf; et les sites « Wartschenbach » dans le bassin « Drau » et « Vordergerbach » dans le bassin « Gail », Autriche. Projet RISKATCH du programme de recherche européen Era-Net CRUE.

³ Géographes cartographes, géomaticiens, gestionnaires de l'espace.

type de communication ne permet, a priori, pas de prendre en compte le lecteur (lors de la conception et la réalisation du document graphique).

L'approche SGE considère avant tout les propriétés et les caractéristiques de la perception visuelle d'un lecteur en général et des sujets destinés à lire le document cartographique en particulier (riverains, touristes, décideurs...). Elle est basée sur l'enregistrement des mouvements oculaires déployés par des sujets face à des diapositives (images, cartes), en l'occurrence face à différentes représentations cartographiques du risque d'inondation. L'approche SGE **est rendue possible grâce** à un appareillage spécifique récent utilisé en ophtalmologie : le **vidéo-oculographe**⁴.

Il existe trois principaux types de mouvements oculaires : les saccades, les poursuites et les fixations.

Les saccades : sont des mouvements volontaires, automatiques ou réflexes (P. LARMANDE, A. LARMANDE, 1989). Une saccade est un mouvement oculaire rapide. Sa vitesse, variable, est fonction de son amplitude. La saccade peut être déclenchée par la vision floue d'un objet (c'est-à-dire l'apparition d'un stimulus rétinien périphérique) ou par un stimulus auditif. Selon son sens, une saccade peut être droite, gauche ou verticale. **Les poursuites** : sont des mouvements oculaires plutôt lents. Ces mouvements sont déclenchés par l'observation d'une cible (ou stimulus) mobile saisie donc en vision centrale (vision des détails). Enfin, lorsque le regard reste « fixe » durant un intervalle de temps compris entre 100 et 1000 millisecondes sur une surface inférieure ou égale à 12 x 12 mm, on parle de **fixation**⁵.

3. Protocole expérimental

Le protocole expérimental décrit les principes méthodologiques sur lesquels s'est appuyée l'étude de la cartographie du risque d'inondation à partir de l'approche SGE.

3.1 Candidats ou sujets tests

L'échantillon global comprenait **21 personnes** au total alors que les précédentes recherches relatives à la SGE (Cf. K. Serrhini, 2000 et 2001) n'ont jamais dépassé 15 sujets. Ces 21 candidats se répartissaient en trois groupes :

⁴ <http://www.metrovision.fr>

⁵ Il reste à noter que durant la fixation d'un stimulus immobile, l'œil ne l'est pas en réalité : on peut enregistrer des micros saccades et des micros tremblements de l'œil (Cf. P. Larmande et al., p. 136).

6 « **spécialistes** », en matière de cartographie et du risque d'inondation, 7 « **initiées** » à la cartographie et/ou à la problématique du « risque d'inondation » et 8 « **témoins** » néophytes en matière de cartographique.

3.2 Durée d'exposition ou d'enregistrement

Les cartes ont été exposées pendant 15 secondes ce qui reste supérieure à tous les tests effectués jusque présent⁶ (au plus 10 secondes)

3.2 Les cartes testées

Après échanges entre les partenaires allemands, autrichiens et français 17 cartes ont été retenues dans le but d'étudier différentes variables visuelles.

3.3 Déroulement des tests et enquête cognitive

L'enregistrement était ponctué de phases durant lesquelles le candidat devait remplir le questionnaire constitué de trois parties permettant de tester l'intérêt des différentes cartes en terme de transmission de l'information du risque d'inondation et de sémiologie graphique (position de la légende, du titre, niveau de discrétisation ...).

3.4 Analyse des résultats expérimentaux

Une triple analyse de l'approche SGE a été mise en œuvre.

Une Analyse statistique a sur la base de calculs effectués à partir des résultats obtenus (fournis par le vidéo-oculographe).

Une Analyse spatiale (ou statique) correspondant à l'analyse de l'enregistrement des tracés des mouvements oculaires relatifs à l'exploration d'une carte.

Une Analyse dynamique permettant de déterminer les éléments les plus attractifs (le titre, l'élément central de la carte...), leur ordre d'accès visuel durant et le temps consacré à leur observation (lecture).

4. Analyse des résultats expérimentaux

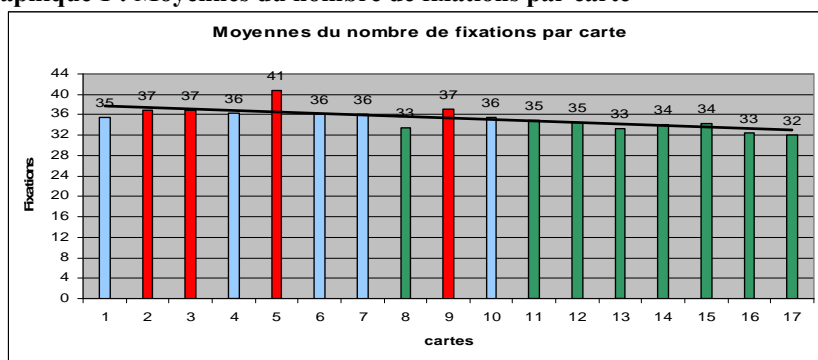
⁶ En effet, étant donné le caractère innovant de la méthode SGE, on a jusqu'à maintenant volontairement limité la durée d'exposition à 10 secondes pour une meilleure exploration de l'intérêt de cette approche.

4.1 Un comportement visuel différencié en fonction des différentes cartes

L'**analyse statistique** portait d'une part sur le nombre de fixation et la durée moyenne des fixations (par carte et par individu) et sur le nombre, l'amplitude et la direction des saccades d'autre part. Dans un souci de synthèse, nous nous sommes limités ici aux principales conclusions issues de l'analyse du nombre des fixations déployées par carte et par groupe de candidats.

L'analyse des résultats statistiques montre que les 17 cartes n'ont pas suscité le même niveau d'exploration visuelle, ce qui apparaît nettement sur le graphique 1 ci-dessous. Il est en effet possible d'identifier trois groupes⁷ de cartes, en fonction de la moyenne du nombre de fixations qu'elles ont générées. Des cartes ont suscitées une exploration visuelle relativement intense (**bâtons en rouge**, Graphique 1) d'autres une exploration visuelle moyenne (**bâtons en bleu**) et certaines ont enfin suscité une exploration visuelle plus faible (**bâtons verts**).

Graphique 1 : Moyennes du nombre de fixations par carte



La carte (sa composition et les choix en matière de sémiologie graphique) influence donc la façon dont le lecteur la regarde. **La perception d'une carte est donc iconographique.**

4.2. Identification et analyse des stratégies visuelles (analyse spatiale)

L'**analyse spatiale** permet d'identifier les stratégies visuelles déployées pour appréhender les différents éléments composant le document cartographique.

⁷ Nous n'avons pas mobilisé une des méthodes de discrétisation (moyenne et écart-type, quantiles...) pour distinguer ces groupes. En effet, l'écart entre chaque groupe reste relativement faible (moins de 10 unités,) surtout que nous considérons qu'une variation de 2 unités ou moins n'est pas significative.

L'Analyse spatiale des mouvements oculaires **pour l'ensemble de l'échantillon** permet d'identifier des régularités dans ces stratégies d'une carte à l'autre.

Le premier constat tiré de cette analyse est que les zones de texte, les zones colorées et celles de fort contraste attirent le regard. Le regard suit en effet les informations écrites (légende et titre), qu'elles soient positionnées à droite ou à gauche de la carte et suivent l'axe de l'élément central de la carte (la rivière).

Les différents éléments de la carte, leur position et leur composition influencent fortement le comportement visuel des lecteurs. Les stratégies visuelles déployées par nos candidats-tests pour **lire la légende** constitue un exemple révélateur.

La légende, est en effet fondamentale car c'est elle qui permet la compréhension de la symbolique graphique de la carte. Les candidats ont d'ailleurs consacré en moyenne, entre 3 et 7 secondes sur 15 à sa lecture (soit entre 20% et 47% du temps total de lecture). **L'analyse statique** des mouvements oculaires permet de déduire qu'une grande « complexité » de la légende (densité et nombre d'information) a obligé la plupart des candidats⁸ à s'y reporter plusieurs fois, ce qui n'a pas été nécessaire quand celle-ci était plus simple.

Une **légende simple** (composée de 4 à 5 classes et de 2 séries d'informations) génère **1 série** de mouvements oculaires (stratégie visuelle). Une légende **très détaillée** (4 classes et plus de 5 séries d'info) génère **3 séries** de mouvements oculaires.

Ainsi, plus les informations contenus dans la légende sont visibles (contraste de couleur) et accessibles (nombre limité d'informations), moins les lecteurs ont besoin d'y revenir et plus celle-ci sera efficiente en termes de transmission visuelle (rapide) de l'information.

L'analyse spatiale révèle donc, à l'image de l'analyse statistique, que la composition des éléments de la carte et les choix en matière de sémiologie, influencent nettement la façon dont ils sont regardés. La perception d'une carte est **donc iconographique**.

L'Analyse spatiale des mouvements oculaires **par groupe de candidats** révèle des particularités dans la façon appréhendé visuellement les cartes.

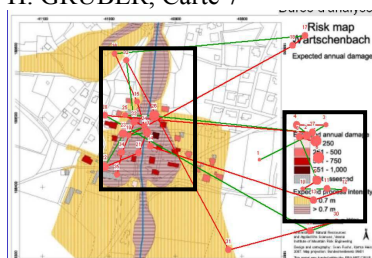
En effet, chacun des trois groupes de candidats (spécialistes, initiés, témoins) voire chaque individu, développe des stratégies visuelles qui lui sont propres en fonction : de sa culture, de ses habitudes visuelles et de son niveau de maîtrise de la cartographie. Pour illustrer ces « particularités » nous reviendrons sur les stratégies visuelles du groupe des « spécialistes » et sur le cas particulier de C. Weber.

Les spécialistes ont tendance à regarder aussi attentivement la légende que la carte (H. GRUBER cartes 7). Au sein de ce groupe, le cas de Christian WEBER constitue en revanche **une exception** car il a consacré **moins de 5 secondes pour la lecture de la légende des 17 cartes** et a même totalement cessé de la regarder à partir de la carte 6. Interrogé sur la particularité de sa stratégie de lecture, C. Weber a expliqué qu'il travaille directement sur les risques de crues des zones cartographiées. Le

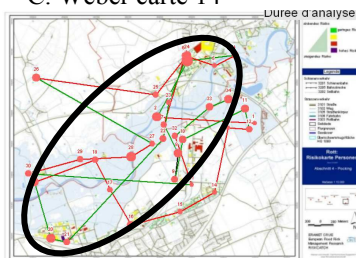
⁸ À quelques rares exceptions près (C. Weber notamment).

niveau d'expertise de ce sujet et sa connaissance du terrain le dispensent donc **de lire la légende.**

H. GRUBER, Carte 7



C. Weber carte 14



L'analyse des résultats obtenus pour les **différents groupes** de candidats a mis en évidence des comportements spécifiques par groupe et parfois selon les individus (cas de C. WEBER). Il apparaît donc clairement que les caractéristiques particulières des individus, influent sur leur manière de lire les cartes. **La perception est donc également anthropique.**

4.3. Analyse dynamique

L'attraction du regard pour les deux principaux éléments (la carte et la légende), est confirmée sans ambiguïté par **l'analyse dynamique**. En effet, si l'on se réfère à la carte 1, dès la troisième seconde, 80 à 90% des sujets ont regardé respectivement l'élément central de la carte et la légende.

Les bandes vidéo permettent par ailleurs de préciser l'ordre dans lequel ces éléments sont appréhendés visuellement. Le regard démarre au centre de la carte⁹. Il se dirige en haut du cartouche de texte et lit successivement le titre (pour peu qu'il soit placé en haut) et la légende sur laquelle il s'attarde. Enfin, le regard se reporte sur la carte où il se focalise sur l'élément visuel principal pour, ensuite se porter sur les éléments visuels secondaire. Ce schéma général correspond à la très nette majorité des candidats. Toutefois, sur la plupart des cartes (en fonction de la densité d'information et de la complexité la légende), ce schéma est plus complexe. Le regard réalise en effet une deuxième voire une troisième série de mouvements visuels.

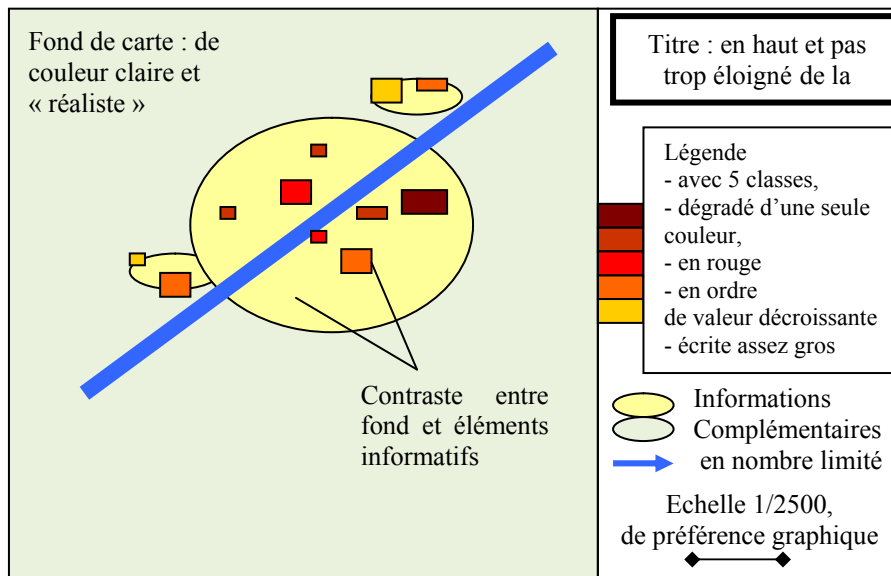
4.4 Analyse cognitive : préférence en matière de sémiologie et de modes de représentation graphique

⁹ Ce qui est en parti lié à la position immobile de la tête dans l'appareillage, et s'y attarde parfois le temps d'une à trois fixations.

L'enquête a notamment fait ressortir de manière claire les préférences des candidats en matière de sémiologie et de représentation graphique. **16 candidats sur 21 (76% des sujets)** préfèrent que la légende se trouve à droite, **15 sur 21 (71%)** préfèrent que le titre se trouve en haut. Pour le nombre de classes (discretisation), **19 personnes sur 21 (90%)** préfèrent une légende simplifiée ne comptant que **5 classes au lieu de 7**. Par ailleurs, **18 candidats sur 21 (85%)** ont préféré le rouge pour représenter le risque. Enfin, **14 personnes sur 21 (plus de 66 % des sujets) ont préféré une échelle plus grande (1/2500)** expliquant qu'elle permettait aux destinataires de mieux se repérer et éventuellement de développer un sentiment de proximité et de familiarité avec la zone concernée.

