
Mise en œuvre d'une mémoire environnementale adaptée aux besoins d'un observatoire du domaine côtier : MEnIr

Cyril Faucher* — Françoise Gourmelon* — Jean-Yves Lafaye**
— Mathias Rouan*

* Laboratoire Géomer (UMR 6554 CNRS LETG), Institut Universitaire Européen de la Mer-OSU (Université de Bretagne Occidentale), Technopôle Brest-Iroise, 29280 Plouzané cedex
<http://letg.univ-nantes.fr>

cyril.faucher@gmail.com, {francoise.gourmelon, mathias.rouan}@univ-brest.fr

** L3i, Université de La Rochelle, avenue Michel Crépeau, 17042 La Rochelle cedex1
<http://www.univ-lr.fr/labo/l3i/>

jylafaye@univ-lr.fr

RÉSUMÉ. Les acteurs de la recherche utilisent pour leurs travaux de multiples sources d'information, qu'elles proviennent de données d'observation ou de références bibliographiques. L'archivage, le catalogage et la description de toutes ces ressources sont les seuls moyens de les pérenniser, de les échanger et d'en garantir une bonne utilisation. Nous présentons ici une démarche de conception d'un système d'information qui répond à ces besoins en s'appuyant sur les concepts d'une "mémoire environnementale" et sur les normes ISO (série 19100). Afin de fournir un outil respectant l'hétérogénéité naturelle de toutes ces informations et de leurs relations nous avons d'abord construit des modèles UML généralistes, que nous particularisons dans une seconde étape. Ces efforts de modélisation sont concrétisés par une phase d'implémentation.

ABSTRACT. Research workers are generally using many various pieces of information in their investigations, e.g. bibliography or experimental observation databases. Archiving, cataloguing, indexing and describing such information resources are the only means to ensure their durability and efficient reuse. In the following, we propose a way to design an information system that is backed up on both the "environmental memory" concept and on ISO standards (collection ISO 19100). We have designed convenient tools for dealing with information complexity, and present here below with pieces of a UML specification that are the basis of the system implementation.

MOTS-CLÉS : archivage, indexation, base de connaissances, ISO 19115, observatoire, systèmes d'information.

KEYWORDS: archiving, indexation, knowledge-based system, ISO 19115, observatory, information systems.

1. Introduction

La plupart des recherches en environnement ont comme objectif principal d'étudier le fonctionnement et l'évolution de systèmes complexes impliquant la prise en compte de multiples variables et de leurs relations. Cette complexité est accrue dans les zones d'interfaces telles que le littoral (Cuq, 2000) où les données recueillies sont très hétérogènes. Dans bien des cas, la réalisation de cet objectif scientifique sur le long terme implique la mise à disposition de séries de données d'observation acquises dans des conditions normalisées. Ce besoin converge avec celui exprimé par l'ensemble de la communauté scientifique relevant des Sciences de l'Environnement qui s'est traduit depuis plusieurs années par la mise en place d'observatoires. On peut citer les 24 structures établies aux Etats-Unis par le programme Long Term Environmental Research (LTER) et, en France, les Observatoires des Sciences de l'Univers (OSU) et les Observatoires de Recherche en Environnement (ORE). Le concept de zones ateliers¹, développé par le Programme Environnement Vie et Sociétés du CNRS² (PEVS) inclut également cette notion de recueil de données d'observation pluridisciplinaire sur le long terme, élargie à la collecte des connaissances dans une perspective de pérennisation, comme en témoigne les réflexions menées au sein du comité scientifique MOTIVE (Modélisation, Transfert d'Informations, Valorisation de l'Environnement) du PEVS. Elles s'appuient sur la notion de « mémoire environnementale », décrite comme la somme des représentations explicites, persistantes et structurées des données, des connaissances, des modèles et des savoir-faire scientifiques attachés à un système d'observation à long terme de l'environnement en vue d'en faciliter l'accès, le partage et la réutilisation (Guarnieri *et al.*, 2003).

Sur la base de ce concept, le projet MEnIr (Mémoire Environnementale du système d'observation à long terme du domaine côtier de la mer d'Iroise) a pour objectif de réaliser un système d'information³ utilisable par un large public (scientifiques, gestionnaires, décideurs), de manière à favoriser une démarche pluridisciplinaire et concertée au sein de la communauté concernée par la gestion intégrée des zones côtières (GIZC) (Cicin-Sain *et al.*, 1998 ; Gourmelon *et al.*, 2005(a)). Les technologies de l'information et de la communication offrent des apports non négligeables au partage et au transfert des connaissances. Il est indiscutable que l'ampleur et les objectifs de la GIZC et des observatoires de

1. Les zones ateliers, d'échelle régionale, sont définies en fonction de questionnements scientifiques impliquant des recherches à long terme sur les anthroposystèmes, et émanant d'une dynamique scientifique ou d'une demande sociale reformulée en termes scientifiques. Elles disposent de moyens techniques d'organisation et de traitement des données ainsi que de centres de compétences nécessaires. Les sites ateliers, d'échelle locale, englobés dans les zones ateliers, font l'objet d'observations et d'expérimentations.

2. Centre National de la Recherche Scientifique : <http://www.cnrs.fr/>

3. Un système d'information est défini comme un ensemble organisé d'éléments qui permet de regrouper, d'indexer et de diffuser de l'information sur un phénomène donné.

recherche en environnement justifient pleinement la mise en œuvre de systèmes d'information opérationnels garantissant d'une part le bon déroulement des programmes, et d'autre part la mise à disposition sur le long terme de données et de compétences (Dwyer *et al.*, 2004).

L'Observatoire du Domaine Côtier (ODC), développé à l'Institut Universitaire Européen de la Mer (IUEM-OSU) de l'Université de Bretagne Occidentale, a pour objectif de recueillir des séries d'observation acquises sur le long terme concernant globalement l'environnement côtier de la mer d'Iroise (entre Manche et Atlantique) dans la perspective de fournir aux scientifiques les données nécessaires à l'interprétation de son évolution sous l'influence conjuguée du climat et des activités humaines. Qualité des eaux côtières, estimation de la biodiversité et suivi de la morphologie des plages sont les thèmes de séries d'observation acquises à des pas de temps différents (de 20 minutes jusqu'à 6 mois), stockées dans un système d'archivage et de catalogage : le Système d'Information pour l'Environnement Côtier (SIEC) consultable *via* le Web (Gourmelon *et al.*, 2004). Le SIEC⁴ utilise le logiciel REPORTS V2 conforme à la norme européenne PR ENV 12657. Accessible via internet, il permet la consultation du catalogue (métadonnées) et le téléchargement de certaines informations en accès libre. On peut citer d'autres initiatives similaires dans des observatoires, à titre d'exemple le programme ROSELT (Réseau d'Observatoires de Surveillance Ecologique à Long Terme) mis en œuvre en Afrique par l'IRD qui utilise l'outil MDWeb⁵ (Desconnets *et al.*, 2007) pour cataloguer les données accessibles via un portail de catalogues institutionnels (Roselt/OSS, 2004). Dans d'autres contextes, de multiples systèmes sont actuellement développés par les organismes producteurs d'informations géographiques, par des collectivités ou des services de l'Etat (Robin *et al.*, 2005). Ils ont pour dénominateur commun de proposer le catalogage et l'archivage des données géoréférencées, d'utiliser la norme ISO 19115 et de fonctionner avec une architecture client/serveur web. Si leur intérêt est réel en termes d'archivage, il est néanmoins limité pour la diffusion des connaissances environnementales. Cette insuffisance est en partie liée à l'absence d'éléments relatifs aux sources, à l'historique, à la généalogie : c'est-à-dire à l'« environnement » de l'information (ou traçabilité).

