

# Perspectives d'utilisation du concept d'entrepôt de données pour les géorépertoires sur Internet

**François Létourneau, M. Sc.<sup>1</sup>**

**Dr Yvan Bédard<sup>1,2</sup>**

**Dr Bernard Moulin<sup>1,3</sup>**

1 : Centre de recherche en géomatique, Université Laval

2 : Département des sciences géomatiques, Université Laval

3 : Département d'informatique, Université Laval

[francois.letourneau@scg.ulaval.ca](mailto:francois.letourneau@scg.ulaval.ca) \*

[yvan.bedard@scg.ulaval.ca](mailto:yvan.bedard@scg.ulaval.ca)

[moulin@ift.ulaval.ca](mailto:moulin@ift.ulaval.ca)

\* adresse actuelle : [francois.letourneau@drdc-rddc.gc.ca](mailto:francois.letourneau@drdc-rddc.gc.ca)

## **Aperçu bibliographique**

M François Létourneau a obtenu récemment sa maîtrise en géomatique de l'Université Laval. Au cours de ses études, il a oeuvré au Centre de recherche en géomatique (CRG). Ses intérêts de recherche incluent les entrepôts de données spatiales, le développement d'applications géomatiques dynamiques sur Internet, les géorépertoires et la modélisation numérique de terrain. Il travaille présentement au Centre de Recherches pour la Défense de Valcartier (CDRV) à titre de scientifique de la Défense.

M Yvan Bédard est professeur titulaire au Département des sciences géomatiques et fut le directeur fondateur du CRG. Très engagé dans la recherche sur les SIG et sur les bases de données spatiales depuis 17 ans, il a participé à plusieurs projets gouvernementaux et industriels, en plus d'avoir été directeur scientifique du Centre de développement de la géomatique.

M Bernard Moulin est professeur titulaire au Département d'informatique. Il travaille depuis près de 15 ans dans les domaines du génie logiciel et de l'intelligence artificielle.

## **Résumé**

L'hétérogénéité des géorépertoires sur Internet entraîne un certain nombre de problèmes pour l'utilisateur qui désire en consulter plusieurs afin de trouver l'information qu'il recherche. Cette hétérogénéité se détaille ainsi : hétérogénéité du contenu, de la structure et des référentiels spatiaux, disparités sémantiques et absence d'outils d'aide à la sélection des documents. Une solution envisagée pour résoudre ces problèmes consiste à intégrer le contenu des géorépertoires hétérogènes

dans un entrepôt de données. Toutefois, nous ne connaissions pas la faisabilité conceptuelle, opérationnelle ou technique de cette approche. Nous avons donc analysé le potentiel de l'approche entrepôt de données pour produire un guichet unique pour un ensemble de géorépertoires sur Internet. Parallèlement à cette étude, nous avons conçu le prototype SOSDS (Système optimal de sélection des données spatiales) accessible via Internet. Ce prototype nous a permis de valider certains concepts de la recherche théorique ainsi que de proposer des méthodes de requête innovatrices pour exploiter à fond les données de l'entrepôt de données. Cet article présente la démarche suivie ainsi que les principaux résultats obtenus au cours de ce projet de recherche réalisé par l'auteur principal, sous la supervision des co-auteurs.