Sur cette base nous avons réalisé une représentation schématique de la carte qui aurait le plus satisfait nos candidats

Légende, titre et échelle à droite



5. Conclusion

Ainsi, l'analyse des résultats statistiques, statiques et dynamiques des mouvements oculaires, complétée par les résultats de l'enquête cognitive a dans un premier temps permis d'identifier des éléments généraux concernant les stratégies visuelles.

Cette recherche a en effet montré le fort pouvoir attractif : du texte et en particulier du titre et de la légende probablement lié à la culture et à l'éducation, d'une part et des couleurs vives ou foncées (au sein de la carte comme de la légende) d'autre part.

Par ailleurs l'analyse des résultats expérimentaux a démontré que les positions respectives des différents éléments (titre, légende, carte) n'étaient pas neutre dans la façon dont ces composantes étaient appréhendés visuellement. Ce qui nous a permis de conclure à une perception iconographique de la carte. Les analyses par groupe de sujets et par individu a d'autre part mis en évidence le caractère « personnel » des stratégies visuelles démontrant ainsi que la perception était également anthropique.

Enfin, certaines « préférences » concernant des éléments particuliers de sémiologie pour la représentation cartographique du risque ont pu être identifiées. Conformément aux objectifs initiaux de cette étude, ces préférences pourront servir de bases (d'outils) à l'élaboration de nouveaux documents cartographiques du risque d'inondation, plus spécifiquement adaptés à leur public destinataire.

L'étude présente toutefois certaines limites, notamment en matière de représentativité de l'échantillon considéré. Ses résultats (statistique, spatial et dynamique) ne peuvent donc être considérés comme un acquis définitif. Il est évidemment nécessaire de confirmer ou infirmer les différents effets constatés au moyen d'un protocole plus exhaustif et surtout adapté aux cartes et documents évalués.

En dépit de ces limites, ces travaux constituent malgré tout une avancée en matière de recherche européenne en sémiologie graphique appliquée à la cartographie du risque d'inondation. Ils ont en effet permis de mettre en évidence certains éléments susceptibles de servir de base à l'élaboration de cartes plus efficaces et adaptées (en termes de communications graphique et cognitive) à différents types de publics. N'évacuant nullement les réserves précédentes et en attendant d'éventuels travaux complémentaires, il nous est possible de synthétiser les résultats obtenus et de tenter de les intégrer dans une représentation schématique de « la carte idéale » selon nos candidats testés (Cf. figure de la page précédente).

ÉLÉMENTS DE BIBLIOGRAPHIQUE

BERTIN J., Dir. de Braudel F., *La graphique et le traitement graphique de l'information*, Paris, Flammarion, 1977.

BERTALANFFY L. V., *Théorie générale des systèmes*, Paris, Dunod. 1973.

BROSSARD T., « Paysage visible et simulation numérique ». *Bulletin de l'association de géographes français*, vol n°2. - p. 163-174, Paris, 1987.

BRUNET R., - « La carte - modèle et les chorèmes », *Mappemonde*, n°4. - p. 2-6 France, 1986.

CAUVIN C. et al, *Cartographie thématique 1, une nouvelle démarche*, Paris, Lavoisier, 2007.

CHESNEAU E., « Propositions pour une cartographie du risque », *Revue du Comité Français de Cartographie*, n°181, Septembre 2004. - p. 55-70.

LARMANDE A. et P., *Neuro-ophtalmologie*, Paris, Masson, 1989.

MANCHE Y. et al, « Un système d'information dans le domaine des risques naturel : projet SIRVA », *Revue internationale de géographie*, vol 12, n°1, Lavoisier, Paris 2002 - P59-76

MOTET S., Correspondance entre représentation, *Revue internationale de géographie*, vol 12, n°1, Lavoisier, Paris 2002 -p. 9-26.

POTTIER N. et al, - « Évaluation de la Politique publique de prévention des risques naturels », *Risques naturels et aménagement en Europe*, Actes colloque 22-24 Oct. Édition Armand Colin, 2002, - p. 51.

PUMAIN D. et al, *La représentation des données géographiques*. Paris, Armand Colin. 1994.

SERRHINI K., 2007, « Evaluation of Covisibility of Planning and Housing Project », in Ph. Mathis Dir, *Graphs and Networks. Multilevel Modeling*, Edition ISTE, p.193-221.

SERRHINI K., « Sémiologie graphique expérimentale : cas des cartes de l'impact visuel d'un aménagement », *5^e rencontres de Théo Quant* (Actes sur CD-ROM), Besançon, 22 et 23 février 2001. - 19 p.

SERRHINI K., Évaluation spatiale de la covisibilité d'un aménagement. Sémiologie graphique expérimentale et modélisation quantitative. - Thèse de doctorat : Aménagement de l'espace et Urbanisme. - Tours : laboratoire du CESA, 2000.

Ministère de l'Aménagement de Territoire et de l'Environnement, Ministère de l'Équipement, des Transports et du Logement, « Plans de prévention des risques naturels PPR. Risques d'inondation. Guide Méthodologique », 1999, Paris, La Documentation Française. - 123 p.

Intégration de données hétérogènes pour un Système d'Aide à la Décision environnemental

Radhwane Kissi *,** — Jean-Christophe Desconnets * — Thérèse Libourel**

* IRD, US ESPACE Maison de télédétection,
500, rue Jean François Breton
34093 Montpellier Cedex
{nom}@teledetection.fr

** LIRMM UMR 5506 (CNRS-UM2)
161, Rue Ada, 34392 Montpellier Cedex 5
{prenom.nom}@lirmm.fr

RÉSUMÉ. L'article porte sur la notion de Coverage (couverture) introduite dans les standards ISO/OGC. Nous pensons que cette notion offre une solution élégante pour un architecte de SAD. Elle doit autoriser une configuration souple du système en intégrant, de manière dynamique, des données issues de différentes sources de données. Les utilisateurs finaux ayant alors la possibilité de croiser données d'actualité et données patrimoniales de manière uniforme. La mise en œuvre de ce concept au sein d'un système décisionnel environnemental bouscule la conception pré-établie et traditionnelle des systèmes multidimensionnels.

ABSTRACT. This article concerns the notion of coverage introduced by the ISO/OGCs standards. We think that this notion offers an elegant solution for the DS's architecture. It can allow an easy system configuration by dynamically integrating data stemming from different data sources. End-users have possibilities for crossing uniformly current data with stored data. The application of this notion into a decision system for environment will change the pre-established and traditional conception of the multidimensional systems.

MOTS-CLÉS : Systèmes d'Information Géographique, Système d'Aide à la Décision Environnemental, Couvertures, Normes ISO et OGC, Interopérabilité, Observation, Capteurs.

KEYWORDS: Geographic Information System, Decision Systems for environment, Coverage, ISO and OGC Norms, Interoperability, Observation, Sensors.

1. Introduction

L'intérêt pour les systèmes d'information environnementaux ne cesse de croître ces dernières années, compte tenu des initiatives nationales et internationales liées notamment au développement durable.

Parmi les principes énoncés au sommet de Rio (UN, 1992) nous avons relevé :

- Le principe 9 relatif à la coopération en matière de renforcement des capacités qui adjoint à « améliorer la compréhension scientifique par des échanges de connaissances scientifiques et techniques et ceci en facilitant la mise au point, l'adaptation, la diffusion et le transfert de techniques, y compris de techniques nouvelles et novatrices. »
- Le principe 10 relatif à l'accès à l'information et participation aux décisions. « La meilleure façon de traiter les questions d'environnement est d'assurer la participation de tous les citoyens concernés. Au niveau national, chaque individu doit avoir dûment accès aux informations relatives à l'environnement que détiennent les autorités publiques, y compris aux informations relatives aux substances et activités dangereuses dans leurs collectivités, et avoir la possibilité de participer aux processus de prise de décision.»

Des conclusions similaires figurent dans l'initiative Agenda 21 (Agora21, 2005) et la Convention d'Aarhus⁸ en 1998 (notamment les articles 4 et 5), ainsi que dans la déclaration concernant le volet de l'information pour la prise de décision du Sommet Mondial sur le Développement durable de Johannesburg en 2002. « Dans le cadre du développement durable, chacun est un utilisateur et un fournisseur d'informations, au sens large. Il faut entendre par là, des données, des renseignements, des expériences présentées de façon appropriée et des connaissances. Le besoin d'information se fait sentir à tous les niveaux, du niveau national et international chez les principaux décideurs au niveau local et à celui de l'individu. . . ».

Finalement, les applications dédiées à une gestion pertinente de l'information dans un programme de Développement Durable, doivent faciliter, par une « présentation appropriée », le partage et l'intégration des connaissances et des ressources existantes dans diverses disciplines scientifiques afin de permettre, à terme, de traiter et analyser au mieux les questions complexes soulevées.

Dans les nombreux travaux réalisés autour des systèmes décisionnels environnementaux, les auteurs se sont focalisés sur l'analyse et la restitution finale de celle-ci (Rivest et al, 2002) (Bédard et al, 2005) (Tchounikine et al, 2005) (Miquel et al, 2002), en se référant aux approches dites « datawarehouse ». Peu, à notre connaissance, ont porté leur attention sur la manière dont les données ont été acquises et mises à disposition ; le plus souvent les expérimentations d'évaluation se limitent à des jeux de données existants propres ou émanant de production nationale.

L'objectif de cet article est, en raisonnant en tant que concepteur génie logiciel, de donner à l'architecte du Système d'Aide à la Décision (SAD) les moyens de construire, de manière la plus aisée possible, le système approprié. Pour cela, nous devons nous focaliser sur la chaîne complète de traitements nécessaires depuis l'acquisition jusqu'à la restitution de données afin de rendre accessible, à tout utilisateur (décideur, ou citoyen) du système, l'analyse décisionnelle à partir du croisement de données pérennes déjà archivées avec des données d'actualité provenant de différentes sources d'acquisition et ceci de manière automatique.

Les champs d'applications concernent les phénomènes naturels ou anthropiques de type continu qui varient spatialement, temporellement ou spatio-temporellement comme la protection de l'environnement, la météorologie, le traitement des épidémiologies, la désertification, la pollution...etc.

Afin de poser les principes de la mise en œuvre de tels systèmes (qui bien sûr doivent assurer l'interopérabilité entre les divers étapes d'acquisition et de transformation des données observées), nous avons analysé les différentes normes existantes, relatives à l'information géographique, aux observations et aux types des données acquises spécifiées conjointement par l'organisation ISO (International Standard Organisation) et le consortium OGC (Open Geospatial Consortium).

L'article est organisé en quatre parties. La section 2 présente le contexte, la problématique et les définitions de base. La section 3 présente le concept de couverture au travers des normes qui le définissent. La section 4 est consacrée à la présentation des normes qui permettent l'obtention de couvertures et finalement on clôt le propos par une conclusion et des perspectives.

2. Contexte et généralités

2.1. GENERALITES

Historiquement, pour prendre des décisions, l'homme utilisait les cartes géographiques et les informations qu'elles contenaient. La construction des cartes passait par plusieurs étapes, de l'information obtenue par des observations faites sur le monde réel par des êtres humains qualifiés d'experts, à la représentation schématisées de ces informations et ceci afin d'offrir aux décideurs qu'ils soient des dirigeants, des experts ou de simples personnes un moyen commode pour la prise de décision.

La vision actuelle de la décision par cartes géographiques a changé avec l'apport de l'outil informatique qui a amélioré en matière de qualité et de rapidité le processus de réalisation de ces cartes. Il a permis aussi une représentation plus précise et réaliste du monde réel, mais un certain nombre de questions restent encore en suspens, notamment autour du traitement général de la vision « champ continu » de l'information.

Dans cet article, nous abordons un des problèmes sous-jacents à l'aide à la prise de décision sous l'angle aide à l'architecte de SAD. Un système d'aide à la décision (SAD) environnemental spatio-temporel doit permettre l'analyse de phénomènes naturels et anthropiques, les données alimentant ces systèmes proviennent de diverses sources et moyens d'acquisition. Le nombre et le type de capteurs dans ce domaine sont variés et hétérogènes, par suite la restitution, des informations et l'échange de ces informations posent un vrai défi en terme d'interopérabilité. Un architecte expert de SAD doit pouvoir construire et mettre à disposition des utilisateurs des éléments d'informations en provenance soit d'observations, soit de données patrimoniales et pour cela il lui faut disposer d'une chaîne de traitements depuis l'observation jusqu'à la restitution (cf. Figure 1).

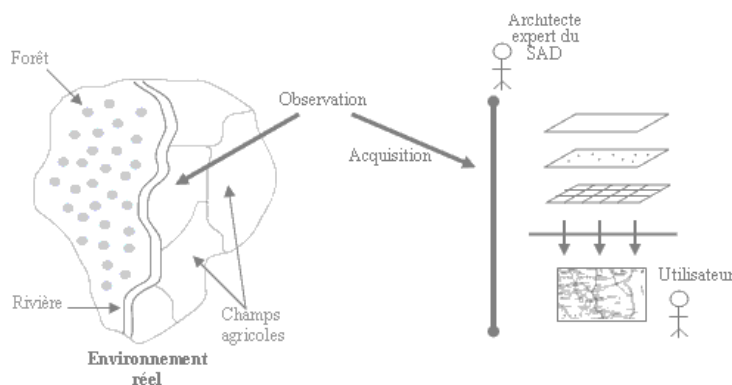


Figure 1. Processus général de mise en œuvre d'un SAD

REMARQUE. — *La vision du SAD que nous envisageons est donc « dynamique » (pas de schéma multidimensionnel a priori).*

2.2. DEFINITIONS

Etant dans le domaine de l'information géolocalisée, nous nous référons à la norme ISO 19101 intitulée *Modèle de référence* (ISO, 2002a), qui sert de cadre à toutes les autres normes de la gamme ISO 19100. Dans ces normes, les entités du monde réel sont abstraites sous la forme de *feature instances* (instances, ou objets) et de *feature types* (types d'objets, ou classes) cf. Figure 2.

NOTE. — *Feature: Abstraction of real world phenomena (ISO, 2005b). A feature may occur as a type or an instance. Feature type or feature instance shall be used when only one is meant.*

Les classes d'objets regroupent des propriétés ou des attributs (*PropertyType*) d'une entité spécialisés en *AttributType* et *AssociationRôle*. Parmi les *AttributType* certains peuvent être valués par des observations.

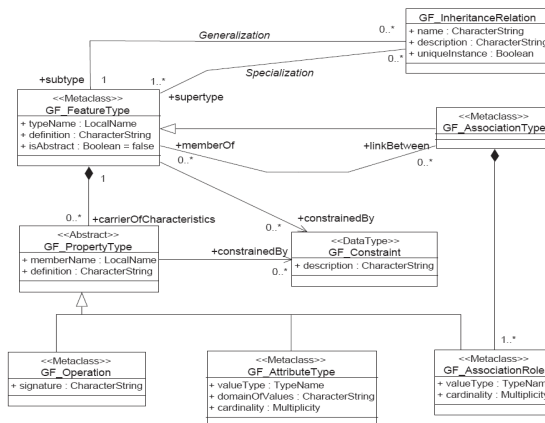


Figure 2. General feature model (ISO 19109)

REMARQUE. — la notion de feature induit souvent une ambiguïté, car elle permet un appariement immédiat avec la notion d'objets du monde vecteur mais elle peut aussi être déclinée pour la vision champs continus.