Le projet MEnIr a été mis en œuvre dans le cadre de l'ODC par le laboratoire Géomer (UMR 6554 CNRS) et a fait l'objet d'un partenariat avec le laboratoire d'informatique (L3I) de La Rochelle. MEnIr a été soutenu par le programme du CNRS intitulé "Archivage et patrimoine documentaire, apports des sciences de l'information et de la cognition" (Gourmelon *et al.*, 2005(b)). Ce projet avait pour objectif de doter l'ODC d'un système d'information de manière à faciliter le porter à connaissance non seulement des produits de l'observation et de la recherche mais aussi des données sur les compétences (Rouan *et al.*, 2005).

4 <http://siec.univ-brest.fr/siec/>

5 développé par l'équipe de l'US Désertification de l'IRD : <http://www.mdweb-project.org/>

Le projet MEnIr a été mené par une équipe composée essentiellement de géographes et d'informaticiens, il s'est déroulé pendant deux ans. Nous présentons dans cet article la méthodologie de mise en œuvre d'une mémoire environnementale en explicitant les moyens en termes de réflexions et de nouvelles technologies de l'information et de la communication qui ont été utilisés.

La suite de cet article présente en section 2 la réflexion conceptuelle qui a conduit à une modélisation du système d'information. La section 3 décrit les normes de métadonnées et leur mise en œuvre dans le processus de développement d'une mémoire environnementale, puis nous présentons dans la section 4 l'implémentation qui a conduit à la réalisation d'un prototype⁶ dédié à l'Observatoire du Domaine Côtier de l'IUEM. Il tient compte de l'étude des besoins d'un système d'information similaire à développer pour l'Institut du Littoral et de l'Environnement (ILE) de l'Université de La Rochelle, dans un souci d'interopérabilité. Puis nous concluons et donnons des perspectives à ce travail en soulignant l'intérêt d'introduire l'ingénierie des modèles et des connaissances au développement des systèmes d'information environnementaux (SIE).

2. Modélisation d'une mémoire environnementale

2.1. Objectifs d'une mémoire environnementale

De façon concrète, une mémoire environnementale a pour but de rendre facilement accessibles certaines connaissances. Elle doit aussi être évolutive, aisément modifiable, extensible au fur et à mesure de l'évolution des connaissances sur une zone atelier donnée et de l'implication de nouveaux acteurs. Elle doit également permettre d'exploiter efficacement des informations pour lesquelles les protocoles de collecte ont changé au cours du temps, ce qui rend cruciale l'exploitation conjointe des données et des métadonnées. Si le recours aux technologies de l'information et de la communication semble incontournable, il reste cependant à déterminer la ou les technologies à mettre en œuvre, sans pour autant oublier la nécessité d'une modélisation conceptuelle. Il ne s'agit pas, ici, de dresser un panorama exhaustif des solutions technologiques, mais de proposer de possibles adéquations entre l'objet de la mémoire et les moyens de sa mise en œuvre.

2.2. Démarche formelle et analyse générale

La démarche est fondée sur un inventaire des acteurs, des projets, des sources de connaissances et des données disponibles. Un travail basé sur des enquêtes et des

6. Mémoire Environnementale Iroise : <http://menir.univ-brest.fr/>

réunions associant les différents acteurs a été réalisé au sein des deux observatoires (IUEM et ILE). La phase de modélisation s'est basée sur ce socle de connaissances.

Le système s'adresse à des chercheurs et à des étudiants évoluant au sein de laboratoires de recherche. Pour mener à bien cet inventaire, nous avons opté pour une modélisation à partir d'un formalisme éprouvé : UML (Unified Modeling Language) (OMG, 2004) standard de l'OMG (Object Management Group) qui permet d'aborder tous les aspects nécessaires à la conception d'un système informatisé (statiques, dynamiques et opérationnels). Vis-à-vis de la définition d'une mémoire environnementale, il apparaît que quatre entités se distinguent dans le domaine à spécifier :

- Les acteurs sont les personnes (chercheurs, étudiants, ...) appartenant à un laboratoire de recherche, qui participent à des projets, et qui dans ce contexte produisent des connaissances et des données d'observation. Nous distinguons trois types d'acteurs : internes, externes à la structure organisationnelle représentée par la mémoire environnementale donnée et ceux qui ont eu une activité vis-à-vis de celle-ci.

- Les productions sont les articles scientifiques, les rapports, les communications à des conférences, ...

- Les données d'observation sont des séries de données, des relevés de terrain, des mesures qui sont en rapport avec la zone géographique sur laquelle opère l'observatoire.

- Un projet est une action développée par des acteurs (projet de recherche, thèse, contrat) au cours de laquelle des productions vont être rédigées et des données vont être produites et utilisées. Afin de conserver l'évolution d'un projet au cours du temps, il est divisé en périodes qui peuvent être soit « en activité », soit « suspendu », ou « stoppé » (figure 1).

Des actions concertées avec les futurs utilisateurs du système ont permis de déterminer les entités constructives et descriptives de la connaissance. Nous avons rencontré l'ensemble des partenaires fournisseurs et utilisateurs d'information. Nos questions lors de ces entretiens portaient notamment sur les ressources possédées et sur les besoins des acteurs en termes d'informations

Connaissant les éléments à considérer et les limites du domaine à traiter, nous avons pu ensuite composer le contenu des mémoires (figure 1). Il est à noter que le concept de projet est partagé entre les mémoires de documentation et globale. En effet, certains attributs documentent des projets comme le contexte, les enjeux, la bibliographie, alors que d'autres attributs ou relations montrent l'évolution au cours du temps du projet, comme : la rédaction de publications, la production de données. Ces dernières informations, relatives au suivi de projet, contribuent à l'enrichissement de la mémoire globale.