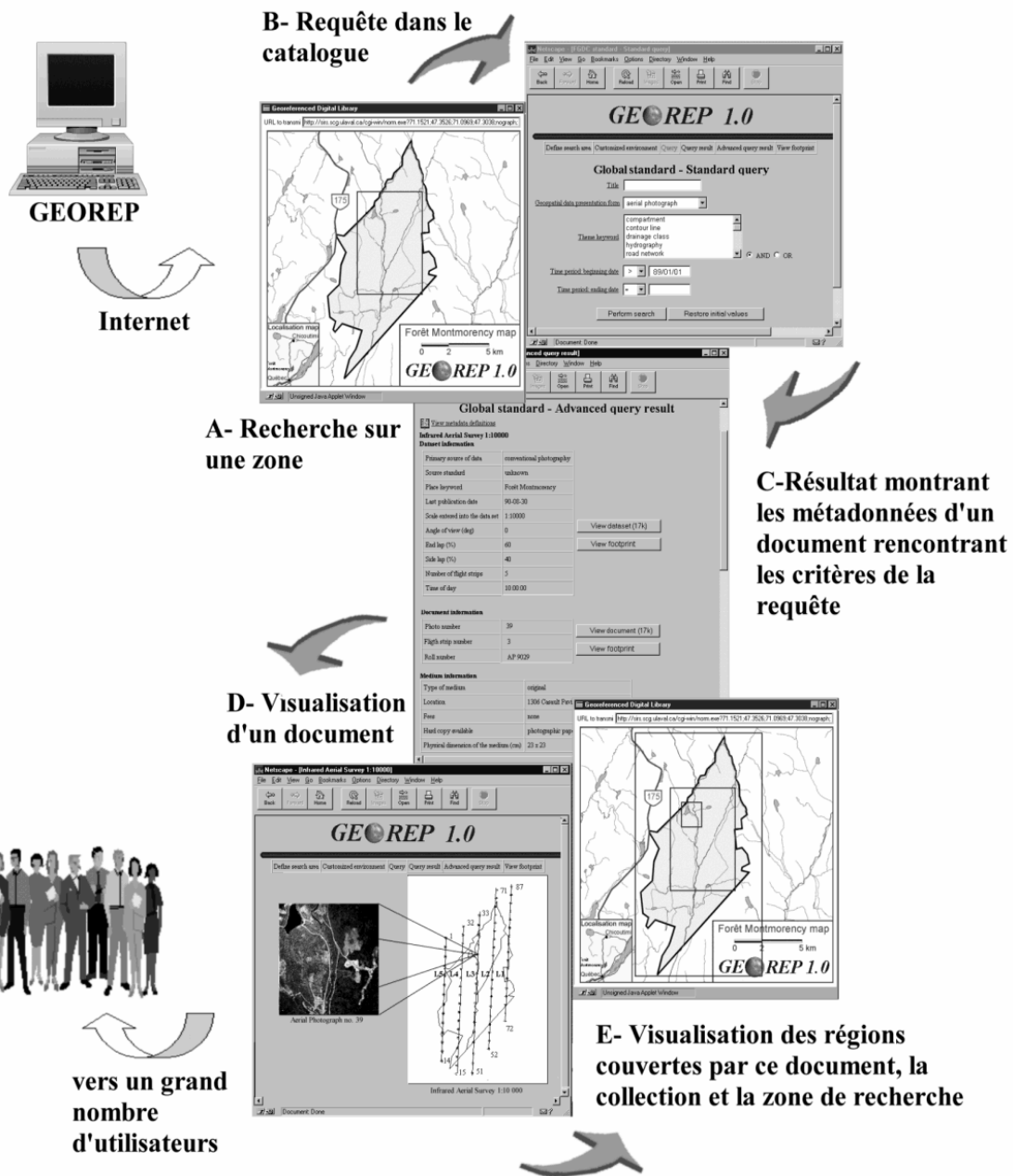
## **Abstract**

The implementation by different organizations of Georeferenced Digital Libraries (GDL) on the World Wide Web creates several problems for users who must query these heterogeneous libraries for a given territory. They face differences in contents, structures, spatial reference systems, semantics, interfaces, etc. In addition, these GDLs usually do not provide on-line support for the users who want to select the best products according to his needs. Our solution to these problems is to import some metadata from these GDLs, to process their data and to integrate the results in an homogeneous spatial data warehouse which serves as a "one-stop" query engine. Such solution is known in computer sciences as "Data Warehousing" and is theoretically very promising. However, there was a high level of uncertainty with regard to such a solution: conceptually, operationally and technically. Consequently, we started a research and development project to further study these theoretical possibilities and to experiment the resulting concepts. This resulted in a prototype called SOSSD (System for the Optimal Selection of Spatial Data) developed by the main author for his M.Sc., with the two co-authors as advisors. This paper presents our theoretical approach as well as the results obtained with SOSSD, including the implementation of a spatial data warehouse on the Web and the use of fuzzy logic to support the users in their queries.