Dans la norme OGC intitulée *Observations and Measurements* (O&M) (OGC, 2007g), un attribut, peut être valué à la suite d'une observation instrumentée obéissant à un protocole d'estimation (*Procedure*) cf. figure 3. Le protocole d'estimation se fait par un système de capture ou un capteur (*Sensor*).

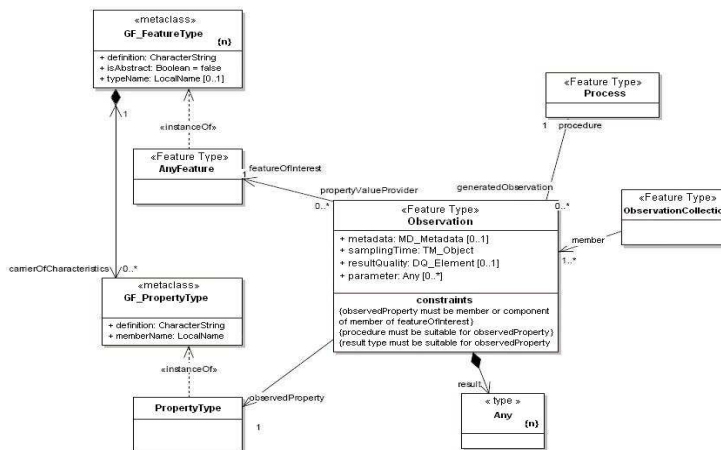


Figure 3. Modèle conceptuel de base d'O&M

La norme O&M désigne comme phénomène une propriété d'une entité du monde réel. La plupart des phénomènes existants sont définis dans l'ontologie *SWEET* (SWEET, 2006).

Exemple : En foresterie, forêt, arbre, etc. sont des *features*, la *feature* une forêt particulière est une instance de la class forêt qui possède comme propriétés : Superficie, Densité des arbres, Age moyen des arbres..., sa propriété Superficie peut avoir comme valeur **50 Km²**, pour le phénomène densité sa valeur est acquise à partir d'une photo obtenue par un capteur qui est un satellite.

2.3. CHAÎNE DE TRAITEMENTS

Le mécanisme général de la chaîne de traitements que nous souhaitons étudier correspond au schéma présenté ci-dessous.

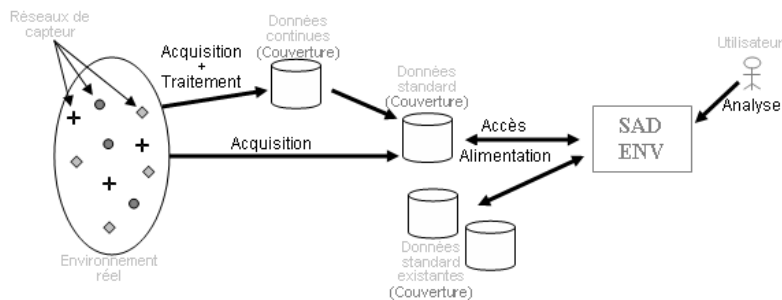


Figure 4. Vue d'ensemble de la chaîne de traitements

En partant d'un ou plusieurs réseaux de capteurs (cf. Figure 4), différents phénomènes continus d'un environnement réel sont observés (ces phénomènes présentent des variations spatiales et/ou temporelles) conformément aux normes de l'initiative SWE (Botts et al, 2007), des données standards sont fournies spécifiées dans la norme ISO 19123 (ISO, 2005c). Cette première étape est dite étape d'acquisition ou d'observation.

Les données issues d'une étape d'acquisition sont inévitablement de type discret. Une deuxième étape, non obligatoire, consiste à convertir les données discrètes en données continues par un traitement conforme, lui aussi, à la norme ISO 19123. Par la suite ces données standards seront soit utilisées comme données d'actualité, ou soit archivées pour intervenir ultérieurement.

Finalement l'étape accès aux données issues des deux précédentes étapes, alimentation du SAD, restitution à l'utilisateur se fait par un mécanisme basé sur une fonction spécifiée par la norme ISO 19123. Ceci fait appel à la notion de couverture (*Coverage*) qui est donc au cœur de notre réflexion.

2.4. EXEMPLE ILLUSTRATIF

Un responsable au sein d'un organisme de santé publique qui s'occupe de la prévention des risques liés aux personnes atteintes par des maladies respiratoires a besoin de croiser sur le territoire de sa commune :

- des données d'actualité issues de capteurs et de détecteurs : les taux d'humidité de l'air, les températures de l'air, les taux de pollution et éventuellement d'autres variables relatives à la caractérisation de la qualité de l'air,
- des données historiques ou pérennes : la répartition des personnes malades déclarées par rapport à leurs adresses dans la ville, les plans cadastraux de la ville, les informations des précédentes situations similaires,

Pour pouvoir répondre aux genres de questions : est ce qu'il faut évacuer les personnes souffrantes de ce type de maladie habitant tel ou tel quartier ? Quels sont les endroits de la ville qui présentent le moins de danger en terme de contamination?

Un citoyen souffrant de ce type de maladie a besoin lui aussi de ces outils pour savoir quelle est la partie de la ville qui lui convient le plus.

Le SAD doit permettre de croiser les données observées entre elles et éventuellement avec d'autres données existantes considérées plus ou moins utiles pour l'utilisateur, et doit, par exemple, restituer des cartes analytiques montrant les zones de la ville, classées en fonction de la gravité du risque.

3. Couverture

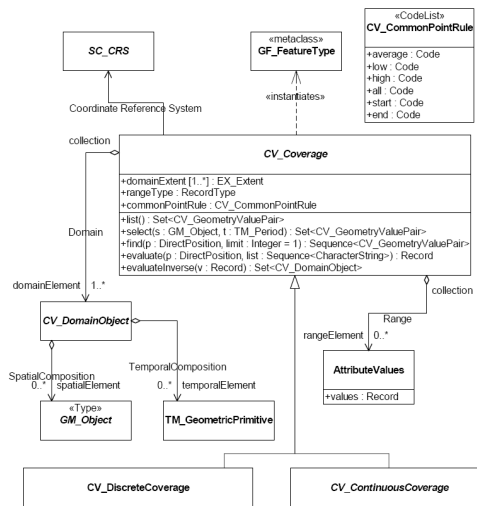


Figure 5. Modèle conceptuel de base d'une couverture (CV_Coverage)

Le principe d'associer des valeurs aux propriétés d'une entité géographique est spécifié dans la norme ISO 19109 (ISO, 2005b) mais dans l'esprit de cette norme la *feature* ne supporte que des valeurs constantes, et c'est pour cette raison que ISO a défini la norme ISO 19123.

La norme ISO 19123 désigne par une couverture (*Coverage*) une *feature* ayant des valeurs de propriétés constantes ou variables et possédant une fonction d'accès à ces valeurs. Nous retrouvons donc ici la volonté de couvrir aussi la vision champ continu évoquée précédemment.

Une couverture permet surtout de faire abstraction du type de ses composants spatiaux, temporels et spatio-temporels. En effet, la norme ISO 19123 - *Schéma de la géométrie et des fonctions de couverture* (ISO, 2005c) définit le type de données couverture (*CV_Coverage*), les différentes structures de découpage d'une couverture et la fonction standard d'évaluation (*evaluate*).

Le domaine (*Domain*) d'une couverture (cf. Figure 5) regroupe un ensemble d'objets appelés *CV_DomainObject*, l'étendue et les limites de la couverture sont définies dans *DomainExtent* qui est de type *Ex_Extent* (ISO, 2003b). *DomainExtent* permet de définir l'étendue et de savoir quelles sont les positions qui appartiennent au domaine d'une couverture. Un *CV_DomainObject* est composé d'un objet géométrique (entité géographique) de type *GM_Object* (ISO, 2003a) et/ou d'un objet temporel de type *TM_GeometricPrimitive* (ISO, 2002b).

La plage (*Range*) d'une couverture contient les valeurs des propriétés des entités géographiques appartenant au domaine de cette couverture, le type de valeurs enregistrement (*Record*) (ISO, 2005a) peut être de n'importe quel type. *CV_GeometryValuePair* associe les éléments du domaine avec les éléments de la plage dans une couverture. Une couverture doit supporter la condition d'homogénéité (c'est-à-dire la propriété concernée doit être définie dans toutes les positions du domaine d'une couverture).

Toute structure de données Couverture supporte la fonction standard *evaluate*. La méthode *evaluate* ou son inverse *evaluateInverse* est un mécanisme unifié d'accès aux données contenues dans une couverture. *evaluate* reçoit en entrées les coordonnées (*DirectPosition*) d'un objet spatial, temporel ou spatio-temporel (*CV_DomainObject*) définies dans le domaine de la couverture et renvoie en sorties des données de type enregistrement, définies dans la plage de la couverture et inversement pour son inverse.

Exemple : supposons que nous disposons de stations météorologiques (cf. Figure 6) se trouvant dans une région administrative; chaque département ayant une station. La couverture couvre la région (*DomainExtent*) et les départements sont les entités géographiques (*GM_Object*). Les valeurs possibles de température de l'air mesurées par les stations dans des instants de temps (*TM_GeometricPrimitive*) représentent la plage de la couverture, le domaine représente l'ensemble des départements et les instants de mesure. *evaluate* permet de connaître les températures acquises, par des

séries d'observations, à partir d'une position (*DirectPosition*) dans la couverture : *evaluate* (position $x, y, z, t = 25^\circ$ et *evaluateInverse* permet de connaître les positions dans la région ou la température est 20° dans les différents instants de temps.

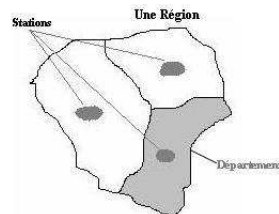


Figure 6. Stations météorologiques d'une région

ISO 19123 distingue deux types, spécialisation du type couverture : le type couverture discrète et le type couverture continue.

Dans une couverture discrète, l'évaluation donne le même résultat pour tout point d'un *CV_DomaineObject* qui englobe la position d'observation, par contre dans une couverture continue l'évaluation est calculée en tout point du domaine par interpolation. Dans la réalité, les données issues des observations sont des données discrètes (section 4.2), une couverture continue n'est qu'une couverture discrète associée à une fonction analytique ou une des méthodes d'interpolation spatiales supportée par la structure de découpage de la couverture pour approximer les valeurs en toute position.

L'approximation se fait à chaque fois que la fonction *evaluate* est appelée, dans certains cas quelques valeurs approximées sont enregistrées et cela en fonction du type de découpage. ISO 19123 définit une collection de méthodes d'interpolation spatiales : *nearestneighbor*, *linear*, *quadratic*, *cubic*, *bilinear*, *biquadratique*, *bicubic*, *lostarea* et *barycentric*.

Les objets géométriques (*GM_Object*) composant le domaine d'une couverture discrète peuvent être de différentes formes mais ISO 19123 simplifie cette complexité en définissant des couvertures discrètes composées que par des objets géométriques de dimension identiques : points 0-D (*CV_discretePointCoverage*), courbes 1-D (*CV_discreteCurveCoverage*), cellules formant un tableau 2-D (*CV_discreteGridPointCoverage*), surfaces 2-D (*CV_discreteSurfaceCoverage*) et solides 3-D (*CV_discreteSolidCoverage*).

L'utilité du découpage est de permettre l'usage de différentes méthodes d'interpolations afin de comparer la qualité des résultats obtenus. ISO 19123 définit plusieurs structures de découpage (qui ne s'applique pas forcément à tous les types de couvertures discrètes) et qui donnent des couvertures continues supportant certaines méthodes d'interpolation :

- Couverture continue à découpage en cellules de formes quadrilatérales (*CV_ContinusQuadrilateralGridCoverage*) : supporte *nearestneighbor*, *bilinear*, *biquadratique* et *bicubic*.
- Couverture continue à découpage en cellules de forme triangulaire irrégulière (*CV_TINCoverage*) : supporte *nearestneighbor* et *barycentric*.
- Couverture continue à découpage en cellules de forme hexagonale (*CV_HexagonalGridCoverage*) : supporte *nearestneighbor* et *lostarea*.
- Couverture continue à découpage en cellules de forme de polygone selon la méthode de *Thiessen* (*CV_ThiessenPolygoneCoverage*) : supporte *nearestneighbor* et *lostarea*.
- Couverture continue de segmentation de courbes (*CV_SegmentedCurveCoverage*) : supporte *nearestneighbor*, *linear*, *quadratic* et *cubic*.

Exemple : une collection d'observations faites par des stations météorologiques réparties dans une région (figure 6) donne une couverture discrète contenant les températures de l'air en certains points à différents instants de temps. Deux possibilités se présentent, soit on fait un découpage du domaine spatial de la couverture en cellules de formes quadrilatérales, soit on fait un découpage (en fonction de la position des points) en cellules de forme de polygone de *Thiessen* où les points représentent les centroïdes des polygones, ou de forme triangulaire irrégulière où les points représentent les sommets des triangles. Dans la première possibilité on peut utiliser les méthodes d'interpolation : *nearestneighbor*, *bilinear*, *biquadratique*, ou *bicubic*, par contre dans la deuxième possibilité d'autres méthodes d'interpolation : *barycentric* ou *lostarea* peuvent être appliquées. L'intérêt de la possibilité d'utiliser différentes méthodes d'interpolation est de permettre à l'utilisateur de comparer la qualité des résultats qu'il a obtenus.

4. Acquisition des couvertures

Vu la diversité et l'hétérogénéité des capteurs qui sont classés en fonction de leur localisation en :

- Capteurs fixes : par exemple un thermomètre dans une station météorologique,
- Capteurs mobiles ou dynamiques : par exemple un satellite orbitaire de surveillance.

L'organisme OGC propose via l'initiative *Sensor Web Enablement* (SWE) (Botts et al, 2007) d'offrir l'interopérabilité grâce à une vue unifiée regroupant un ensemble de standards pour l'exploitation, le contrôle et l'accès aux observations de capteurs et de tous autres types de système de capture déployés sur le web (cf. la Figure 7).