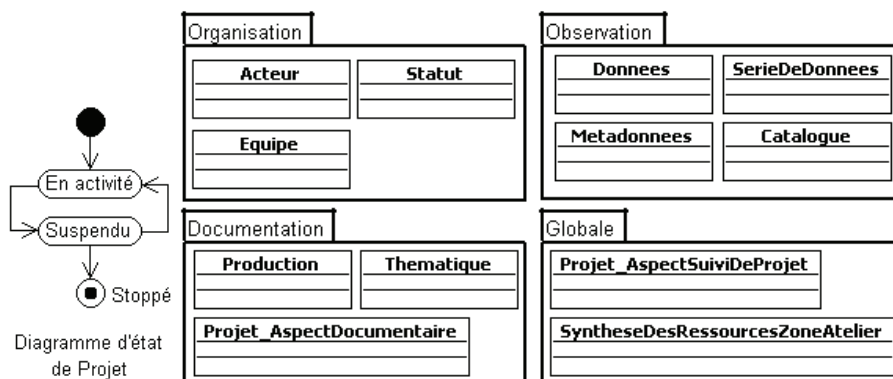


Figure 1. Diagramme d'état du concept de projet et vue package du contenu des mémoires de MENr

Le système doit proposer l'accès en consultation et en saisie des ressources (acteurs, projets, productions, données et métadonnées). Les droits d'accès en consultation sont choisis par le producteur de la ressource et sont spécifiques aux couples ressource-acteur autorisés.

2.3. Composition de modèles

Une mémoire environnementale trouve son originalité dans sa proposition de décomposer le domaine en mémoires spécialisées, qui constituent des modules fortement cohérents en minimisant les dépendances. On constate deux types de dépendances, celui des liens nombreux entre objets d'une même mémoire, et celui du lien plus ténu entre entités appartenant à des mémoires différentes. La notion de « projet », plus particulièrement son rôle fédérateur, contribue pleinement à fournir des relations inter-mémoires.

L'originalité de la modélisation réside dans la composition de modèles permettant la séparation des préoccupations (fonctionnalités). Cette composition de modèles est réalisée en empruntant les techniques du génie logiciel (patrons de conception) ou celles des documents numériques (normes et formats de métadonnées).

2.4. Modèles et patrons de conception

Un patron de conception ou « design pattern » en anglais est un objet conceptuel décrit en UML et conçu pour répondre à des besoins récurrents de conception en

informatique (Gamma *et al.*, 1995). Ces patrons fournissent de plus des gabarits d'implémentation, des méthodes pour y accéder et les manipuler (par exemple écrites en code Java). Nous avons principalement utilisé le patron de conception « composite » (Riehle, 1997) pour les aspects statiques et les patrons « Temporel Object », « Effectivity » et « Snapshot » pour les aspects dynamiques.

2.4.1. Patron de conception « composant-composite »

Il permet la construction de structures hiérarchiques (dans lesquelles chaque élément connaît la liste des éléments de niveau inférieur et l'élément de niveau supérieur. Ce patron donne de la flexibilité et de la profondeur au système en offrant des points de vue sur le contenu à différentes granularités comme par exemple pour le catalogue de données où nous avons des catégories de données qui jouent le rôle de nœuds et des séries de données qui jouent le rôle de feuilles.

2.4.2. Patrons de conception temporels

Les aspects dynamiques du système concernent les mémoires organisationnelle, documentaire et globale. En effet, c'est dans la gestion des statuts des acteurs et dans celle de leurs liens avec les équipes de rattachement que l'on observe des évolutions fréquentes dont on souhaite garder la mémoire. Ce caractère évolutif touche également les projets auxquels les acteurs participent et les productions scientifiques qui en résultent. Une question importante est de savoir délimiter l'intérêt que l'on porte aux aspects temporels. Dans certains cas, les compétences des acteurs sont cumulatives, et l'on ne souhaite pas conserver les informations sur les dates d'acquisition des dites compétences, ni celles concernant le contexte dans lequel ces compétences ont été maîtrisées. Les objets sont donc dépourvus de dimension temporelle. Un choix différent consiste à enregistrer la chronologie des états et des transitions des objets manipulés. Pour stéréotyper les « objets qui changent avec le temps », nous nous référons aux patrons de conception proposés par Fowler (Fowler, 2002, Carlson *et al.*, 1999) qui distingue la notion d'objet temporel (`temporalObject`) et celle de propriété temporelle (`Effectivity`) d'un objet.

Le patron « Temporal Object » permet de localiser un objet dans le temps à travers un rôle de « continuity », et d'assurer la gestion des versions successives de cet objet au cours de son cycle de vie. Ces versions possèdent un attribut « date » qui correspond à la création de l'instance. La date de fin de validité pour une instance correspond à la date de création de l'instance de version suivante. Ceci permet la reconstitution de l'évolution d'un phénomène au cours du temps.

Le patron « Snapshot » s'adresse aux objets temporels qui n'ont pas de durée d'existence tels que des événements non mémorisés mais pour lesquels il est utile d'avoir une date de validité (dates de création, de mise à jour, ...).

Le patron « Effectivity » permet de préciser le caractère valide d'une propriété pendant une période de temps avec des dates de début et de fin d'activité. Toutes les

instances des périodes seront conservées pour former des historiques. A noter que les périodes peuvent se superposer temporellement.

Ces trois patrons de conception ont pour principal intérêt d'encapsuler les attributs temporels et les méthodes associées. Ces propriétés donnent accès aux fonctionnalités telles que l'archivage de chaque état jugé significatif de l'objet stéréotypé, l'accès à l'objet dans l'état courant, précédent ou suivant, l'identification des dates de transition et des événements déclencheurs.

2.5. Utilisation des patrons de conception temporels pour le modèle d'archivage

La figure 2 illustre l'utilisation des patrons de conception temporels pour le stéréotypage des principales classes et associations de la mémoire d'organisation.