## **1. Introduction**

Le développement des technologies géomatiques au cours des trente dernières années a mené à la production d'une quantité phénoménale de données à référence spatiale, que ce soit par les gouvernements ou par l'industrie. Rapidement, des besoins particuliers sont apparus afin de mieux gérer et diffuser ces données.

Aujourd'hui, la connaissance et la diffusion du patrimoine géo-informationnel d'une organisation reposent sur des moyens modernes, soit les géorépertoires et Internet. Un géorépertoire est un catalogue décrivant et géoréférençant des jeux de données géomatiques (ex. cartes, photographies aériennes, images satellitaires). De façon plus précise, un géorépertoire est « un index spatial informatisé de documents géoréférencés contenant les métadonnées permettant d'en connaître le contenu et la nature et de visualiser ces documents au complet ou sous forme d'extraits. Il possède une interface graphique qui permet de définir ou de sélectionner la région d'intérêt sur laquelle porte la recherche documentaire ainsi que les fonctions nécessaires à la définition des paramètres de recherche » [LÉT 98], figure 1. Le géorépertoire doit nécessairement permettre des requêtes



spatiales, sans quoi il ne constitue qu'un simple répertoire géodocumentaire.

Le principal problème de la première génération de géorépertoires était leur incapacité à permettre à un grand nombre d'utilisateurs d'accéder à l'information qu'ils contenaient. En effet, ils étaient pour la plupart basés sur des systèmes d'information géographique (SIG) accessibles uniquement sur un seul poste ou au mieux sur un réseau local. Le couplage Internet et géorépertoire s'est fait tout naturellement, à partir de l'année 1994. En effet, les technologies Web (ainsi que Java, les systèmes de bases de données relationnelles en ligne et les logiciels de cartographie dynamique sur Internet)

ont rendu possible la diffusion des informations contenues dans les géorépertoires à un grand nombre d'utilisateurs, peu importe leur localisation et l'heure d'accès. Dès lors, les géorépertoires sont devenus des applications très en vogue dans le domaine des technologies géomatiques sur Internet. Il en résulte la mise en place de plus en plus de systèmes qui couvrent ou non les mêmes territoires mais surtout, qui sont très hétérogènes. L'utilisateur ne s'y retrouve qu'avec difficulté et ne peut que souhaiter l'apparition d'un guichet unique.

Parmi les solutions possibles à ce problème, notre équipe de recherche a expérimenté deux approches : la première est basée sur l'utilisation de l'approche « entrepôt de données » (*data warehouse*) [LÉT 98] alors que la deuxième exploite les agents logiciels dans un contexte d'interopérabilité [MAA 97], [MAA 98]. Notre article vise à présenter les résultats obtenus avec l'approche « entrepôt de données ». Il s'inspire des travaux de recherche effectués par l'auteur principal au Centre de recherche en géomatique (CRG) de l'Université Laval, sous la supervision des co-auteurs. Ainsi, après avoir présenté le contexte et la problématique de notre recherche, nous présenterons le concept d'entrepôt de données et son extension pour les données spatiales. Par la suite, nous introduisons notre solution de guichet unique appelée SOSDS, laquelle utilise justement ce concept d'entrepôt de données. Entre autres, nous y traiterons du module d'aide à la sélection des meilleurs jeux de données obtenus parmi ceux provenant de tous les géorépertoires hétérogènes couvrant le secteur désiré.

## **2. Mise en contexte et problématique**

Cette section vise à présenter le contexte actuel d'utilisation des géorépertoires sur Internet et la problématique générale qui en découle, soit la présence de géorépertoires hétérogènes nécessitant de la part des utilisateurs des efforts constants dans leur recherche d'information spatiale.