Figure 7 : Vue d'ensemble des normes de l'initiative SWE

Cette initiative autorise le déploiement de systèmes de capture, relatifs à l'observation de phénomènes continus, communicants avec une interface web via des services conformes aux spécifications *Sensor Observation Service* (SOS) (OGC, 2007a), *Sensor Planning Service* (SPS) (OGC, 2007f), *Sensor Alert Service* (SAS) (OGC, 2007c) et *Web Notification Service* (WNS) (OGC, 2003a). SOS fournit une interface (API) utilisant un langage de description : *Sensor Model Language* (SensorML) (OGC, 2007e) ou *Transducer Markup Language* (TML) (OGC, 2007b) qui renseigne sur les caractéristiques des capteurs utilisés et sur les résultats d'observations obtenus conformes à la norme *Observations and Measurements* (O&M).

Nous ne rentrerons pas dans le détail de chaque standard, et par la suite nous nous intéresserons qu'à O&M et SensorML pour la réalisation de l'étape d'acquisition des observations.

4.1. O&M

Le protocole d'expérimentation ou la procédure qu'utilise O&M peut être soit liée à un instrument, un capteur, un observateur humain, soit un algorithme d'estimation de la valeur de la propriété observée, soit la description du processus utilisé pour l'estimation du résultat. Cette description qui est utile pour le calcul des taux d'erreur se fait par le langage SensorML qui est définie dans la norme SensorML (section 4.2).

Le résultat de l'étape d'acquisition peut être soit constituée d'une observation faite sur une propriété d'une entité géographique (*featureOfInterest*), soit d'une série d'observations (*ObservationCollection*) menées sur une ou plusieurs propriétés d'une entité géographique en utilisant des capteurs qui peuvent être de toutes sortes.

Une série d'observations comportant l'ensemble des observations faites sur les propriétés d'un ensemble d'objets de *DomainObject* en une période de temps identique donne comme résultat une couverture discrète (*CV_DiscreteCoverage*). La

particularité des propriétés observées réside dans le fait qu'elles varient spatialement, temporellement ou spatio-temporellement.

NOTE. — Dans O&M trois types de couvertures discrètes sont définies : couverture discrète de points (*CV_DiscretePointCoverage*) qui est spécifiée dans ISO 19123, couverture discrète d'éléments (*CV_DiscreteElementCoverage*), et couverture discrète en instant de temps (*CV_DiscreteTimeInstantCoverage*) qui sont définies dans le document *GML Encoding of Discrete Coverages* (OGC, 2007d) qui spécifie des spécialisations additionnelles de couverture discrète (*CV_DiscreteCoverage*) par rapport à ISO 19123.

4.2. SENSOR ML

SensorML est un langage qui permet de décrire en *eXtensible Markup Language* (XML) les informations relevant de métadonnées relatives à un capteur (ou à un système de capture) : fabricant, numéro de série, position, SRC spatial, SRC temporel, etc..., les phénomènes observables, les services offerts et le protocole ou la procédure d'estimation du résultat. Il est utilisable pour décrire n'importe quel type de capteur, allant du simple thermomètre visuel au microscope électronique complexe. Il est compatible avec O&M et aussi avec d'autres fournisseurs de données observées mais par contre il ne décrit ni le résultat ni la description matérielle d'un capteur.

Les couvertures acquises peuvent donc constituer des données pérennes accumulées ou peuvent constituer des données d'actualité. Le croisement de ces couvertures est donc celui que l'on propose de présenter à des fins d'analyse. La définition et l'accès aux informations sont alors traitées de manière uniforme.

5. Conclusion

Dans cet article nous présentons l'idée qui, à nos yeux, préfigure comment l'architecture d'un SAD environnemental pourrait être conçue. Nous nous sommes intéressés au cheminement des données au sein de cette architecture, de l'acquisition à la restitution en veillant à ce que ceci soit conforme aux normes internationales existantes. L'intérêt est que des données d'actualité issues de capteurs peuvent être croisées avec des données patrimoniales. Le concept de couverture représente l'élément phare de la mise en place de cette architecture. En effet, il peut être perçu comme une notion générique d'accès à des informations géoréférencées contenues dans des données spatiales, temporelles ou spatiotemporelles de n'importe quel type (raster, vecteur, ...). La combinaison des fonctions *evaluate* ou *evaluateInverse* correspondent au mécanisme d'accès. La notion de couverture est essentielle car ce concept peut encapsuler de manière homogène tout type de données.

Les couvertures discrètes en provenance de différents capteurs, pourront être obtenues en faisant appel à un des services web spécifiés par l'initiative SWE : SOS ou SPS. Par contre le passage par un service web, tel que WCPS (Baumann, 2006) (OGC, 2006b), qui offre des interfaces pour accéder aux couvertures afin d'alimenter le SAD, est très contraignant. Pour cela on se focalisera pour un premier temps dans le développement d'un mécanisme nécessaire à l'accès aux données couvertures.

Dans les différents domaines relevant de l'environnement, les analyses décisionnelles nécessitent l'intervention de plusieurs données hétérogènes d'une manière instantanée. Les solutions existantes utilisent une intégration des données matérialisée sous forme d'entrepôt (*Datawarehouse*) qui nécessite une conception a priori plus ou moins complexe et figée. L'implantation de la notion d'une couverture, au sein d'une architecture logicielle, devrait rendre l'intégration des données plus dynamique, l'analyse multidimensionnelle plus naturelle et surtout la conception de SAD plus aisé. Le concept de couverture ouvre la perspective de système multidimensionnel dynamique et de ce fait rend la conception pré-établie et traditionnelle caduque.

Au-delà de cette première analyse, notre objectif final est de concevoir un ensemble de bibliothèques logicielles nécessaires à l'implémentation de l'architecture proposée dans notre article. Par la suite, nous testerons cette implémentation en évaluant les qualités des données restituées, notamment en mesurant de manière précise les biais introduits par les transformations et approximations effectuées durant les différentes étapes de l'exécution.

6. Bibliographie

- Agora21., Information et gouvernance, 2005, URL: <http://www.agora21.org/information.html>.
- Baumann P., « Towards a Standard for Interoperable Earth System Raster Services », *23rd FIG Congress*, Munich, Germany, 8 – 13 October 2006.
- Bédard, Y., Proulx M.J., Rivest S., « Enrichissement du OLAP pour l'analyse géographique : exemples de réalisation et différentes possibilités technologiques », *Revue des Nouvelles Technologies de l'Information - Entrepôts de données et l'Analyse en ligne*, Cepaduès-Éditions, France, pp. 1-20, 2005.
- Botts M., Percivall G., Reed C., Davidson J., « Sensor Web Enablement: Overview And High Level Architecture », *OGC 06-050r2 version 2.0 OpenGIS® White paper*, 19 juin 2007.
- ISO., ISO-TC211 - 19101 information géographique - Modèle de référence, Standard international, 2002a.
- ISO., ISO/TS - 19103 information géographique - Schéma de langage conceptuel, Standard international, 2005a.

- ISO., ISO-TC211 - 19107 - information géographique - Schéma spatial - Standard international, 2003a.
- ISO., ISO-TC211 - 19108 information géographique - Schéma temporel, Standard international, 2002b.
- ISO., ISO-TC211 - 19109 - information géographique - Règles de schéma d'application - Standard international, 2005b.
- ISO., ISO-TC211 - 19115 - information géographique - Métadonnées - Standard international, 2003b.
- ISO., ISO-TC211 - 19123 - information géographique - Schéma de la géométrie et des fonctions de couverture - Standard international, 2005c.
- Marconett B., A Prototype Network-Distributed Sensor Web Observation Service, Aero Associates Project Paper No. SOS-Study-DMarconett-17July06, Computer Science Department, University of California, Davis, July 2006.
- Miquel M., Bédard Y., Brisebois A., « Conception d'entrepôts de données géospatiales à partir de sources hétérogènes, exemple d'application en foresterie », *Ingénierie des Systèmes d'information*, Vol. 7, No. 3, 2002, pp. 89-111.
- OGC., OGC 03-008r2 - Web Notification Service - OpenGIS® Discussion Paper, 2003a.
- OGC., OGC 06-009r6 - Sensor Observation Service - OpenGIS® Implementation Standard, 2007a.
- OGC., OGC 06-010r6 - Transducer Markup Language - OpenGIS® Implementation Specification, 2007b.
- OGC., OGC 06-028r5 - Sensor Alert Service - OpenGIS® Implementation Specification, 2007c.
- OGC., OGC 06-035r1 - Web Coverage Processing Service - OpenGIS® Implementation Specification, 2006b.
- OGC., OGC 06-188r1 - GML Encoding of Discrete Coverages - OpenGIS® Best Practices Document, 2007d.
- OGC., OGC 07-000 - Sensor Model Language (SensorML) Implementation Specification - OpenGIS® Implementation Standard, 2007e.
- OGC., OGC 07-014r3 - Sensor Planning Service - OpenGIS® Implementation Specification, 2007f.
- OGC., OGC 07-022r1 - Observations and Measurements - Part 1 - Observation schema, OpenGIS® Implementation Standard, 2007g.
- Rivest S., Bedard Y., Marchand P., « Toward better support for spatial decision making: Defining the characteristics of spatial on-line analytical processing (SOLAP) », *Geomatica*, 2001, vol. 55, n°4, pp. 539-555.
- SWEET., Semantic Web for Earth and Environmental Terminology ontology, 2006, URL: <http://sweet.jpl.nasa.gov/ontology/property.owl/>.

Tchounikine A., Miquel M., Laurini R., Ahmed T., Bimonte S., Baillet V., « Panorama de travaux autour de l'intégration de données spatio-temporelles dans les hypercubes ». *Revue des Nouvelles Technologies de l'Information - Entrepôts de données et l'Analyse en ligne*, France, Cépaduès-Éditions, 2005, 21-33 p.

UN, Report of the United Nations Conference on Environment and Development, Rio de Janeiro, United Nations, 3-14 June 1992. URL: <http://www.un.org/documents/ga/conf151/aconf15126-1annex1.htm>

Un système d'aide pour les sciences de la terre : un exemple géotechnique

Faure Nicolas

Equipe MODEME

Centre de recherche de l'IAE

Université Jean Moulin Lyon III

Nicolas.faure@univ-lyon3.fr

RÉSUMÉ. Cet article présente certains aspects de la conception d'un système d'aide en géotechnique dans le cadre d'un projet industriel. Il présente un certain nombre de réflexions et d'observations issues de l'expérience relatives à l'adéquation d'un tel système et d'une communauté d'experts dans le cadre d'une science empirique, c'est-à-dire portant sur le façonnement d'un environnement naturel, avec toutes les incertitudes que cela suppose au niveau des connaissances.

ABSTRACT. This paper deals with some aspects of a knowledge-based system for geotechnics, mainly regarding its use by a community of subject experts, and some considerations about how such systems can be adapted to the uncertainties of such a community, interacting with a natural environment.

MOTS-CLÉS : géotechnique, représentation des connaissances, aide à la modélisation, approche ontologique

KEYWORDS: geotechnics, knowledge representation, modelling support, ontological approach

1. Introduction

Le projet RAMCESH est un projet ayant trait à la réalisation d'un système d'aide dans le cadre de la géotechnique, mais dont les conclusions sont transposables à tout domaine présentant des spécificités similaires, notamment un domaine d'études empiriques ayant trait à l'environnement rassemblant des disciplines variées.

Issu d'une volonté industrielle et réalisé en collaboration étroite avec des experts du domaine, son objectif est de produire un outil d'aide à la représentation et manipulation des connaissances qui s'adresse directement aux experts de domaine.

2. Géotechnique : notre approche

La géotechnique est la science qui se préoccupe des interactions d'un sol et d'un construit, route, bâtiment ou ouvrage d'art. A ce terme de géotechnique est parfois substitué celui de géosciences, mais cette dernière appellation apparaît moins éclairante sur la nature empirique de l'activité géotechnique, c'est-à-dire essentiellement fondée sur la relation à un contexte donné. Toute démarche géotechnique ne peut se concevoir qu'en fonction d'une situation réelle ou supposée, d'une contextualisation d'un problème, et ce généralement par analogie avec une situation similaire déjà connue. Cela traduit, de fait, l'activité géotechnique comme un ensemble de savoir-faire et de techniques, comparables au tour de main d'un artisan (Magnan 2002).

Le terme de géosciences est pourtant utilisé parce qu'il est aussi révélateur d'une des caractéristiques fondamentales de ce domaine, à savoir son hétérogénéité de pratiques. De fait, les géosciences constituent l'ensemble des sciences, ou éléments scientifiques, mobilisées dans la pratique géotechnique.

3. Les hétérogénéités, caractéristiques majeures du domaine

Technique

La géotechnique est en effet le point de rencontre de spécialités distinctes : géologie, mécanique des milieux continus et discontinus, chimie, hydraulique, etc... Chacune de ces spécialités implique une certaine conception a priori des éléments géotechniques, une terminologie et des pratiques particulières. Par exemple, le concept, aussi fondamental soit-il, de sol, ne sera pas forcément compris a priori de la même manière par un géologue ou un spécialiste de stabilité des pentes. Pour le premier, un sol est un ensemble de roches situé sous la couche meuble de la surface. Pour le second, en revanche, cette couche meuble, parfois appelée sol détritique et située au-dessus du substratum rocheux rigide, est l'objet de son étude.

Régionale

Un deuxième élément fonde l'hétérogénéité de la géotechnique, à savoir sa disparité régionale ; les types de roches et les configurations de sol sont liées à des régions, ce qui donne donc naissance à des terminologies et des concepts régionaux : en France, le granit du mont d'Ambin est un exemple et évoque une forme particulière de granit et certaines de ses propriétés ; le Toki Granite, autre exemple, est évoqué dans certains articles traitant de projets géotechniques au Japon et définit un autre type de granit, non caractérisé autrement que par sa provenance. Dans l'un et l'autre cas, les spécialistes peu familiers de la géologie alpine ou japonaise sont confrontés à des données incomplètes, car la dénomination particulière du granit implique l'importance de propriétés particulières, de surcroît variables selon les occurrences du concept.

La généralisation des caractérisations régionales n'est pas nouvelle, et transparaît assez facilement avec l'échelle des temps géologiques, le « calendrier géologique », où les étages et sous-étages géologiques portent les noms dérivés des régions typiquement représentatives des sols des époques correspondantes : Maestrichtien (Maestricht), Hettangien (Hettange), Coniacien (Cognac), etc... Ces travaux de dénomination et de calibrage correspondant sont souvent compliqués par le caractère évolutif de ces éléments qu'on pourrait croire figés : l'échelle des temps géologiques est ainsi remise à jour tous les quatre ans.

Ces quelques exemples permettent de brosser le tableau d'un environnement professionnel où les usages ont une importance non négligeable. Dans le cas de l'échelle des temps géologiques, cela correspond à l'enrichissement d'une colonne « autres dénominations », aux limites parfois mal identifiées, mais d'une manière générale cela correspond à l'enrichissement informel d'une polysémie déjà rendue redoutable par l'étendue du domaine géotechnique.

Une connaissance possible ?