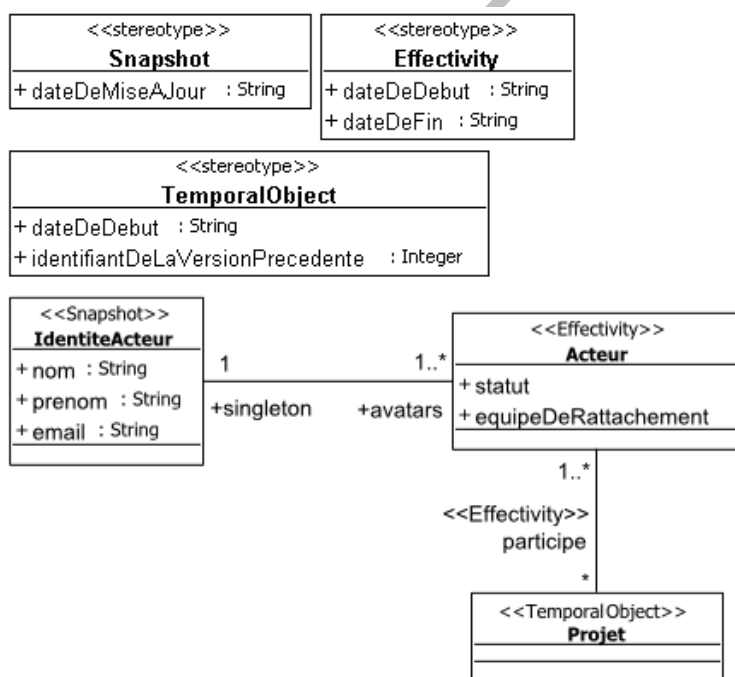


Figure 2. Utilisation des stéréotypes d'objets temporels pour MENIr

La classe « IdentityActor » représente l'aspect singleton du concept d'acteur du système et contient à la fois les informations invariantes (prenom, etc) et la valeur courante de certains paramètres (email, etc) dont l'historique n'est pas archivé. Cette

classe est stéréotypée par « Snapshot » qui témoigne de la date de la dernière mise à jour de l'objet considéré.

Au cours de son cycle de vie, l'acteur pérenne, instance d'« IdentiteActeur », joue des rôles variés dont on souhaite mémoriser la dynamique à travers la classe « Acteur », stéréotypée par « Effectivity ». Chaque rôle donne lieu à la création d'une instance de la classe « Acteur ». Toutes ces instances sont liées à l'unique « IdentiteActeur » qui leur correspond par un lien possédant un rôle d'association UML nommé « singleton ». Le rôle « avatars » se comporte comme un itérateur et permet de naviguer directement tous les rôles d'un acteur au cours de sa vie dans le système. Pour la classe « Projet », le stéréotype est « TemporalObject » car un projet traverse différents états successifs caractérisant son cycle de vie et dont on souhaite conserver la chronologie. Son cycle de vie est décrit en UML par un diagramme état transition (figure 1). Pour cette classe, il n'a pas été prévu de créer une classe mère (jouant le rôle de singleton) comme cela a été fait pour la classe « Acteur ». En effet, il a été choisi de dupliquer les informations de la classe « Projet » à chaque nouvelle version de celui-ci. Enfin, l'association « participe » stéréotypée par « Effectivity » rend compte de la participation d'un acteur à un projet sur une période de temps bornée.

3. Les normes de l'information géographique (ISO 19100) dans le processus de développement d'une mémoire environnementale

3.1. *Usage des normes de métadonnées pour les documents numériques et pour l'information géographique spécifiquement*

La nécessité de faire accompagner les données d'observation par une méta-information est acquise depuis longtemps. L'avènement des modèles objets, puis celui des informations semi-structurées ont renforcé ce principe en intégrant plus intimement les informations brutes avec leurs types, leurs structures, et les méthodes afférentes. Dans le domaine des sciences expérimentales ou humaines, il est inconcevable d'analyser une donnée hors d'informations complémentaires sur le contexte de l'obtention de cette donnée.

Parallèlement à cette préoccupation, la gestion statique et dynamique des données (et respectivement des métadonnées⁷) nécessite la connaissance d'informations contextuelles non pas sur la nature du processus de capture de la donnée, mais sur le statut de la donnée dans le modèle et dans la base persistante, par exemple pour le contrôle de sa cohérence (intégrité), de son évolution (contrainte de référence) et de sa distribution (droit d'accès).

7. Les métadonnées représentent un moyen pour documenter le contenu, le contexte, la qualité, la structure et l'accessibilité d'un lot de données numériques ou non. Elles jouent le rôle de mode d'emploi des données et assurent leur traçabilité.

Ce constat ne précise pas en quoi consiste la description du contexte, et la simplicité de l'idée cache la difficulté de dresser une liste consensuelle de ces métadonnées pertinentes et indispensables. On comprend bien que le contexte que l'on veut descriptif peut être dépendant du contexte d'exploitation de la donnée. L'existence de normes est une réponse à ces besoins communs de disposer d'une liste exhaustive, structurée, documentée et exploitable de ces métadonnées. Le terme de structure est ici fondamental car un simple catalogage linéaire serait très difficilement exploitable en pratique.

Nous avons choisi la série de normes ISO 19100⁸ largement utilisée et qui sert couramment de référence dans le domaine spatio-temporel. Cependant, l'essentiel de notre travail pourrait être développé avec d'autres normes qui offrent des concepts et des fonctionnalités en général similaires. La série ISO 19100 est développée depuis 1994 par le groupe de travail TC211 de l'International Organization for Standardization (ISO).

Les normes ISO 19100 adoptent un formalisme objet, décrit en UML pour le niveau conceptuel et (éventuellement) en XML pour le niveau logique. Elles s'articulent autour d'un document central ISO 19115 (ISO, 2003) qui offre l'ensemble des concepts et fonctionnalités pour décrire les métadonnées attachées à des données. Des normes spécifiques à certains aspects (mesures physiques, données temporelles, ...) existent et sont entièrement compatibles avec l'ISO 19115.

Ces normes sont auto-documentées en cela qu'elles incluent une part de leur propre sémantique (à travers l'usage de dictionnaires, de stéréotypes UML, ...) ainsi que des règles (texte en langage naturel) sur la manière de les utiliser. L'ISO 19115 affiche, dès sa première page, sa vocation à décrire une information géographique. Par expérience, on constate que cet axiome semble rebuter de nombreux utilisateurs potentiels, non nécessairement prioritairement intéressés par l'aspect géoréférencement de leurs données. En fait, s'il est vrai que le noyau (*core*) décrit l'ensemble minimal des métadonnées obligatoires pour décrire un élément de donnée géographique (*geographical dataset*), le reste est applicable à des données de domaines quelconques.

Enfin, et cela concerne une part importante de notre travail actuel et sans doute de ses développements futurs, l'esprit et le formalisme dans lesquels sont décrites les normes peuvent être exploités pour développer des extensions spécifiques à certains aspects métiers spécialisés. On notera à ce sujet que les normes ISO 19100 spécifient elles-mêmes leur mode d'extension, lui aussi ainsi normalisé.

8. <http://www.isotc211.org>

3.2. *Autres apports de la normalisation*

Nous ne revenons pas sur l'intérêt fondamental de disposer d'un standard lors de la conception d'un système d'information concernant des utilisateurs et des pourvoyeurs de données hétérogènes. La norme est l'élément de base pour toute volonté d'intégration de données et de construction d'un système de médiation (Bigagli *et al.*, 2006 ; Libourel, 2003). Par contre nous insisterons sur les apports corollaires dont nous détaillerons des exemples d'exploitation dans les paragraphes à suivre.