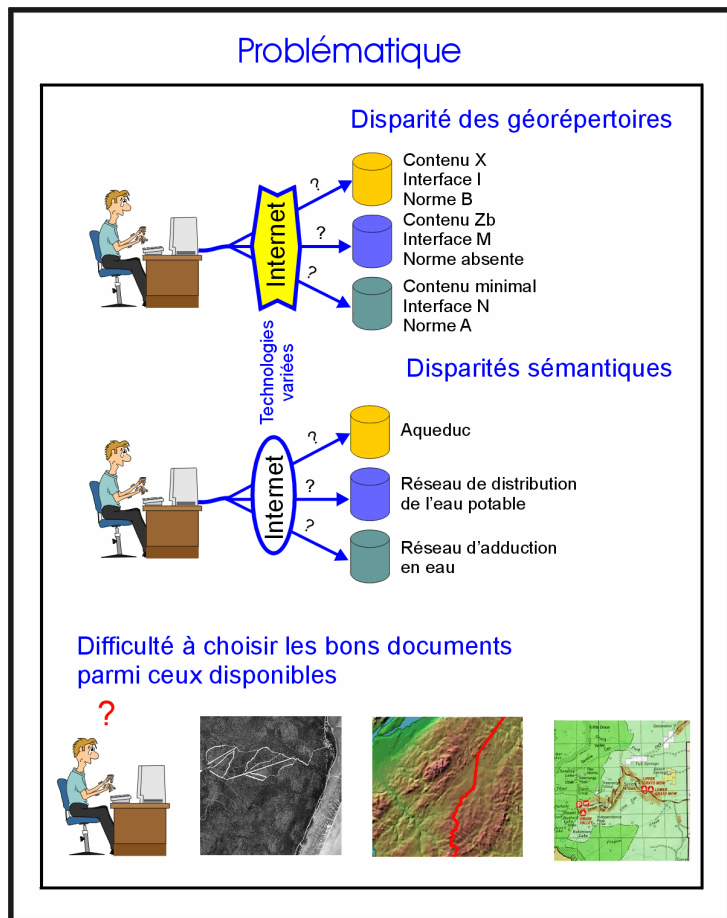
### **2.1. Cadre de développement des géorépertoires sur Internet**

Les premiers géorépertoires qui ont vu le jour sur le Web, au début de 1994, étaient de très simples factures. En effet, les technologies disponibles et le niveau d'évolution du langage HTML ne permettaient pas l'implantation de fonctions de recherche avancées. Ainsi, les quelques géorépertoires disponibles ne se contentaient-ils que de fournir à l'utilisateur une liste statique des documents disponibles, accompagnés de quelques métadonnées. Rapidement toutefois, avec l'apparition du langage Java et le développement des bases de données relationnelles en ligne, les géorépertoires se sont diversifiés et ne présentent plus la même image qu'il y a 4 ans. De plus, l'utilisation d'engins cartographiques dynamiques permet désormais de faire certaines opérations d'analyse spatiale via Internet. Les géorépertoires, de par leur structure, leur implantation et leur contenu sont devenus des applications hétérogènes malgré l'introduction de normes de métadonnées dans la communauté géomatique [FGDC, CGSB et ANZLIC]. Voyons plus en détail la nature de cette hétérogénéité.

Au cours des années 1995, 1996 et 1997, notre groupe de recherche a effectué une série d'études [PRO 96 et LET 98] visant à détailler la nature et la structure des géorépertoires sur Internet. Ces études ont été réalisées parallèlement au développement d'un géorépertoire au CRG (GÉOREP 1.0, <http://sirs.scg.ulaval.ca>) visant à explorer et étendre certains concepts

propres à cette technologie. Il ressort de ces études les points suivants (représentés partiellement à la figure 2) :

1 La majorité des géorépertoires ne font référence qu'à des documents cartographiques. Peu d'entre eux réfèrent à l'ensemble des documents à référence spatiale disponibles sur le marché (images satellites, photos aériennes, modèles numériques de terrain, documents descriptifs, relevés géodésiques, etc.). L'utilisateur doit donc consulter plusieurs sites différents s'il veut obtenir toute l'information spatiale disponible pour un territoire particulier.