Cet état de choses n'est pas seulement le fait de la double hétérogénéité du domaine qui vient d'être décrite, mais aussi de la nature même des projets géotechniques. Ceux-ci débutent généralement par la conception d'un construit inexistant à implanter dans un environnement inconnu. Le sol, en effet, est un ensemble qui échappe aux méthodes d'observations usuelles, et notamment visuelles ; on ne peut en connaître les caractéristiques qu'au travers de tests et de sondages¹ qui, pour aussi précis qu'ils soient, ne permettent qu'une approximation de la réalité. Par ailleurs, un sol donné n'est pas reproductible en laboratoire, et les sondages ont un coût élevé. De fait, un sol est un système complexe dont l'étendue

¹ A ce sujet, la communauté géotechnique conserve à l'esprit que le volume des sondages et le volume du sol auquel on applique leur résultat possède un rapport qui peut surprendre, de l'ordre de un pour un milliard

physique elle-même est soumise à hypothèse, puisque ses caractéristiques peuvent être générées via des interactions à partir d'éléments dont il est a priori impossible de déterminer l'emplacement, la proximité ou même la nature.

On trouve là les raisons fondamentales de l'empirisme mentionné en premier lieu dans cet article, les causes de l'importance de l'expertise et la motivation principale pour le raisonnement par analogie, avec tous les risques qu'il comporte pour les projets de conception. C'est pourquoi la démarche pragmatique (au sens premier du terme, c'est-à-dire dont la validité est déterminée par les résultats) est très souvent utilisée en géotechnique².

La documentation et informatisation du domaine

Paradoxalement, la communauté géotechnique est une communauté fortement mondialisée et relativement communicante, dans la mesure où pléthore de documents d'une nature ou d'une autre sont produits, dans des revues spécialisées ou à l'occasion d'un des nombreux congrès et symposiums internationaux. Un premier constat est que l'on retrouve dans la forme générale documentaire ce qui vient d'être dit : on distingue les documents à haute technicité, généralement relevant d'une spécialité (le plus souvent en physique des matériaux) et les rapports de chantier, essentiellement descriptifs, décrivant situations rencontrées, mesures prises en fonction de ces situations et résultats de ces mesures.

De manière comparable, et pour ce qui concerne la pratique géotechnique en France, le recours à des systèmes d'aide informatique se limite bien souvent à des logiciels de calcul, consacrés à des points spécifiques : dimensionnement de tunnels en fonction d'un sol donné, calcul des confortations nécessaires à la stabilisation d'un talus, estimation des trajectoires de chutes de blocs en fonction de certains paramètres, etc...

La communauté géotechnique française a connu un vaste engouement pour les systèmes à base de connaissance dans les années 1980 et 1990. De nombreux systèmes experts ont donc été mis à l'étude (CESSOL, XPENT, CASTOR, SOUT, PROSPECTOR, etc...), mais rarement utilisés en pratique. Les causes de ce fait sont multiples et ont parfois été analysées (Magnan 1992), mais deux constantes apparaissent :

² C'est l'objet de la méthode principalement employée à l'heure actuelle durant la réalisation de projets géotechniques, baptisée « méthode observationnelle » et qui consiste, pour pallier partiellement à l'inconnaissance de l'environnement, à étudier en temps (presque) réel l'interaction de chaque élément de la structure projetée avec le sol, à mesure de l'avancement du projet.

- le besoin pour l'utilisateur de comprendre l'outil qu'il manipule, au moins ses principes fondamentaux, le rejet des intermédiaires pour le traitement de la connaissance et surtout une adéquation de l'outil à sa pratique usuelle

- l'investissement souvent trop important (sur de longues durées), au moins en temps et en énergie, pour alimenter une base de connaissances en mesure de fournir des réponses satisfaisantes

Les tendances les plus récentes concernant les systèmes d'information visent à faciliter le travail collaboratif, avec notamment l'utilisation de plus en plus accrue des réseaux et de systèmes d'annotations et commentaires concernant des documents de travail, ou de mise à disposition de cas pour constituer une base de références communes. Ce dernier type de système rejoint ce qui était dit plus haut en ce qui concerne la littérature du domaine et la distinction entre articles à haute technicité et descriptifs de cas. On constate du reste que la communauté géotechnique véhicule à travers ces descriptions nombre d'histoires qui augurent favorablement de ce que des techniques comme le storytelling pourraient apporter au domaine.

4. Notre proposition

Face aux spécificités du domaine, le projet RAMCESH s'est plutôt engagé sur la voie de la représentation du réel tel qu'évoqué au sein des différents discours documentaires plutôt que sur un réel modélisé, consensuel et potentiellement formalisé selon les attentes d'une communauté homogène.

Par conséquent, notre proposition ne repose pas sur une conceptualisation a priori du domaine, duquel les éléments intervenants dans le discours ne seraient que des instances, comme il est d'usage dans de nombreux systèmes à base de connaissances (de nombreuses définitions des ontologies le demandent, par exemple). Au contraire, on s'est attaché à ne représenter qu'un consensus minimal, demandant le moins possible d'engagement sémantique ou ontologique (Bachimont, 2000), en bâtissant tout d'abord à partir des documents un réseau lexical, c'est-à-dire un ensemble des termes recensés dans le corpus organisés selon une hiérarchie de spécialisation, rarement remise en question par les experts du domaine³ et tenue pour vraie sur l'étendue du domaine.

Afin de faciliter la prise en main de l'outil par lesdits experts, ce réseau lexical obéit à un certain nombre de contraintes simplificatrices :

- Il ne comporte que des substantifs ou des syntagmes nominaux
- Il est subdivisé en plusieurs « couches métier » ne supportant qu'un seul terme général par terme donné (pour former au final une arborescence), construites selon

³ Lorsque c'est le cas, on conserve les variations en les rapportant à une « couche métier » (cf. *infra*) et en les traitant comme des « points de vue » possibles

une approche par spécialisation et généralisation des termes les plus courants (baptisée middle-out)

- Il permet d'associer à chaque terme un ensemble de synonymes, cas d'usage recensés et variations flexionnelles utilisé pour la reconnaissance des termes pendant l'analyse documentaire

- Il permet également d'associer à chaque terme un ensemble de définitions et commentaires en langage naturel afin notamment de notifier les variations de définition selon les experts (chaque définition est explicitement attribuée à l'utilisateur qui l'a proposée)

- Il ne supporte aucune description catégorique formelle, par propriétés ou attributs

- Il autorise exceptionnellement une relation indéterminée (« de voisinage » ou « d'association ») si elle est estimée nécessaire et surtout vérifiée sur l'étendue du domaine

La deuxième partie de notre proposition consiste en un construit à tendance plus conceptuelle, le granule de connaissances. La même logique sémasiologique⁴ prévaut pour le granule que pour le réseau lexical, à ceci près qu'il s'agit d'exprimer un concept en fonction d'éléments lexicaux présents dans le corpus documentaire. Pour ce faire, on se concentre sur une forme de causalité (fréquente dans les documents, notamment descriptifs et familière aux experts sous la forme de règles de production) qu'on extrait sous la forme de fragments de texte. Chacun de ces fragments de texte est ensuite formalisé selon ses constituants lexicaux à l'aide d'un ensemble de relations sémantiques baptisées « relations métier », au nombre d'une cinquantaine.

Le granule comporte également une partie Univers et Modèle d'Univers⁵, définissant chaque instance et relations d'équivalences entre instance décrite dans le fragment de texte. Le granule (figure 1) et la théorie qui le sous-tend sont décrits plus exhaustivement dans (Faure 2007a, 2007b).

L'ensemble des granules ainsi composés constitue une base de connaissances aisément consultable, dont les divers éléments sont agrégeables selon un processus de scoring déterminant une ressemblance structurelle. Ce processus se justifie puisqu'à la différence des graphes conceptuels (Sowa, 1984), par exemple, la simple

⁴ Entendue ici comme l'ensemble des méthodes visant à faire émerger la signification d'un terme à partir de son occurrence, a contrario d'une approche onomasiologique, qui supposerait la préexistence d'un concept dont le terme ne serait qu'une désignation (et qu'on aurait spécifié par exemple dans une conceptualisation explicite et partagée)

⁵ On s'inspire ici de méthodes de représentation formelles du discours comme la Discourse Representation Theory (Kamp et Reyle, 1993 ; Corblin, 2002) qui emploient notamment ce type d'approche pour résoudre les anaphores et autoriser une représentation dynamique du discours, importantes dans le cas qui nous occupe pour autoriser les agrégations de granules

équivalence des labels ne suffit pas à déterminer une équivalence sémantique en l'absence de conceptualisation consensuelle sur le domaine⁶ (les termes peuvent avoir des valeurs différentes selon leurs occurrences en l'absence d'engagement sémantique voire ontologique, à la différence du support des GC). Il revient à l'expert de domaine, l'utilisateur du système, de valider ou non les ressemblances à l'issue de ce processus. Il peut ainsi construire un ensemble cohérent et pertinent relativement à son projet à partir des connaissances disparates, incomplètes et globalement incohérentes de la base (telles qu'issues de la lecture des documents d'origine en raison des hétérogénéités mentionnées en première partie). On assimile cette construction à la modélisation d'un problème spécifique, d'où la désignation de « système d'aide à la modélisation ».

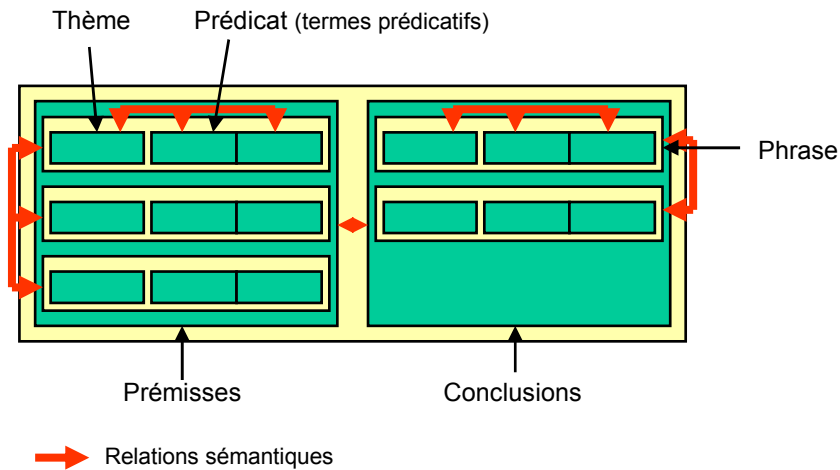


Figure 1. Les composants du granule. Les relations sémantiques sont de trois types : relations métiers entre terme thématique et termes prédicatifs, relations phrastiques (un sous-ensemble d'opérateurs booléens) entre phrases (un thème induit une phrase) et relation d'implication (augmentée d'une modalité) entre prémisses et conclusions. Chaque phrase supporte la négation. Cette figure n'inclut pas univers et modèle d'univers.

⁶ Ce qui correspond à la définition des ontologies selon Studer : une conceptualisation formelle explicite et partagée (Studer, 1998)

5. Prototype, premiers constats, perspectives

Ce modèle de représentation des connaissances a fait l'objet d'un développement sous la forme d'un prototype, dont les premiers essais ont amené à certains constats recensés ci-dessous.

Pour qu'un tel outil soit profitable à la profession, il doit être collaboratif ; le prototype issu du projet RAMCESH est donc accessible via un site Web moyennant une identification et l'obtention d'un droit d'accès auprès d'un modérateur.

L'ensemble du système étant amené à être implémenté et utilisé directement par les experts du domaine (pour dépasser le phénomène « boîte noire » constaté par certains experts, ainsi qu'indiqué par (Magnan, 1992)), il est très tôt apparu comme important de leur permettre d'utiliser des outils dont ils aient la maîtrise et qui leur soient familiers. C'est l'une des raisons qui ont mené au découpage du réseau lexical en « couches métiers » arborescentes, facilement visualisables via des outils aisément disponibles. Plus qu'une simple visualisation, il est apparu important de permettre aux experts de manipuler les données via ces outils. Le prototype permet en conséquence d'exporter et d'importer les données lexicales et associées (synonymes, définitions) dans des formats type tableur (.csv) ou graphiques (formats propriétaires) via un format xml.

Cela permet un confort d'utilisation qui est profitable aux experts (la possibilité de manipuler directement les arborescences à travers des interfaces graphiques est ainsi très appréciée), mais demande une gestion prudente des mises à jour et des versions ; le prototype permet de « geler » les accès en écriture aux parties ainsi exportées et en cours de modification. Tous les utilisateurs n'ont pas les droits pour cette manipulation, ce qui a nécessité la mise en place de droits d'accès divers régis à travers des profils donnant accès en modification aux diverses parties du formalisme (réseau lexical, granules). Ces profils permettent aussi de régler la partie validation de l'implémentation, fondamentale pour un travail de représentation des connaissances.

La notion de « couches métier » a également fait surgir la nécessité d'aligner les arborescences au sein de la base de connaissance globale, et de préserver les variations que l'on pourrait trouver au sein de celles-ci, ce qui a donné lieu à des travaux spécifiques (Ziani, 2007).

L'ensemble de ces mesures autorise une certaine séquentialisation du travail des experts relativement à l'implémentation de la base de connaissances. Très prosaïquement, il a été constaté que les experts participant au projet utilisaient leur temps de trajets professionnels pour effectuer ce travail ; c'est sans doute l'une des raisons majeures pour laquelle les possibilités d'import sont appréciées.

Certains éléments d'interface sont apparus également prépondérants : ainsi, l'interface, essentiellement fondée sur du glisser-déposer, évite les saisies manuelles

et les fautes que cela pourrait engendrer, ainsi qu'une frappe clavier souvent vécue comme une astreinte par des experts de terrain.

Parmi les possibilités d'amélioration du prototype en un outil pleinement fonctionnel, on trouve le traitement automatique des documents, fondé sur un extracteur terminologique statistique dont les résultats sont affinés par divers filtres. A l'heure actuelle, cet extracteur analyse chaque document dans son ensemble et propose un ensemble de candidats-termes pas toujours adapté au sujet traité. Une amélioration majeure serait d'effectuer un traitement thématique préalable des documents pour isoler les fragments de texte qui serviront de base aux granules de connaissance, et confiner l'analyse à ces fragments pour obtenir une liste de candidats-termes plus pertinente.

Un autre point majeur à améliorer est la constitution semi-automatique des granules, travail parfois fastidieux pour les experts, quoiqu'il soit facilement assimilé (quelques minutes de formation suffisent). Une automatisation accrue de la constitution des granules, qui proposerait les relations sémantiques et certains éléments d'univers en fonction d'une analyse documentaire plus poussée (la présence de certains termes induit certaines relations) est à l'étude, notamment en imposant des contraintes aux relations (une restriction à des sous-domaines, par exemple), mais l'équilibre délicat entre complexité croissante du système et facilité de prise en main serait alors bousculé.