Dès que l'on dispose d'un système de métadonnées normalisées structuré comme par exemple la série ISO 19100, d'un ensemble organisé de termes et de concepts qui s'avèrent extrêmement utiles non seulement pour documenter les éléments du modèle de données – ce qui est leur rôle premier – mais aussi pour servir d'information de base à la constitution d'une ontologie du domaine (Gruber, 2003). Tout au moins, selon la richesse du modèle, d'un thesaurus fortement structuré comme le prévoit la norme avec un système natif de mots-clés que la navigation peut également exploiter.

L'ingénierie dirigée par les modèles (IDM) ou plus généralement de l'approche MDA (*Model Driven Architecture*) décrite par Miller *et al.* (2003) permet de manipuler systématiquement par programme les modèles de données et de métadonnées. Cette approche assure automatisation, rejouabilité et traçabilité. C'est à dire finalement, ergonomie et garantie de cohérence dans le processus d'intégration de données. Des outils dédiés à la métamodélisation ont été développés, tel que ATL (Bézivin *et al.*, 2003) ou encore Kermeta (Muller *et al.*, 2005). Nous avons déjà expérimenté ces outils sur l'ISO 19100 dans Faucher *et al.* (2007).

La structure des métadonnées construit également *de facto* une documentation du système conçu, et ceci d'autant plus efficacement que les classes du modèle du domaine peuvent tout naturellement être considérées comme des données complexes et se voir associer des métadonnées en toute conformité avec le modèle général de la norme.

Les points précédemment évoqués (ontologies, thesaurus, méta-description du modèle du domaine,...) sont naturellement parmi les arguments de base sur lesquels s'articule un système de navigation. L'indexation et la recherche de données se basent à la fois sur les valeurs (données) et sur les métadonnées (abstractions).

3.3. *Présentation des normes utilisées : ISO 19115 Geographical information : Metadata*

L'ISO 19115 décrit précisément l'ensemble des métadonnées susceptibles d'accompagner les données dans le but :

- d'aider les producteurs de données à définir les informations utiles pour caractériser leurs données ;
- de faciliter l'organisation et la gestion des métadonnées ;
- de permettre aux utilisateurs de connaître les données qu'ils manipulent ;
- de faciliter la fouille de données, la restitution d'information et leur réutilisation ;
- de permettre aux utilisateurs de juger rapidement si une donnée d'un entrepôt est d'intérêt ou non pour eux.

La structure de l'ensemble des métadonnées est représentée par plusieurs packages de diagrammes de classes UML. Chaque package est spécialisé suivant un usage et les métadonnées sont également structurées de façon transversale aux packages, selon le niveau d'abstraction auquel elles s'appliquent (de « l'agrégat » supérieur jusqu'au « dataset »).

Il est important d'insister sur le fait que la totalité des métadonnées décrites n'est pas nécessaire. Les métadonnées s'utilisent en référence à des « profils » qui ne retiennent que les métadonnées strictement utiles (obligatoires/optionnelles) dans un contexte particulier. Ces profils sont définis dans l'ISO 19106 et la manière d'en construire est décrite dans l'ISO 19115 Annexe C. Le sous-ensemble minimal de métadonnées obligatoires est d'ailleurs très restreint (dix champs tout au plus).

3.4. Articulation domaine-normes (extension, tissage, ...)

Le contenu des mémoires est par nature voué à évoluer au cours du temps. Il est par conséquent indispensable de raisonner en termes de système minimal auquel peuvent être ajoutées des extensions. L'étude menée conjointement dans deux instituts s'intéressant à l'observation à long terme du littoral a permis d'élargir le spectre des besoins, d'affiner les modèles et d'obtenir ainsi une vision complète et partagée. Il apparaît que la norme ISO 19115 pourrait constituer une mémoire environnementale minimale, en y ajoutant tout de même les notions de projet, l'enregistrement de l'historique des « CI_ResponsableParty » (société, laboratoire, personne) et l'indexation des lots de données en « DS_Aggregate ». Ce niveau de base peut ensuite être étendu par raffinement notamment par des ajouts d'attributs comme pour les spécialisations de production qui joueraient le rôle de « CI_Citation ». Ces extensions peuvent être directement réalisées grâce aux mécanismes d'extension propre à la norme, ou encore obtenues par spécification de liens entre les éléments du modèle du domaine et les classes de la norme. Nous proposons d'utiliser une technique de l'ingénierie des modèles qui consiste à réaliser un tissage entre le modèle métier et la norme (Didonet *et al.*, 2005). Le modèle résultant du tissage et le modèle original du domaine sont distincts, ce qui évite l'introduction de nouvelles entités dans la structure de la norme, contrairement à l'utilisation du mécanisme d'extension.

La figure 3 montre quelles sont les dépendances et les origines des types d'attributs et en particulier s'ils proviennent des normes ou du domaine d'application (*from MemEnv*). Ainsi on remarque que tous les attributs relatifs aux projets sont issus du domaine d'application alors que les attributs des productions et des acteurs proviennent à la fois des normes et du domaine d'application. Nous montrons à partir de cette figure une utilisation des mises en relation domaine-normes (tissage) décrites précédemment.