2 Une grande hétérogénéité existe dans les géorépertoires en ce qui a trait au nombre de métadonnées disponibles pour les documents à référence spatiale. En 1997, selon notre évaluation, 20 % des géorépertoires ne faisait qu'afficher la liste du nom des documents disponibles, 24 % affichait cette même liste mais en y ajoutant la référence spatiale (coordonnées, zone, territoire, code de feuillet), 32 % des géorépertoires affichaient l'équivalent d'une norme minimale de métadonnées tandis que 24 % affichaient l'ensemble des métadonnées d'une norme. Parmi les normes de métadonnées utilisées, la norme américaine FGDC arrive au premier rang. Les normes canadienne CGSB et australienne ANZLIC n'étaient utilisées chacune que par un site en 1997.

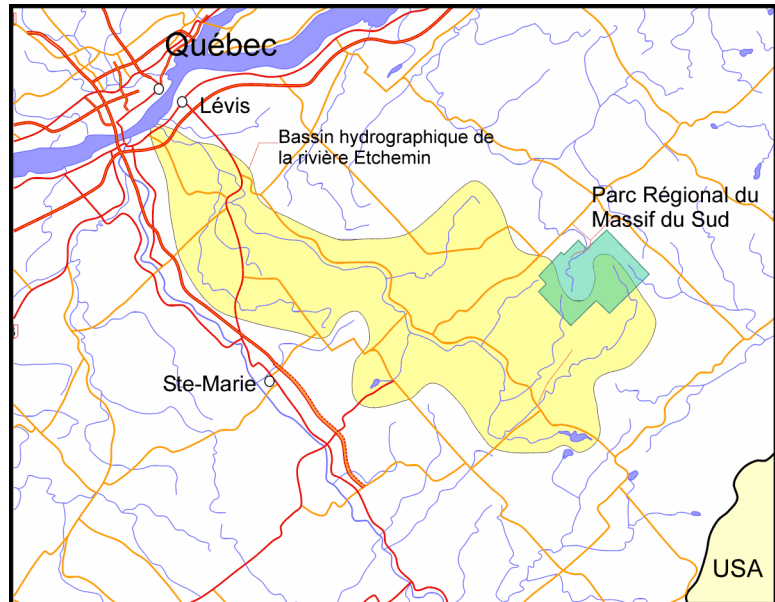
3 L'utilisation de bases de données relationnelles en ligne a permis une plus grande souplesse pour la recherche d'informations par l'utilisateur. Toutefois, ce ne sont pas tous les géorépertoires qui les utilisent. Ainsi, en 1997, nous évaluons que près de la moitié des géorépertoires sur Internet n'utilisaient pas de bases de données relationnelles en ligne. Les données étaient simplement stockées dans des pages Web statiques.