6. Conclusion

A l'heure actuelle, le prototype a permis de bâtir un réseau lexical d'environ 6000 termes et une base de connaissances de 800 granules à partir d'un corpus d'une dizaine de documents. La possibilité de réaliser des projets complexes à partir de la base de connaissances devrait être testée sur un cas de projet de tunnel ferroviaire.

L'intervention de l'expert de domaine directement dans l'implémentation d'un tel système d'aide semble donc possible et satisfaisant, évitant au moins partiellement le goulot d'étranglement en matière de systèmes à base de connaissances relevé par certains auteurs⁷. A ce titre et à la lueur de l'expérience acquise pendant le projet, nous pensons que les éléments-clés d'une telle approche pour des sciences empiriques consiste en la réduction maximale de l'engagement nécessaire autour d'un consensus conceptuel, et une démarche générale visant à faciliter la compréhension du système par des experts peu au fait des problématiques d'ingénierie et de représentation des connaissances. Une réflexion sur la dualité *transfer process* / *modelling process* (décrits par Studer,

⁷ « Despite some successes, the lack of tools to allow subject matter experts to directly enter, query, and debug formal domain knowledge in a knowledge-base (KB) still remains a major obstacle to their deployment ». (Clark et al., 2001)

1998) et l'émergence d'une approche hybride serait à cet égard sans doute profitable aux communautés d'experts des domaines des sciences de la terre.

7. Bibliographie

- Clark et al., 2001 : Clark, Peter ; Thompson, John ; Barker, Ken ; Porter, Bruce ; Chaudhri, Vinay ; Rodriguez, Andres ; Thoméré, Jérôme ; Mishra, Sunil ; Gil, Yolanda ; Hayes, Pat ; Reichherzer, Thomas ; Knowledge entry as the graphical assembly of components ; Proc. of K-Cap 01, Canada, 2001
- Corblin, 2002 : Corblin, Francis ; Représentation du discours et sémantique formelle ; Presses Universitaires de France, Paris, 2002
- Faure, 2007a : Faure, Nicolas ; Modèles du signe et ontologies ; atelier Modélisation des Connaissances, EGC 2007, Namur, 2007
- Faure, 2007b : Faure Nicolas ; Un système d'aide à la modélisation des connaissances en géotechnique, thèse de doctorat, Université J.Moulin Lyon 3, 2007
- Kamp&Reyle, 1993 : Kamp, Jans ; Reyle, Uwe ; From discourse to logic ; Kluwer, 1993
- Magnan, 1992 : Magnan Jean-Pierre ; CESSOL : bilan du développement d'un système expert ; Actes du colloque Géotechnique et Informatique, ENPC, Paris, 1992
- Magnan, 2002 : Magnan, Jean-Pierre ; L'organisation du travail en géotechnique : normalisation, développement et artisanat ; Lettre de la géotechnique (26-27), Société Internationale de la Mécanique des Sols et de la Géotechnique, 2002 (<http://www.geotechnique.org>)
- Sowa, 1984 : Sowa, John F. ; Conceptual Structures: Information Processing in Mind and Machine ; Addison Wesley, 1984
- Studer et al., 1998 : Studer, R. ; Benjamins, V. ; Fensel, D. ; Knowledge engineering: Principles and methods. IEEE Transactions on Data and Knowledge Engineering, 1998
- Ziani, 2007 : Ziani Mina ; Construction d'une ontologie globale par fusion d'ontologies métiers. Application au projet RAMCESH, Mémoire de Recherche de Master, Université J. Moulin Lyon 3, 2007

Une architecture orientée services pour la gestion sémantique des données géologiques pour le stockage de CO₂

Nabil Belaid(*, **) — Idir Ait-Sadoune(*) — Yamine Ait-Ameur(*)
— Jean-François Rainaud(**) — Stéphane Jean(*)

(*) *LISI / ENSMA Téléport 2 - 1 avenue Clément Ader
BP 40109 86961 Futuroscope Chasseneuil cedex
{nabil.belaid, idir.aitsadoune, yamine, jean}@ensma.fr*

(**) *Institut Français du Pétrole - DTIMA, 1-4 Av. de Bois-Préau,
92852 Rueil-Malmaison,
j-francois.rainaud@ifp.fr*

RÉSUMÉ. Dans cet article, nous étudions deux problèmes de gestion des données en ingénierie pétrolière et stockage de CO₂: résoudre l'hétérogénéité sémantique des données causée par les multiples représentations produites par les disciplines de la géologie ; et accéder à l'information pertinente par une recherche sémantique dans les modèles géologiques et les documents de géologie. L'objectif de notre travail est ainsi double : d'une part, proposer un système de gestion de ces données basé sur une architecture orientée services (SOA) ; et, d'autre part, utiliser des ontologies pour résoudre l'hétérogénéité sémantique.

ABSTRACT. In this article, we focus on two problems in the CO₂ storage and in the petroleum engineering data management: resolve the semantic data heterogeneity caused by the multiple representation produced by the diverse geology disciplines; and access the relevant information through a semantic search in the geological models and in the geological documents. The aim of our work is twofold: make a proposal for a data management system based on a services oriented architecture; and use ontologies to resolve the in the data semantic heterogeneity.

MOTS-CLÉS : ingénierie pétrolière, ontologie, annotation sémantique, services web, base de données à base ontologique, architecture orientée services

KEYWORDS: petroleum engineering, ontology, semantic annotation, web services, ontology based data base, services oriented architecture

1. Introduction et problématique

Dans le domaine de l'ingénierie pétrolière, les géologues sont appelés à manipuler des données environnementales de natures et de formats divers. La représentation de ces données est souvent réalisée au travers de modèles spécifiques différents dits modèles métiers (stratigraphique, sismique, géologique structurale, lithologique, etc). Le traitement réservé à ces données et les différentes interprétations données par les géologues posent le problème de l'hétérogénéité de ces données et de leurs représentations.

Pour résoudre le problème de l'hétérogénéité, nous proposons d'organiser les données et les informations manipulées dans une architecture globale basée sur des ontologies de domaines et sur la recherche sémantique.

Dans leurs campagnes de collecte de données, les géologues se basent essentiellement sur des modèles métiers et sur des documents. Ces modèles et ces documents sont produits puis, c'est aux géologues que revient la tâche d'interprétation et d'annotation sémantique au travers de services web. Ces différents documents (publications, articles, rapports, graphiques etc.) conservent ces interprétations. Les géologues doivent les consulter afin de reconstituer les connaissances associées à une zone particulière. Nous proposons pour cela d'intégrer les différentes activités des géologues (annotation, stockage et recherche) dans une architecture orientée services (SOA) où chaque activité est réalisée par des services web. L'enchaînement et la composition de ces services permettent de définir différents workflows de la chaîne de modélisation.

Les différents travaux décrits dans cet article s'inscrivent dans le cadre du projet ANR *e-Wok Hub*¹ qui traite les problèmes liés au stockage du CO₂ par des approches fondées sur la modélisation à base ontologique². L'objectif de cet article est de donner une description globale des ontologies identifiées dans le domaine de l'ingénierie pétrolière (section 2) et de présenter le système de gestion des documents manipulés par les géologues (sections 3 et 4).

2. Les ontologies

Pour répondre aux différents besoins en représentation et en modélisation de connaissances, nous avons contribué, avec l'aide d'experts géologues, à la construction d'ontologies de domaine qui regroupent de nombreux concepts de la géologie.

Pour la construction de ces ontologies, nous avons compté sur diverses sources d'information. Un ensemble de modèles et de documents représentatifs a été choisi la

1. <http://www-sop.inria.fr/edelweiss/projects/ewok/>

2. Ce projet regroupe les partenaires suivants : LISI (Laboratoire d'Informatique Scientifique et Industrielle) de l'ENSMA (Ecole Nationale Supérieure de Mécanique et d'Aérotechnique), IFP (Institut Français du Pétrole), BRGM (Bureau de Recherches Géologiques et Minières), INRIA (Institut National de Recherche en Informatique et en Automatique), EADS (European Aeronautic Defence and Space) et le CRITT Informatique.

première fois par les experts dans le domaine de la géologie. Ils ont choisi manuellement le vocabulaire approprié contenu dans ces documents et relatifs aux aspects géographiques et géologiques du stockage de CO₂.

A partir de ces sources d'information et la participation des membres experts du projet eWokHub (BRGM, IFP, ENSMP), les ontologies de domaine suivantes ont été définies (Mastella *et al.*, 2008) :

- une ontologie *des concepts géographiques*³ qui repose sur le découpage administratif et sur la définition (polygonale) spatiale de secteurs. Elle définit les différents secteurs administratifs du territoire. On y trouve les cantons, les communes, les régions, les départements etc. ;
- une ontologie pour définir et gérer *les âges géologiques* pour la datation. Elle permet de définir l'échelle des temps géologiques ;
- une ontologie pour décrire *les bases de la géologie, les unités géologiques, les propriétés géologiques et les processus géologiques* (figure 1). Elle définit entre autres les horizons, les failles, etc.

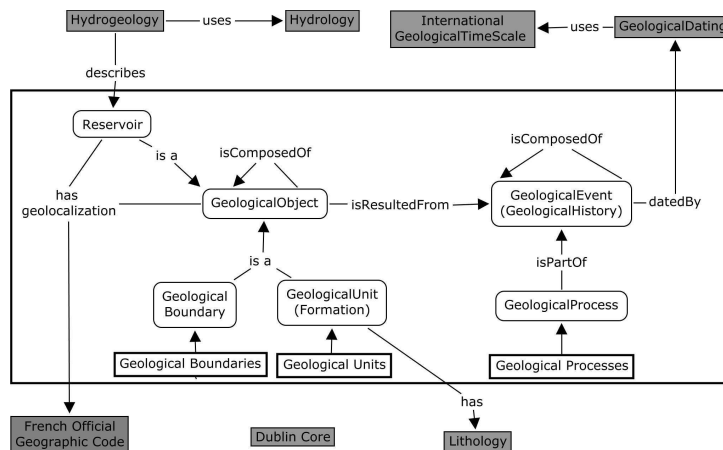


Figure 1. Une vue de l'ontologie des bases de la géologie, des unités géologiques, des propriétés géologiques et des processus géologiques.

Ces ontologies sont représentées dans notre système en RDF(S)/OWL (Manola *et al.*, 2004, Brickley *et al.*, 2004, McGuinness *et al.*, 2004).

3. Cette ontologie est basée sur le Code officiel géographique (COG) fourni par l'Institut National de la Statistique et des Études Économiques (INSEE) http://www.insee.fr/fr/nom_def_met/nomenclatures/cog/index.asp

Nous nous intéressons dans la suite de cet article aux services qui traitent les documents employés par les géologues.

3. Système de gestion de documents

Pour offrir une assistance aux géologues dans la phase de collecte, d'annotation, de sauvegarde et de recherche des documents relevant du domaine de l'ingénierie pétrolière, nous proposons un système de gestion de documents basé sur une architecture orientée services (SOA). Le système proposé permet d'analyser syntaxiquement les documents, de les annoter sémantiquement, de stocker ces annotations dans une base de données à base ontologique et enfin, d'effectuer une recherche sémantique sur ces documents.

Les annotations définies proviennent des différentes ontologies obtenues et évoquées en section 2. Nous proposons, dans le cadre de cet article, de détailler les différents services qui concernent l'annotation de la géolocalisation des documents, c'est-à-dire, qui permettent de situer géographiquement le contexte d'un document.

Le système proposé offre ainsi trois opérations principales :

- 1) annotations géographiques des documents puis sauvegarde de ces annotations ;
- 2) recherche sémantique des documents à partir d'une requête ontologique ;
- 3) affichage d'un document.

Avant de détailler les différentes opérations réalisées par les différents services, nous présentons, dans la figure 2, un modèle UML simplifié des données échangées entre les différents services. Ce modèle définit un format d'échange (RDF-XML) unifié pour les différents services mis en place. Dans ce système, nous manipulons des *ressources* qui peuvent être de diverses natures : *Annotation* (annotation géographique en RDF), *Query* (requête écrite en SPARQL) et *MediaUnit* (un document transformé en document XML conforme au format d'échange). Ce dernier constitue l'enveloppe de la ressource. Chaque *Ressource* est identifiée par une *URI* et peut être annotée par une ou plusieurs *Annotations*.

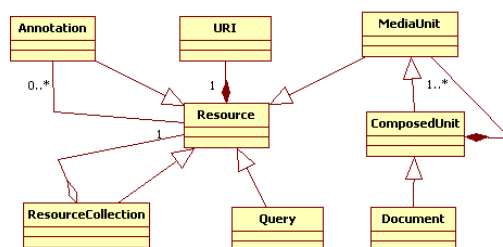


Figure 2. Modèle de données manipulées par les services

Décrivons à présent les trois opérations mises en place.

3.1. Annotations géographiques de documents et leur sauvegarde

Cette opération définit un traitement agissant sur les divers documents produits par des travaux en géologie. Ces documents, qui traitent de zones géologiques géolocalisées, sont rangés dans un répertoire identifié. Ce traitement permet de transformer ces documents dans le format d'échange défini sur la figure 2, d'annoter ces documents par des annotations géographiques et de leur attribuer une URI unique. Cet article ne décrit pas comment ces annotations sont produites. Des annotations sémantiques permettent de les produire de façon automatique sous forme de triplets de la forme (URI_document, géolocalisé_par, URI_onto_géolocalisation). Les annotations sont, par la suite, sauvegardées dans la base de données à base ontologique ONTODB (Dehainsala *et al.*, 2007).

Ce traitement consiste en l'enchaînement des services suivants⁴ (figure 3) :

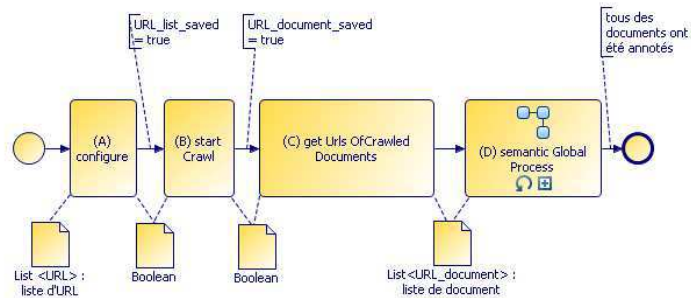


Figure 3. Orchestration des services d'annotation géographique de documents et de leur sauvegarde

- (A) *configure* : sauvegarde une liste d'URL de répertoires contenant des documents non annotés. Un booléen est renvoyé pour indiquer le bon déroulement du service.

- *configure*(List<URL>) : Boolean

- (B) *startCrawl* : parcourt la liste d'URL des répertoires sauvegardée par le service *configure* et génère une liste d'URL des documents trouvés dans ces répertoires. Un booléen est renvoyé pour indiquer le bon déroulement du service.

- *startCrawl*(Boolean) : Boolean

- (C) *getUrlsOfCrawledDocuments* : récupère la liste d'URL des documents sauvegardée par le service précédent.