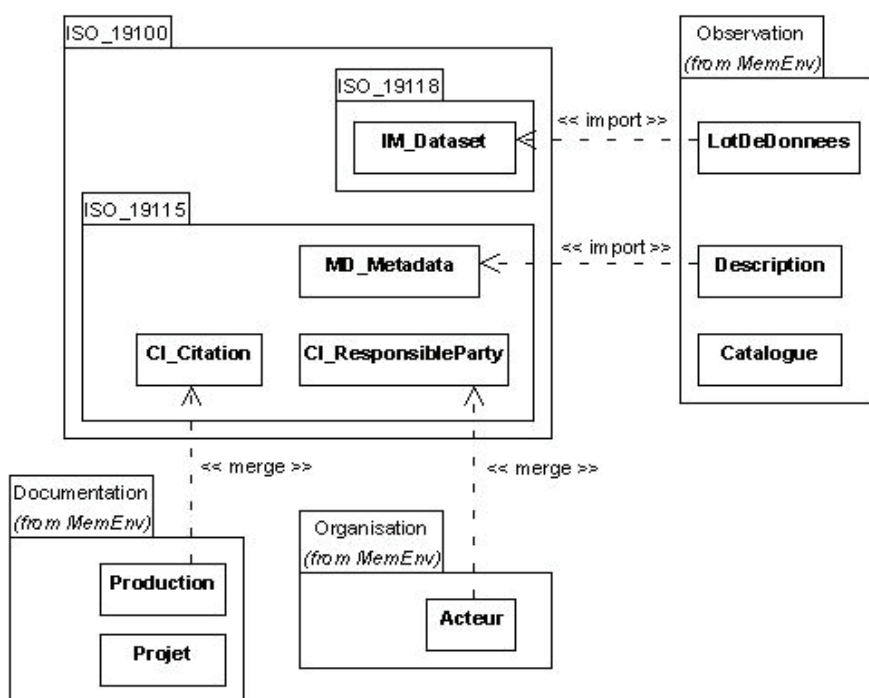


Figure 3. Provenance des informations intégrées dans les mémoires

3.5. Interopérabilité (compatibilité entre systèmes)

En ce qui concerne l'information géographique, beaucoup de travaux ont été menés ces dix dernières années afin de développer l'interopérabilité. Le format d'échange de données géoréférencées GML (Geographic Markup Language ou encore ISO 19136), basé sur XML, réunit à la fois les données et leurs métadonnées. L'ISO, par l'intermédiaire du groupe de travail TC211, propose des standards notamment l'ISO 19115, l'ISO 19118 précitées et l'ISO 19139 qui est

l'implémentation en XML de l'ISO 19100. L'Open Geospatial Consortium⁹ (OGC) donne des recommandations et des exemples pour l'utilisation de ces standards. Ces outils de standardisation sont de plus en plus utilisés dans la communauté et prennent une grande importance dans la conception de nouveaux systèmes d'information. Par contre, le « temps d'incubation » de ces normalisations et protocoles peut être pénalisant aussi bien pour les informaticiens que pour les thématiciens.

Les bases de données réparties, l'interopérabilité et la médiation deviennent incontournables pour améliorer et accélérer les échanges, ce qui suppose de mettre en place des systèmes d'information utilisant des normes et des formats d'échange standards, susceptibles d'être facilement traités par les systèmes de médiation. On peut ainsi concevoir des systèmes de recherche d'information dirigés par le contexte qui s'adaptent à l'environnement habituel de l'utilisateur.

4. Implémentation d'une mémoire environnementale

Après la modélisation du système d'information, un prototype de site Web dédié à l'Observatoire du Domaine Côtier de l'IUEM a été implémenté en tenant compte des besoins également identifiés par l'ILE, dans un souci de future interopérabilité. La phase de prototypage a permis de valider les modèles et de mettre en évidence les problèmes conceptuels et opérationnels.

4.1. Architecture système

L'architecture du système s'appuie sur les techniques de diffusion et d'exploration du Web, ainsi sur une architecture *3-Tier* client-serveur. Elle permet d'avoir un client léger et ainsi de limiter les logiciels nécessaires pour l'utilisation de l'application, puisque seul un navigateur Web est nécessaire. Les solutions choisies sont des composants libres de droits dit « OpenSource » et qui sont très utilisées dans la communauté : Serveur Apache, base de données MySql et le langage Php.

4.2. Points de vue

Une fois les modèles conçus et implémentés, il est nécessaire de réfléchir aux interfaces utilisateurs propres à les rendre opérationnels afin de combiner les aspects textuels, graphiques, symboliques, cartographiques, ergonomiques pour concevoir la partie visible du système par l'utilisateur final. Ainsi la vue du « client » doit refléter l'idée que se fait l'utilisateur de la zone étudiée par le SIE, i.e. son environnement,

9. <http://www.opengeospatial.org/>

ses caractéristiques, ses structures, ses acteurs et ses résultats. Les points de vue « utilisateur » et « structurel » sont différents, en effet la structuration en mémoires diffère de l'attente de l'utilisateur, mais se prête bien du point de vue du modélisateur du domaine et de la persistance des données. La figure 4 montre ces deux points de vue d'une mémoire environnementale et précise le rôle de celle-ci en tant qu'« organisateur » ou en tant que « diffuseur » d'information. Ainsi la couche modèles de données s'organise selon le principe des mémoires et la couche de présentation se compose selon les structures (observatoire, organisme), les hommes (chercheurs, fournisseurs de données) et les travaux (projets de recherche, publications). L'implémentation a bien entendu suivi ce schéma et plus particulièrement le patron de conception « Modèle-Vue-Contrôleur » (MVC).

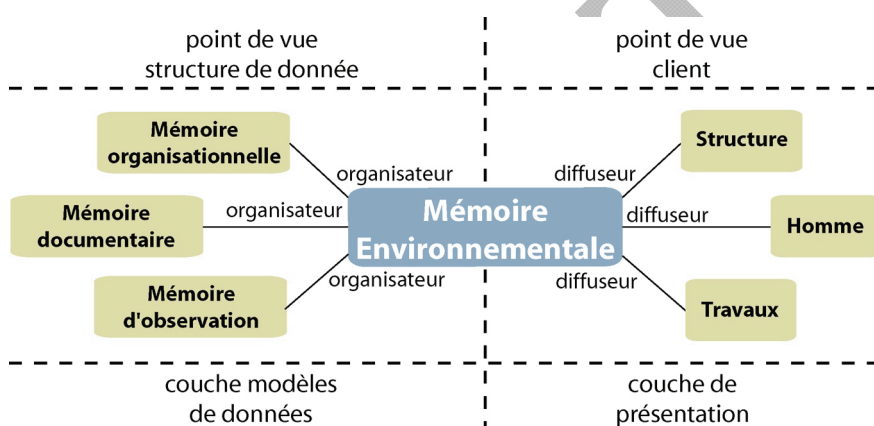


Figure 4. Points de vue "structure de donnée" et "client" sur une mémoire environnementale

4.3. Résultats d'implémentation

4.3.1. Patron de conception « composant-composite » appliqué au catalogue de données

Le catalogue de données de MEnIr est basé prioritairement sur l'aspect navigation des données, puis dans un second temps des métadonnées (pointeurs d'une donnée vers sa métadonnée). En effet, ce sont les données d'observation qui sont affichées alors que les métadonnées servent à construire le catalogue (mots-clés, catégories, etc). Le catalogue de données se visualise sous la forme d'un graphe dans lequel les nœuds sont des catégories de données et les feuilles des séries de données. Les lots de données (ou dataset) sont liés à ces séries. L'utilisation du patron de conception « composant-composite » permet d'attacher une même « catégorie de données » à plusieurs nœuds du graphe. La catégorie n'est

pas dupliquée dans la base de données, c'est sa référence qui est utilisée plusieurs fois. Ainsi le gestionnaire de données peut faire évoluer le catalogue en proposant aux utilisateurs de nouveaux chemins d'accès aux données sans induire des changements sur les données ou les métadonnées. Il est à noter que le catalogue qui a été développé permet, de la même manière, de positionner des lots de données sous plusieurs catégories. Les catégories de plus haut niveau (les plus abstraites) qui ont été utilisées correspondent à celles fournies par l'ISO 19115, i.e. : « Domaine, Thème, Localisation, Temporel ». Les catégories inférieures ont été déterminées suivant les besoins des utilisateurs.