- 4 La plupart des géorépertoires permettent de visualiser soit un extrait ou soit la totalité d'un document, généralement sous forme d'images GIF ou JPEG. Toutefois, un nombre plus restreint affichent la couverture territoriale des documents ou des collections qu'ils référencent.
- 5 Les référentiels spatiaux utilisés par les géorépertoires sont très différents. Les systèmes de coordonnées, la forme des documents géoréférencés et les cartes de fond varient en fonction des territoires couverts, des documents disponibles et des systèmes de référence géodésique en vigueur.
- 6 Un certain nombre de géorépertoires permettent de télécharger les fichiers graphiques qu'ils référencent, particulièrement aux États-Unis où l'accessibilité aux données spatiales gouvernementales est très grande. Toutefois, peu de ces sites permettent l'importation des fichiers de métadonnées accompagnant les fichiers graphiques.
- 7 On constate un certain nombre de disparités sémantiques entre les différents géorépertoires. Ainsi, dans un premier géorépertoire, le champ *titre* peut correspondre à *nom officiel* dans un autre géorépertoire. Des différences fondamentales sont susceptibles d'être identifiées. Par exemple, dans un premier géorépertoire, une carte du réseau hydrique peut représenter l'ensemble des éléments hydriques d'un territoire (rivières, lacs, étangs, marais, ruisseaux), tandis que dans un autre géorépertoire, un autre document portant le même nom ne représentera que les lacs et cours d'eau du réseau hydrographique supérieur. Ces différences compliquent la tâche de l'utilisateur lorsqu'il recherche des documents dans différents géorépertoires.
- 8 Finalement, la variété des technologies utilisées dans les géorépertoires sur Internet force l'utilisateur à employer les deux principaux navigateurs Web s'il désire exploiter à fond l'ensemble des propriétés des sites. Les différentes versions du langage HTML utilisées par Netscape et Microsoft, la multiplication des plugiciels (*plug-ins*), Java et Active-X sont les principales causes de l'hétérogénéité technologique des géorépertoires sur le Web.

Ainsi, les géorépertoires sur Internet diffèrent beaucoup entre eux, que ce soit au niveau du contenu, de la structure, de la sémantique ou de la technologie. À la prochaine sous-section, nous démontrerons l'impact de cette hétérogénéité pour l'utilisateur.

## 2.2. Problématique spécifique liée à la consultation de plusieurs géorépertoires sur Internet

Lorsqu'un usager doit consulter plusieurs géorépertoires pour un territoire défini, il est confronté à cette hétérogénéité décrite ci-avant (figure 2). Prenons l'exemple d'un organisme désirant faire une étude complète sur l'aménagement et la conservation d'un bassin versant d'une importante rivière. Pour les fins de l'exemple, nous choisissons la rivière Etchemin, s'écoulant



des hauteurs des Appalaches jusqu'au Fleuve Saint-Laurent, face à Québec (figure 3). Pour les fins de l'étude, ces aménagistes auront besoin d'un ensemble diversifié de documents à référence spatiale (cartes topographiques, modèles numériques de terrain, images satellites, photos aériennes, rapports environnementaux, cartes éco-forestières et agricoles, pour ne nommer que ces derniers). Pour obtenir cette information, ils devront interroger entre autres les sites du ministère des Ressources naturelles du Québec, de Géomatique Canada, du ministère de l'Environnement et de la Faune et du consortium SPOT, Landsat ou RadarSat pour obtenir les images de télédétection. Chacun de ces sites est susceptible d'utiliser des contenus différents, selon des normes différentes, avec des interfaces différentes, etc. Ainsi l'utilisateur devra se familiariser avec chaque géorépertoire. Après avoir obtenu les informations désirées, il devra comparer celles-ci afin de choisir les jeux de données les plus appropriés pour ses besoins, ceci malgré que ces résultats soient également très hétérogènes. C'est donc un exercice fastidieux...

Face à ce problème d'hétérogénéité, deux types de solutions sont envisageables. La première solution vise à régler une partie du problème d'hétérogénéité en amont à l'aide de normes, tandis que la seconde le traite en aval. La prochaine sous-section traitera de ces possibilités complémentaires que l'on a avantage à combiner.