4. Nous représentons par $f(x)$: y un service f qui prend en entrée une donnée de type x et renvoie une donnée de type y

- *getUrlsOfCrawledDocuments*(Boolean) :List<URL_document>

• (D)*semanticGlobalProcess* : est un service qui sera appelé autant de fois qu'il y a d'URL en sortie du service précédent. C'est un service composé. Il consiste en l'enchaînement des services suivants (figure 4).

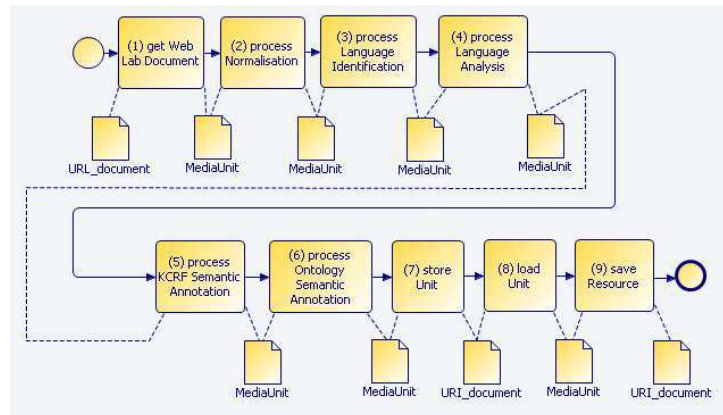


Figure 4. Orchestration des services composant le service "semanticGlobalProcess"

• (1)*getWebLabDocument* : construit une ressource correspondant au format d'échange de type MediaUnit qu'il annote avec l'URI physique du document natif.

- *getWebLabDocument*(URL_document) :MediaUnit

• (2)*processNormalisation* : extrait le contenu textuel du document natif puis l'insère dans la MediaUnit renvoyée par le service précédent.

- *processNormalisation*(MediaUnit) :MediaUnit

• (3)*processLanguageIdentification* : annote la MediaUnit renvoyée par le service précédent avec la langue du document natif.

- *processLanguageIdentification*(MediaUnit) :MediaUnit

• (4)*processLinguisticAnalysis* : analyse syntaxiquement la MediaUnit renvoyée par le service précédent et annote ses termes avec des tags (nom, verbe, adjectif .etc). Un exemple de tag pourrait être "Poitiers est une ville".

- *processLinguisticAnalysis*(MediaUnit) :MediaUnit

• (5)*processKCRFSemanticAnnotation* et (6)*processOntologySemanticAnnotation* : permettent d'annoter la MediaUnit renvoyée par le service précédent par des annotations de localisation géographique. Ils se basent sur des statistiques portant sur les constructions syntaxiques fournies par le service précédent et sur l'ontologie des concepts géographiques. Un exemple d'annotations pourrait être "Poitiers a comme code commune (86000)".

- *processKCRFSemanticAnnotation*(MediaUnit) :MediaUnit

- *processOntologySementicAnnotation(MediaUnit) :MediaUnit*
- (7)*storeUnit* : sauvegarde la MediaUnit renvoyée par le service précédent avec une URI unique.
 - *storeUnit(MediaUnit) :URI_document*
- (8)*loadUnit* : charge la MediaUnit correspondante à une URI renvoyée par le service précédent.
 - *loadUnit(URI_document) :MediaUnit*
- (9)*saveResource* : sauvegarde dans une base de données à base ontologique les annotations sur la MediaUnit renvoyée par le service précédent.
 - *saveResource(MediaUnit) :URI_document*

3.2. Recherche sémantique de documents à partir d'une requête

Cette opération permet de retrouver les URI de documents préalablement annotés et sauvegardés dans la base de données par l'opération décrite dans la section 3.1. Elle est réalisée par un service nommé *search* détaillé dans la suite de cet article. Ce service prend en entrée une requête sémantique exprimée en SPARQL (Prud'hommeaux *et al.*, 2008) dans laquelle un paramètre de localisation géographique doit être spécifié et renvoie en sortie une liste d'URI de documents.

– *search(Query) :List<URI_document>*

Notre choix a porté sur le langage de requête SPARQL car nous avons estimé qu'il était assez général pour être interprété aussi bien sur les moteurs sémantiques comme CORESE (Corby *et al.*, 2004) que sur la base de données à base ontologique ONTODB (Dehainsala *et al.*, 2007). Cela rend les accès à ces outils complètement transparents.

3.3. Affichage d'un document

Cette opération permet d'afficher une MediaUnit préalablement annotée et sauvegardée à partir de son URI. Le service qui traite cette opération est *loadUnit*.

– *loadUnit(URI_document) :MediaUnit*

4. Exemple d'implémentation de services

Nous avons choisi de montrer comme exemple d'implémentation des différents services de notre système, le service de recherche sémantique *search*. Ce dernier permet de retrouver une liste d'URI de documents préalablement annotés et sauvegardés dans une base de données à base ontologique. Il prend en entrée une requête séman-

tique exprimée en SPARQL et renvoie en sortie une liste d'URI de ressources de type document.

En termes d'implémentation, les échanges entre services se font par messages. Sur le tableau 1, nous montrons la partie du code WSDL(Booth *et al.*, 2007) qui définit les messages manipulés et la signature du service *search*. Ce dernier reçoit un message de type *query* et renvoie un message de type *resourceCollection*.

```

<message name="searchIn">
  <part name="request" element="types :query"/>
</message>

<message name="searchOut">
  <part name="response" element="types :resourceCollection"/>
</message>

<portType name="Search">
  <operation name="search">
    <input message="searchIn">
      <output message="searchOut">
    </operation>
  </portType>

```

Tableau 1. Code WSDL décrivant le service *search*

Le code WSDL du tableau 1 est implémenté par la méthode JAVA décrite sur le tableau 2. Le service récupère la requête écrite en SPARQL et fait appel à une API permettant d'interroger une base de données à base ontologique OntoDB(Dehainsala *et al.*, 2007) par le langage de requête OntoQL(Jean *et al.*, 2006).

```

public SearchOut search(SearchIn request) throws exception{
  String querySPARQL = request.getQuery().getData();

  String url = "jdbc :postgresql ://localhost :5432/eWokHub";
  String user = "postgres";
  String pwd = "root";

  Class.forName("org.postgresql.Driver");
  Connection connection = DriverManager.getConnection(url, user, pwd);
  OntoQLSession session = new OntoQLSessionImpl(connection);
  session.setReferenceLanguage(ModelUtil.ENGLISH);

  OntoQLResultSet resultSet = statement.executeSPARQLQuery(querySPARQL);
  while(resultSet.next()) {
    String resultQuery = resultSet.getString(1);
    Document doc = new Document();
    doc.setUri(resultQuery);
    result.getResources().add(doc);
  }

  ResourceCollection result = new ResourceCollection();
  SearchOut resultOut = new SearchOut();
  resultOut.setResults(result);
  return resultOut;
}

```

Tableau 2. Code Java décrivant le service *search*

Sur le tableau 3 un exemple de requête SPARQL reçue par le service de recherche sémantique. Cette requête consiste à rechercher les documents (*rdf:type geo :document*) annotés par le prédicat de localisation géographique (*geo :geolocalized*) avec une zone géologique (*rdf:type geo :zone_geologique*) qui a comme valeur la région Poitou-Charentes (*<http://rdf.insee.fr/geo/REG_54>*).

```

<query>
<data>
PREFIX geo : <http://rdf.insee.fr/geo/>
SELECT ?doc
WHERE {
  {?doc rdf:type geo :document .
  ?doc geo :geolocalized?commune .
  ?region rdf:type geo :zone_geologique .
  ?region geo :subdivision?commune .
  FILTER (?region = <http://rdf.insee.fr/geo/REG_54>)}
UNION
  {?doc rdf:type geo :document .
  ?doc geo :geolocalized <http://rdf.insee.fr/geo/REG_54>
  }}
</data>
</query>

```

Tableau 3. Exemple de requête SPARQL reçue par le service search

Le résultat de la requête du tableau 3 est un ensemble d'URI de documents annotés par la région de la Poitou-Charentes et les communes contenues dans cette région (Jaunay Clan et Chasseneuil Du Poitou) (tableau 4).

```

<results>
<resources uri="http://rdf.insee.fr/doc/ChasseneuilDuPoitou" xsi:type="ns3:document"/>
<resources uri="http://rdf.insee.fr/doc/JaunayClan" xsi:type="ns3:document"/>
<resources uri="http://rdf.insee.fr/doc/PoitouCharentes" xsi:type="ns3:document"/>
</results>

```

Tableau 4. Exemple de résultats renvoyée par le service search

5. Conclusion et perspective

Dans cet article, nous avons traité le problème lié à l'hétérogénéité des données manipulées dans le domaine de l'ingénierie pétrolière et de leurs représentations. La solution proposée par le projet *e-Wok Hub* est basée sur des ontologies de domaines pour annoter les différentes informations et données manipulées par les géologues.

Le second problème traité est lié aux activités d'annotation, de stockage et de recherche des documents collectés. Nous avons proposé d'intégrer ces différentes activités dans une architecture orientée services (SOA) basée sur les services web.

Dans l'architecture proposée, la définition de l'enchaînement des activités et leur orchestration pour réaliser la tâche désirée, nécessite une connaissance a priori des signatures des services et de leurs rôles. Comme perspective à ce travail, nous proposons

d'étudier les problèmes liés à l'interopérabilité des activités et à leur enchaînement ainsi que la validation a priori des compositions de services. Nous voulons également automatiser le processus de découverte sémantique des services en intégrant dans des annuaires leurs descriptions sémantiques.

6. Bibliographie

- Booth D., Liu C.-K., Web Services Description Language, Technical report, W3C Recommendation, 2007. <http://www.w3.org/TR/2007/REC-wsdl20-primer-20070626/>.
- Brickley D., Guha R., McBride B., RDF Vocabulary Description Language 1.0 : RDF Schema, Technical report, W3C Recommendation, 2004. <http://www.w3.org/TR/2004/REC-rdf-schema-20040210/>.
- Corby O., Dieng-Kuntz R., Faron-Zucker C., Querying the semantic web with the corese search engine, Technical report, Valencia, 2004.
- Dehainsala H., Pierra G., Bellatreche L., « OntoDB : An Ontology-Based Database for Data Intensive Applications », *of the 12th Int. Conf. on Database Systems for Advanced Applications (DASFAA'07)*, LNCS. Springer, p. 497-508, 2007.
- Jean S., Ait-Ameur Y., Pierra G., « Querying Ontology Based Database Using OntoQL (an Ontology Query Language) », *Ontologies, Databases, and Applications of Semantics (ODBASE'06)*, p. 704-721, 2006.
- Manola F., Miller E., McBride B., RDF primer, Technical report, W3C Recommendation, 2004. <http://www.w3.org/TR/2004/REC-rdf-primer-20040210/>.
- Mastella L., Perrin M., Ait-Ameur Y., Abel M., Rainaud J.-F., « Formalising geological knowledge through ontologies and semantic annotation », *extended abstract accepted at the 70th EAGE Conference and Exhibition*, 9-12 June, 2008.
- McGuinness D., Harmelen F., OWL Web Ontology Language, Technical report, W3C Recommendation, 2004. <http://www.w3.org/TR/2004/REC-owl-features-20040210/>.
- Prud'hommeaux E., Seaborne A., sparql query language for rdf, Technical report, W3C Recommendation, 2008. <http://www.w3.org/TR/rdf-sparql-query/>.

Système pour l'acquisition et la gestion de données environnementales

Gil De Sousa* — **Hai-Ying Zhou**** — **Kun-Mean Hou ***** —
Christophe de Vault*** — **Jean-Pierre Chanet***

*Cemagref, UR TSCF/UMR TETIS, France

gil.de-sousa@cemagref.fr

jean-pierre.chanet@cemagref.fr

**Harbin Institute of Technology, Chine

haiyingzhou@hit.edu.cn

***LIMOS UMR 6158 CNRS, Université Blaise Pascal, Clermont-F^d II, France

kun-mean.hou@isima.fr

christophe.devault@isima.fr

RÉSUMÉ. L'acquisition de données environnementales fait partie des nombreux domaines d'applications des réseaux de capteurs sans fil (RCSF). Dans cet article, un système complet pour l'acquisition et la gestion de ce type de données est présenté. Il comprend un capteur sans fil, un noyau temps-réel et un micro-système de fichiers. Chaque élément apporte des innovations par rapport aux solutions existantes comme le système LiveFile dédié à la fois à la gestion de la mémoire Flash et à l'interrogation simple de données.

ABSTRACT. Environmental data collection is part of the various applications of Wireless Sensor Networks (WSN). In this article, a complete system dedicated to collect and to manage this kind of data is presented. It is based on a wireless sensor, a real-time kernel and a micro-file system. Each element brings innovative concepts compared to existing solutions such as the LiveFile system dedicated both to Flash memory management and simple data query.

MOTS-CLÉS : acquisition de données environnementales, réseaux de capteurs sans fil, noyau temps-réel, micro-système de fichiers interrogatif

KEYWORDS: environmental data collection, wireless sensor networks, real-time kernel, interrogative micro-file system

1. Introduction

Les réseaux de capteurs sans fil (RCSF) sont une thématique de recherche récente. De plus en plus d'applications font ou envisagent de faire appel à eux soit pour acquérir des données soit pour substituer une infrastructure de réseau fixe inexistante. Les domaines d'application de ce type de technologie sont, entre autres, la santé et le bien-être des personnes (« smart care, smart home »), l'intervention des secours, les déploiements militaires et la surveillance ou l'étude de l'environnement (Akyildiz *et al.*, 2002).

Les problématiques soulevées par les RCSF sont diverses et variées. Certaines sont nouvelles comme le développement de protocole de routage pour les réseaux Ad Hoc. D'autres sont issues de problématiques couramment rencontrées qui ont été résolues pour des cas donnés mais différents des RCSF. L'élaboration et le fonctionnement des systèmes d'exploitation embarqués doivent être ainsi repensés pour s'adapter aux contraintes associées aux RCSF. Les ressources limitées en mémoire, en énergie et en puissance de calcul présentes au sein de chaque capteur représentent la principale contrainte. Le processus d'acquisition d'une donnée peut être découpé en quatre étapes :

- l'acquisition des données,
- le stockage des données,
- le traitement et l'agrégation des données,
- la transmission des données ;

L'étape d'agrégation est située soit avant le stockage soit avant la transmission. Chaque ressource est associée principalement à une étape du processus d'acquisition. La mémoire l'est au stockage, l'énergie à la transmission et la puissance de calcul à l'agrégation.

Un dispositif utilisant les RCSF est proposé pour l'acquisition et la gestion de données environnementales. Il comprend un capteur sans fil LiveNode, un système d'exploitation LIMOS et un microsystème de fichiers muni de fonctionnalités pour l'interrogation de données LiveFile. Ces deux derniers éléments feront l'objet d'une présentation détaillée respectivement dans les sections 2 et 3. La section 4 présentera l'utilisation de l'ensemble du dispositif. La section 5 viendra conclure l'article.