4.3.2. Consultation d'objets à paramètres temporels

M. Fowler propose des implémentations en langage objet des patrons de conception « TemporalObject », « SnapShot » et « Effectivity » - déjà évoqués en section 2.5. - et fournit notamment des méthodes de construction et de parcours des structures de données correspondantes. Il s'agit de déterminer les méthodes de stockage dans la base de données et d'adapter les scripts en conséquence. Les règles de passage d'un modèle UML à un modèle relationnel suffisent pour satisfaire nos exigences et la traduction entre le code « java » qu'utilise Fowler et le langage de programmation que nous avons utilisé : « php » est dans ce cas assez simple.

La restitution des informations temporelles aux utilisateurs a été implémentée de manière à construire des vues sur les objets stéréotypés « temporel » que nous avons nommées « historiques ». Ces historiques sont diffusés sous forme d'images, le visuel étant un bon moyen d'interpréter rapidement ce type d'information. Le temps est représenté en abscisse et chacune des formes employées correspond à un modèle temporel. Ainsi nous représentons les « SnapShots » par une forme ponctuelle avec soit une barre verticale (i.e. acquisition d'une compétence) (figure 5), soit un pictogramme (i.e. utilisation d'une donnée) (figure 6). Les « Effectivity » seront quant à elles symbolisées par des barres horizontales (i.e. liaison à une équipe de rattachement, participation à un projet) (figure 5). Enfin un « TemporalObject » forme une succession de barre horizontale contiguë et rend compte de l'évolution d'un projet au cours du temps (figure 6).

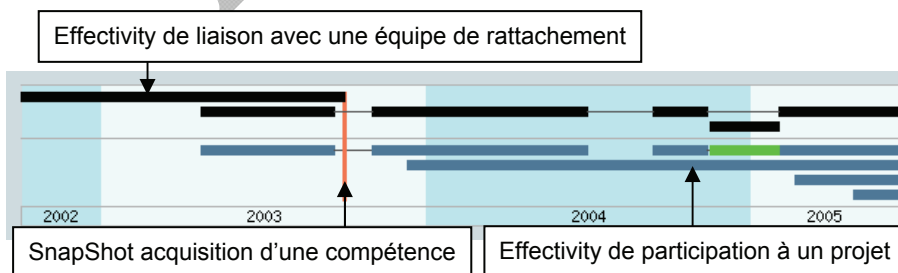


Figure 5. Historique d'un acteur

Pour l'historique de projet, le nombre trop important de données générées par les « Effectivity » entre projet et acteur nous a amené à afficher séparément ces informations (dates d'arrivée et de départ d'un acteur dans un projet). Ainsi la symbolique d'une barre horizontale pour un « Effectivity » est écartée et remplacée par des notions « d'entrée » et de « sortie » dans un projet qui s'assimilent à des « SnapShot », par conséquent à du ponctuel (figure 6). L'information ainsi fournie est quelque peu dégradée mais plus compréhensible. A noter que la structure de données des « Effectivity » reste inchangée, seule la forme de l'information restituée diffère.

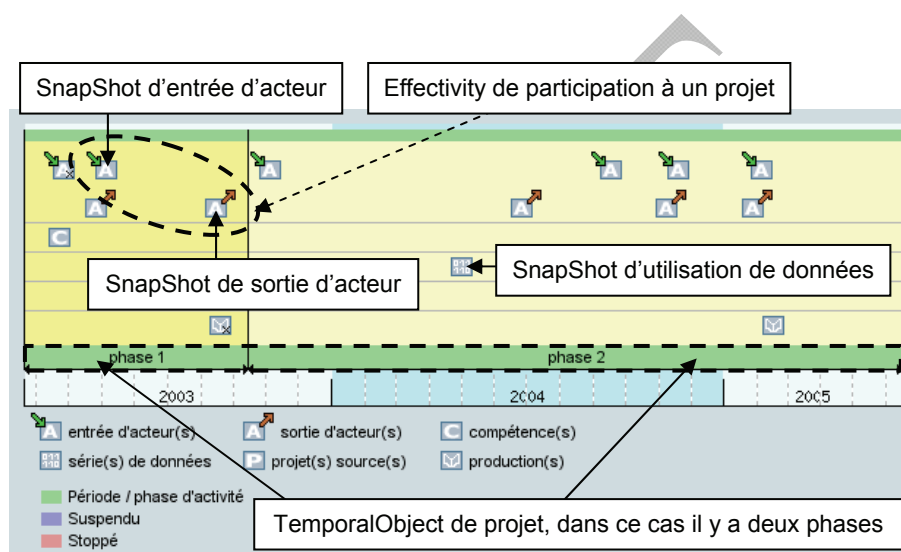


Figure 6. Historique d'un projet

5. Conclusions et perspectives

Nous avons décrit dans cet article la réalisation d'un système d'information développé selon la notion de mémoire environnementale. En plus du catalogage et de l'archivage de l'information, il procure à l'utilisateur tout un ensemble d'éléments permettant de replacer l'information dans son contexte scientifique et généalogique. De l'analyse des liens entre les mémoires mises en œuvre pourra émerger une mémoire globale qui fournira aux futurs utilisateurs des clefs pour améliorer leur connaissance de l'environnement.

La mise en œuvre informatique d'un tel système a demandé une phase conséquente de modélisation tant au niveau des concepts généraux qu'au niveau de la formalisation des parties métier. Cet ensemble de modèles a permis de mettre en évidence des modèles de base qui pourraient être généralisés et réutilisés sur de

nouvelles zones d'étude. L'utilisation d'un langage unifié (UML), de modèles « préfabriqués » comme les patrons de conception temporels ou les normes de la série ISO 19100 confère un caractère exhaustif et rigoureux à la modélisation, favorise l'interopérabilité avec des systèmes voisins et la mutualisation de ressources réparties. C'est dans la composition entre métamodèles, modèles, patrons et normes que ce projet offre un processus original de conception de système d'archivage et de recherche d'information. Le prototypage du système d'information a, dans un premier temps, contribué à valider les modèles et les principes d'archivage, puis dans un second temps à implémenter la norme de métadonnées ISO 19115.