## 2.3. Deux avenues de solution possibles : régler le problème en amont ou en aval.

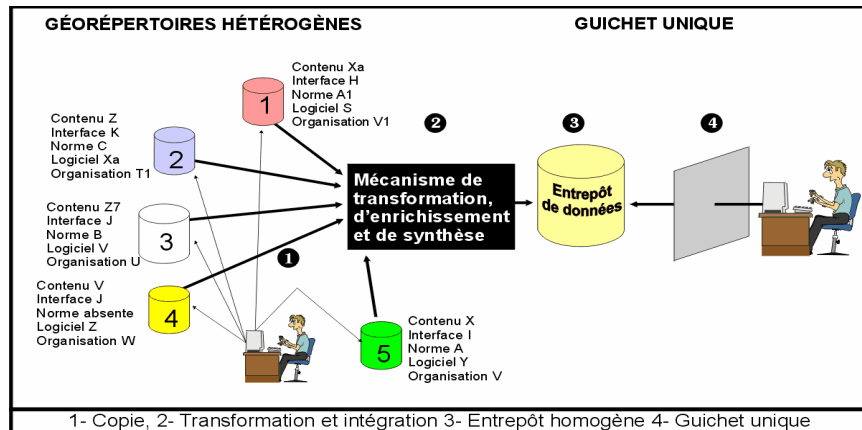
L'utilisation grandissante de normes de métadonnées (ex. FGDC, CSGB, ANZLIC) et de formats d'échange de données (ex. DIGEST, SAIF) permet d'amoindrir certains problèmes d'hétérogénéité pour les années futures. C'est la solution « amont » : minimiser l'hétérogénéité avant que l'utilisateur n'accède aux géorépertoires. Toutefois, tel que le mentionnaient Gouin et al [GOI 97], des milliards de dollars ont été investis dans des jeux de

données avec format propriétaire et il nous apparaît incertain que ces jeux de données soient remplacés par des données normalisées. De plus, une normalisation des données ne pourrait résoudre tous les problèmes (ex. interface graphique, mode d'interrogation des données, visualisation des documents, hétérogénéité sémantique...).

Il faut donc rechercher une solution qui permettrait de résoudre les problèmes qui sont illustrés à la figure 2. La solution que nous envisageons consisterait à intégrer un sous-ensemble du contenu hétérogène des géorépertoires sur le Web dans une base uniforme et homogène. Ici, par intégration, nous n'entendons pas centralisation, mais plutôt intégration en une base unique de copies de certaines données sources, ce qui laisse une entière autonomie aux géorépertoires sources existants.

Face à cet objectif, il nous est apparu que l'utilisation d'un entrepôt de données (ainsi que l'utilisation des agents logiciels dans nos autres travaux) permettrait de solutionner une bonne part des problèmes identifiés précédemment. En effet, ce nouveau type d'architecture de système semble tout à fait approprié pour intégrer des données provenant de plusieurs sources. Selon Inmon

[INM 96], « les entrepôts de données offrent des fonctionnalités d'intégration dans un monde de systèmes d'applications non intégrées » (ex. : les géorépertoires sur Internet). Ainsi, l'entrepôt de données permettrait d'intégrer



le contenu hétérogène des géorépertoires sur le Web dans une base uniforme et homogène. La figure 4 présente l'architecture générique du système que nous proposons.

Cette architecture nous semble appropriée pour solutionner plusieurs problèmes identifiés précédemment. Ainsi, les problèmes liés à l'hétérogénéité des géorépertoires ne seraient plus apparents à l'utilisateur car les données seraient copiées et transformées, depuis les géorépertoires sources vers l'entrepôt de données, où les données seraient stockées dans une base uniforme et homogène. Cette base serait constituée d'un sous-ensemble de métadonnées correspondant à l'équivalent d'une norme minimale de métadonnées. L'utilisation d'une base uniforme et homogène permettrait aussi de régler certains problèmes de disparités sémantiques, notamment celles liées à la structure des géorépertoires. Notons finalement qu'en accédant à une seule interface graphique pour trouver l'information, l'utilisateur n'est plus confronté à l'hétérogénéité des interfaces des géorépertoires. Il a toutefois le loisir d'accéder aux données sources via un hyperlien.