2. Systèmes d'exploitation pour RCSF

La recherche dans la thématique des RCSF amène à repenser et à adapter des solutions élaborées auparavant à d'autres types d'utilisation. Les systèmes d'exploitation embarqués sont un domaine où de nombreux travaux ont été réalisés et ont abouti à des solutions complètes et répondant à un ensemble de besoins. Pour gérer les capteurs sans fil, l'emploi de systèmes d'exploitation existants a d'abord été envisagé. Cependant, leur empreinte mémoire et leur fonctionnement général

s'accordant mal avec les ressources limitées des capteurs, de nouveaux systèmes ont été développés (Stojmenovic 2005). Ils sont répartis en deux grandes catégories :

- les systèmes multitâches,
- les systèmes basés sur les événements ;

Les premiers sont une adaptation des systèmes multitâches au RCSF. Le système MANTIS (Multimodal system for Networks of In-situ wireless Sensor) qui appartient à cette catégorie, utilise une politique d'ordonnancement avec gestion de priorités et partage de temps répandue dans les systèmes temps-réel embarqués.

Contrairement à un processus, un événement ne peut pas être préempté c'est-à-dire être remplacé au niveau du processeur, pendant un instant, par un autre de priorité plus élevé. Le mécanisme de changement de contexte n'est donc pas nécessaire et tous les événements partagent une même et unique pile d'exécution. Ce mode de fonctionnement fait des systèmes basés sur les événements une solution intéressante pour les RCSF. Le noyau TinyOS est le plus connu de ce type de système. Les problèmes de blocage au sein d'un événement sont l'inconvénient majeur de ces systèmes.

Un nouveau noyau nommé LIMOS (Lightweight Multithreading Operating System) a été développé en prenant en compte les avantages et les inconvénients des deux catégories de systèmes qui viennent d'être présentées (Zhou *et al.* 2006a). La principale caractéristique du système LIMOS est son architecture hybride à la fois multitâches et basées sur les événements. Trois modes de fonctionnement sont donc possibles : multitâches, basé sur les événements ou hybride. Ces différentes configurations sont illustrées respectivement par les figures 1, 2 et 3. Quel que soit le mode choisi, le conteneur principal est l'événement. Une configuration où chaque événement a un seul processus correspond à un système basé sur les événements. A l'inverse, la présence d'un seul événement avec plusieurs processus représente un système multitâche.

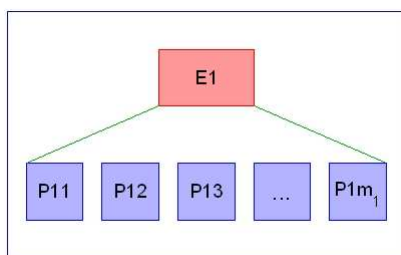


Figure 1. Configuration multitâches

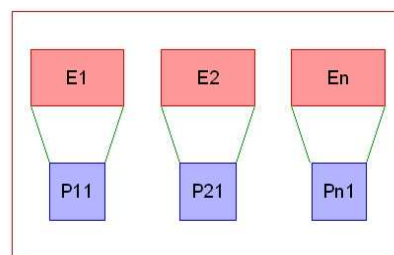


Figure 2. Configuration basée sur les événements

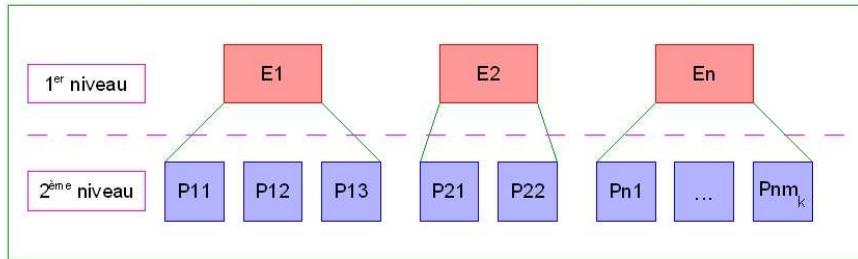


Figure 3. Configuration hybride

La politique d'ordonnancement est gérée à 2 niveaux. Le gestionnaire d'événements sélectionne l'événement courant suivant différents critères tels que la priorité ou le nombre de messages à traiter en attente. Au sein de chaque événement, un ordonnanceur intervient pour gérer l'ordre d'exécution des processus.

3. Gestion des données dans les RCSF

3.1. Gestion de la mémoire

Dans les applications de RCSF surtout celles d'acquisition de données, le stockage d'informations au sein de chaque capteur est parfois requis. Cette fonctionnalité implique de plus en plus l'utilisation de la mémoire Flash non volatile. La gestion de ce type de mémoire est complexe et a nécessité le développement d'un nouveau module dédié et intégré au système d'exploitation LIMOS : le microsystème de fichiers LiveFILE (LIMOS Versatile Embedded File system). Au-delà des traitements associés à la mémoire Flash, ce module comprend également des fonctionnalités pour l'interrogation des données. Par conséquent, le système LiveFile réunit, dans une optique de gestion intelligente des ressources, les gestionnaires de mémoire Flash (Dai *et al.* 2004) et d'interrogations de données (Bonnet *et al.* 2001) (Madden *et al.* 2005).

Le système LiveFile a été conçu de manière à respecter le plus possible les principes suivants :

- différents formats pour le stockage des données,
- gestion intelligente de la mémoire Flash,
- niveau d'abstraction pour l'accès aux données ;

Généralement, dans les applications d'acquisition de données environnementales, le ou les types de données manipulées sont connus à l'avance avant le déploiement in-situ. Une pré-configuration du système LiveFile est ainsi possible en spécifiant la taille et le format des données ainsi que les traitements associés. De plus, les données sont réparties en trois catégories : « RECORD »,

« FILE », « CHECKPOINT ». Ces catégories correspondent respectivement à des enregistrements simples, des données regroupées en fichier et des informations systèmes pour la gestion mémoire.

Différents principes ont été définis pour utiliser de manière optimale la mémoire Flash disponible dans un capteur sans fil. Chaque page de mémoire Flash utilisée est munie d'un en-tête contenant des informations nécessaire à sa gestion. La principale est le nombre d'opérations d'écritures subies par celle-ci. En effet, l'une des particularités de la mémoire Flash est le nombre restreint d'écritures que chaque page ou bloc de pages peut recevoir. Ce nombre varie entre 10.000 et 100.000 programmations. Des mécanismes ont donc été implémentés pour augmenter la durée de vie de ces pages. Le premier consiste à n'écrire dans la mémoire Flash que par bloc de données dont la taille est proche de celle d'une page plutôt que d'écrire informations par informations. Comme le montre la figure 4, un buffer de mémoire RAM dit de pré-écriture est utilisé pour stocker les données temporairement. Les données de ce buffer sont ensuite copiées dans la mémoire Flash soit quand il est plein soit à la suite d'une demande effectuée par le système d'exploitation. Le second mécanisme vise à préserver les pages qui auraient été trop utilisées. Un seuil d'usure est défini et correspond à un nombre d'écritures à partir duquel une page doit être économisée.

Dans le noyau LIMOS, le concept LINDA (Rowstron 1996) est utilisé, par l'intermédiaire des primitives In() et Out(), pour assurer à la fois la communication et la synchronisation entre événements ou processus et la gestion des périphériques. Dans le cadre de l'intégration du système LiveFile dans le noyau LIMOS, ce concept a été étendu pour offrir un niveau d'abstraction pour la gestion des données stockées dans la mémoire Flash. Désormais, suite aux différentes modifications que nous avons réalisées, l'écriture de la mémoire Flash passe par la primitive Out() et la lecture des données insérées par la primitive In().

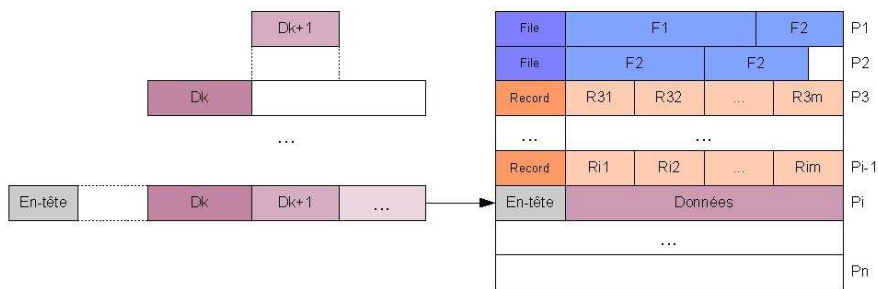


Figure 4. Programmation Différée de la mémoire Flash

3.2. Interrogation des données stockées

L'interrogation de données au sein des systèmes de gestion de bases de données est, la plupart du temps, réalisée à l'aide de requêtes élaborées en langage SQL. Par conséquent, la primitive In() a été, une nouvelle fois, modifiée de manière à pouvoir interpréter des requêtes utilisant ce langage. Généralement, dans une application environnementale, un seul type d'entité peut être considéré pour stocker l'ensemble des données collectées. Partant de cette hypothèse, les opérateurs ensemblistes n'ont pas d'utilité. De plus, la présence d'une seule relation limite mais n'interdit pas l'intérêt des opérateurs de « jointure » et de « division ». Par conséquent, de manière standard, sont présents les opérateurs de « projection » et de « sélection ». Ils peuvent être complétés en cas de nécessité par des fonctions mathématiques pour effectuer, par exemple, la moyenne ou la somme des données et de la clause ORDER BY. Cette dernière est obtenue par l'intermédiaire d'un algorithme de tri.

Pour accélérer l'interrogation de données, la notion de propriétés a été introduite. Son principe est l'indexation de données à partir de critères préétablis. La définition de ces critères est de la responsabilité des experts du domaine étudié. Dans une application environnementale, différents indicateurs obtenus à partir des grandeurs observées peuvent être associés à un phénomène clairement identifié. Si l'on prend l'exemple d'une application de surveillance d'incendie, différents capteurs physiques dédiés à l'observation de la température, de la fumée et de la luminosité peuvent être embarqués au sein d'un même capteur sans fil. Des modèles peuvent être définis pour déterminer suivant les valeurs de ces grandeurs le niveau de risque d'incendie associé. La recherche des moments où le niveau de risque était le plus élevé est ainsi accéléré. Pour chaque donnée collectée, le respect ou non d'une propriété définie se résume au stockage d'un bit à valeur 0 ou 1 dans une structure prévue à cet effet de type soit tableau ou soit liste chaînée.

4. Applications

Un capteur sans fil nommé LiveNode (LIMOS versatile embedded wireless sensor Node) a été développé à partir d'un microcontrôleur ARM7 AT91SAM7S256. Pour le stockage et la gestion des données, il dispose de 64Ko de mémoire RAM et 256Ko de mémoire Flash. Si les besoins de l'application le requièrent, plusieurs capteurs LiveNode peuvent être associés dans une approche dite multi-composants pour fédérer leurs ressources. Ce capteur sert de support au noyau LIMOS et au système LiveFile.

Une première version de l'ensemble du dispositif a été évaluée dans le cadre du projet MobiPlus dont l'objectif est l'amélioration de l'accessibilité aux transports publics pour les personnes handicapées (Zhou *et al.* 2006b). Une infrastructure placée au niveau d'un abribus transmettait des données sur la présence

ou non de personnes handicapées. Grâce à ces informations, le conducteur pouvait décider de la manœuvre à effectuer pour faciliter l'accès au bus.

Actuellement, des travaux portent sur l'adaptation de ce dispositif à différentes applications environnementales. La première vise à étendre la zone de couverture d'un réseau Wi-Fi et à le rendre interopérable avec un RCSF permettant ainsi un management agro-environnemental des exploitations agricoles. Suivant l'état du réseau, les capteurs ont la possibilité de stocker les données et de différer leur transmission. La seconde concerne la surveillance de grandeurs environnementales pour la prise de décision. Elle s'inscrit dans une démarche complète partant de la phase d'acquisition à celle de l'interprétation.

5. Conclusion et perspectives

L'acquisition de données environnementales fait partie des applications les plus répandues pour les RCSF. L'ensemble du dispositif composé du capteur sans fil LiveNode, du noyau LIMOS et du système LiveFile étant dédié au RCSF, son utilisation pour ce type d'applications paraît naturelle. Le capteur LiveNode par son approche dite multi-composants, le noyau et son architecture hybride, le système LiveFile combinant gestion de mémoire et de données apporte des innovations vis-à-vis de l'existant.

Des fonctionnalités doivent être ajoutées au système LiveFile pour pouvoir le considérer comme un système de gestion de bases données dédiés au RCSF. Par exemple, la possibilité de gérer plusieurs relations en fait partie et demande une profonde réflexion sur l'utilisation des ressources du capteur pour atteindre cet objectif. Une interface pour configurer l'ensemble du dispositif est également un domaine à étudier car elle faciliterait et accélérerait le déploiement du réseau de LiveNode.

6. Bibliographie

- Akyildiz I. F., Su W., Sankarasubramaniam Y., Cayirci E., « A Survey on Sensor Networks », *IEEE communications Magazine*, vol. 40, n° 8, August 2002, pp. 102-114.
- Bonnet P., Gehrke J., Seshadri P., « Towards Sensor Database Systems », *2nd International Conference on Mobile Data Management*, Hong Kong, January 2001.
- Dai H, Neufeld M., Han R., « ELF: An Efficient Log-Structured Flash File System For micro Sensor Nodes », *2nd ACM Conference on Embedded Networked Sensor System (SenSys)*, Baltimore, MD, USA, November 3-5, 2004, pp. 176-187.
- Madden S., Franklin M. J., Hellerstein J. M., Hong W., « TinyDB: an Acquisitional Query Processing System for Sensor Networks », *ACM Transactions on Database Systems (TODS)*, vol. 30, n° 1, March 2005, pp 122-173.
- Rowstron A. I. T., Bulk primitives in LINDA run-time systems, PhD Thesis, University of York, UK, October 1996.

Stojmenovic I., *Handbook of Sensor Networks, Algorithms and Architectures*, Wiley Series on Parallel and Distributed Computing, 2005.

Zhou H.-Y., Hou K.-M., Chanet J.-P., de Vault C., De Sousa G., « LIMOS: a Tiny Real-Time Micro-Kernel for Wireless Objects », *2nd IEEE International Conference on Wireless Communications, Networking and Mobile Computing WiCOM 2006 (WCNM 2006)*, Wuhan, China, September 22-24, 2006a.

Zhou H.-Y., De Sousa G., Chanet J.-P., Hou K.-M., Li J.-J., de Vault C., Kara M., « An Intelligent Wireless Bus-Station System Dedicated to Disabled, Wheelchair and Blind Passengers », *IET International Conference on Wireless, Mobile & Multimedia Networks ICWMMN 2006*, Hangzhou, China, November 6-9, 2006b.

Mai 2008