Désormais nous disposons des bases conceptuelles et logiques pour exploiter et maintenir sur le long terme les structures persistantes qui ont été mises en place. Les besoins s'orientent maintenant vers l'adaptation du contenu en fonction de l'utilisateur afin d'assurer une assistance plus ou moins forte, voir proposer un mode « expert » pour la navigation. Dans un second temps, les modèles, ontologies et thesaurus dont nous disposons ouvrent la voie à la conception de systèmes d'intégration de données avec des systèmes exogènes. Il convient alors de spécifier des architectures de médiation, de concevoir des correspondances d'ontologie et d'exploiter les modèles abstraits communs pour faire communiquer des systèmes développés dans des environnements hétérogènes. L'ingénierie des modèles peut nous y aider.

Si les structures logiques et physiques développées jusque là sont bien adaptées à la gestion de la persistance et prennent effectivement en compte l'évolution des besoins, elles ne peuvent efficacement être exploitées telles quelles à des fins de navigation. Nous étudions d'ores et déjà la spécification de modèles adaptés aux objectifs de navigation (OWL¹⁰, RDF¹¹). L'intégration de notre démarche dans un processus de conception traçable tel que MDA confèrera une souplesse d'évolution au système.

Remerciements

Nous saluons la mémoire de F. Cuq, initiateur du projet. Nous remercions les « acteurs » du projet, membres de l'IUEM, notamment : D. Gac, E. Hardy, E. Giraudet, I. Le Berre et C. Tissot. Remerciements également à E. Feuteun et aux équipes de l'ILE pour le soutien qu'ils nous ont apporté.

10. Web Ontology Language : <http://www.w3.org/2004/OWL/>

11. Ressource Description Framework : <http://www.w3.org/RDF/>

6. Bibliographie

- Bézivin J., Dupé G., Jouault F., Pitette G., Rougui J.E., « First Experiments with the ATL model transformation language: Transforming XSLT into XQuery », *OOPSLA 2003 Workshop*, Anaheim, USA, 2003, p. 1-18.
- Bigagli L., Nativi S., Mazzetti P., « Mediation to deal with information heterogeneity – application to Earth System Science », *Advances in Geosciences*, vol. 8, 2006, p. 3-9.
- Carlson A., Estep S., Fowler M., « Temporal Patterns », In Harrison, Foote, Rohnert (eds), *Pattern Languages of Program Design 4*, Addison-Wesley, 1999, p. 241-262.
- Cuq F., « Systèmes d'information géographique et gestion intégrée des zones côtières », *CoastGIS'99 : Geomatics and coastal environment*, Ifremer/SHOM, 2000, p. 18-29.
- Desconnets J.C., Libourel T., Clerc S., Granouillac B., « Cataloguing for distribution of environmental resources », *AGILE 2007 : the 10th AGILE Conference*, Aalborg, Denmark, 2007.
- Didonet Del Fabro M., Jouault F., « Model Transformation and Weaving in the AMMA Platform », *Proceedings of the Generative and Transformational Techniques in Software Engineering (GTTSE'05) Workshop*, Portugal, 2005, p. 71-77.
- Dwyer N., O'Dea L., Cummins V., « The marine irish digital atlas : a web portal to coastal and marine data in Ireland », *Proceedings of the international conference Littoral*, http://mida.ucc.ie/assets/documents/littoral_2004_mida.pdf, 2004.
- Faucher C., Lafaye, J.Y., « Model Driven Engineering for implementing the ISO 19100 series of international standards », *CoastGIS'07 : the 8th International Symposium on GIS and Computer Mapping for Coastal Zone Management*, Santander, Spain, 2007, p. 424-433.
- Fowler M., « Patterns for things that change with time », <http://www.martinfowler.com/ap2/timeNarrative.html>, 2002.
- Gamma E., Helm R., Johnson R., Vlissides J., *Design Patterns : Elements of Reusable Object Oriented Software*, Addison-Wesley, 1995.
- Gourmelon F., Le Berre I., « Towards an institutional GIS for Iroise Sea (France) », *GIS for Coastal Zone Management*, Taylor & Francis, 2004, p. 141-152.
- Gourmelon F., Robin M., *SIG et littoral, Traité IGAT*, Paris, Hermès, 2005(a).
- Gourmelon F., Lafaye J.Y., Rouan M., Faucher C., « MEnIr : Mémoire Environnementale Iroise », *Actes du colloque Société de l'information*, ENS-LSH Lyon, France, 2005(b), p. 128-131.
- Gruber T.R., « Toward principles for the design of ontologies used for knowledge sharing », In Guarino N., Poli R. (eds), *Formal Ontology in Conceptual Analysis and Knowledge Representation*, Padova, Italy, 1993.
- Guarnieri F., Garbolino E., Houllier F., Cuq F., Lévêque C., Weill A., Matarasso P., « Contribution à la définition opérationnelle et à la modélisation de la mémoire environnementale des zones ateliers », In Lévêque C., Leeuw S. (eds), *Quelles natures voulons-nous ? Pour une approche socio-écologique du champ de l'environnement*, Paris, Elsevier, 2003, p. 296-307.

- International Organization for Standardization, Geographic Information – Metadata, ISO 19115:2003(E), Draft International Standard, Genève, 2003.
- Libourel T., Autour de la conception de systèmes complexes : modélisation, évolution, infrastructures, Habilitation à Diriger des Recherches, Université Montpellier II, 2003.
- Miller J., J. Mukerji J., MDA Guide, Version 1.0.1, 2003.
- Muller P.A., Fleurey F., Jézéquel J.M., « Weaving executability into objectoriented meta-languages », In Kent S. and Briand L. (eds), *Proceedings of MODELS/UML'2005*, volume 3713 of LNCS, Montego Bay, Jamaica, Springer-Verlag, 2005, p. 264-278.
- OMG, Unified Modeling Language: Superstructure, Version 2.0, 2004.
- Riehle D., « Composite Design Patterns », *Proceedings of the Conference on Object-Oriented Programming Systems, Languages and Applications (OOPSLA '97)*, 1997, ACM Press, p. 218-228.
- Robin M., Morsetti V., Pichereau D., Babillot P., Colas S., Texier C., « L'information géographique de référence en zone côtière : la situation française », In Gourmelon F., Robin M. (coord.), *SIG et littoral*, Traité IGAT, Paris, Hermès, 2005, p. 49-83.
- Roselt/OSS, Système de circulation de l'information ROSELT : Définition des métadonnées et élaboration des catalogues de référence. Collection contributions techniques, 2004, 73p.
- Rouan M., Gourmelon F., Lafaye J.Y., Faucher C., Giraudet E., « An Information System for the Coastal Domain Observatory within the framework of the European Institute for Marine Studies », *Proceedings of the International Marine Data and Information Systems conference (IMDIS)*, Ifremer and COI, Brest, France, 2005, p. 88.