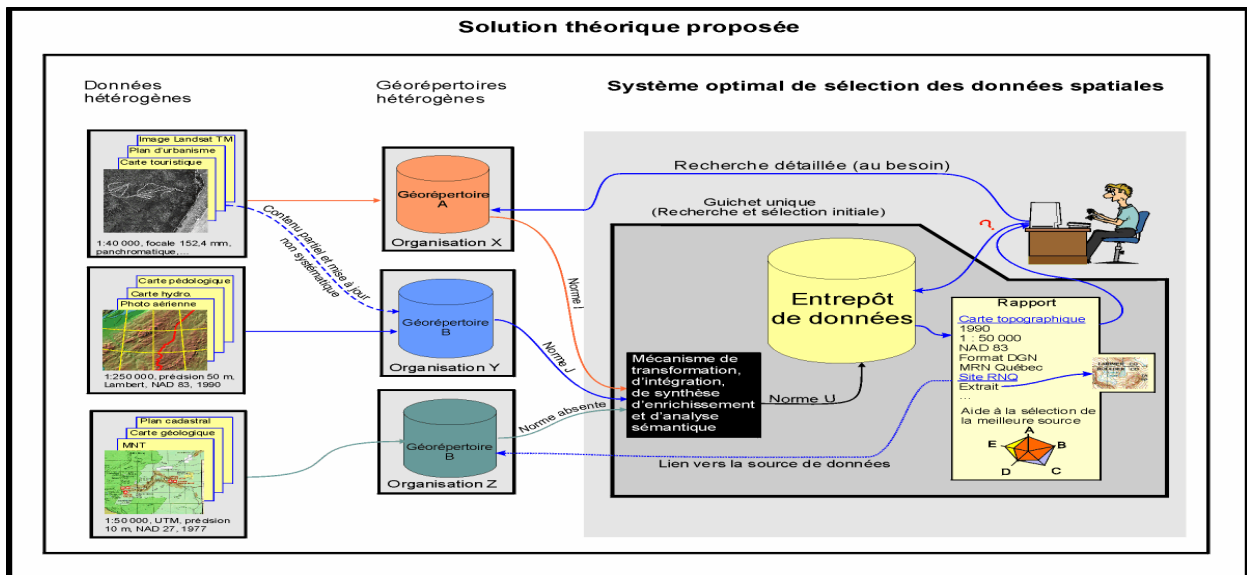
Bref, une telle architecture permettrait à l'utilisateur de consulter les métadonnées dans une base uniforme et homogène et lui fournirait des outils performants pour définir ses besoins et



choisir les meilleurs documents. Cette solution s'apparente au principe du guichet unique où l'utilisateur peut obtenir à un seul endroit à peu près toute l'information dont il a besoin pour un sujet particulier. La figure 5 synthétise ces propos.

La problématique générale abordée dans cette recherche se présente ainsi : quoique la solution de l'entrepôt de données proposée ici semble être une façon efficace d'atteindre l'objectif du guichet unique, la difficulté à laquelle nous faisons face au début des travaux est qu'une telle architecture n'avait pas encore été testée sur le Web pour les données spatiales (et sûrement pas pour les géorépertoires) et que conséquemment, nous n'en connaissons pas la faisabilité conceptuelle, opérationnelle (ex.: transformation, enrichissement et intégration des métadonnées) et technique (ex.: couplage d'un entrepôt de données et d'un site Web).

La prochaine section vise à présenter plus en détail les concepts relatifs à l'entrepôt de données. L'emphase sera mise sur l'architecture de l'entrepôt de données, sur son mécanisme



d'intégration des données et sur l'extension du concept pour la gestion des données spatiales.

### 3. Présentation du concept d'entrepôt de données

L'entrepôt de données (en anglais *Data Warehouse*) est un concept mis de l'avant au milieu de la décennie 80 par Barry Devlin [DEV 97]. On identifie aussi William Inmon [INM 96] comme le père de l'entrepôt de données. Terme en vogue à la fin de cette décennie, l'entrepôt de données mérite une définition rigoureuse, afin de bien cerner la nature de ce concept.

### 3.1. Définition de l'entrepôt de données

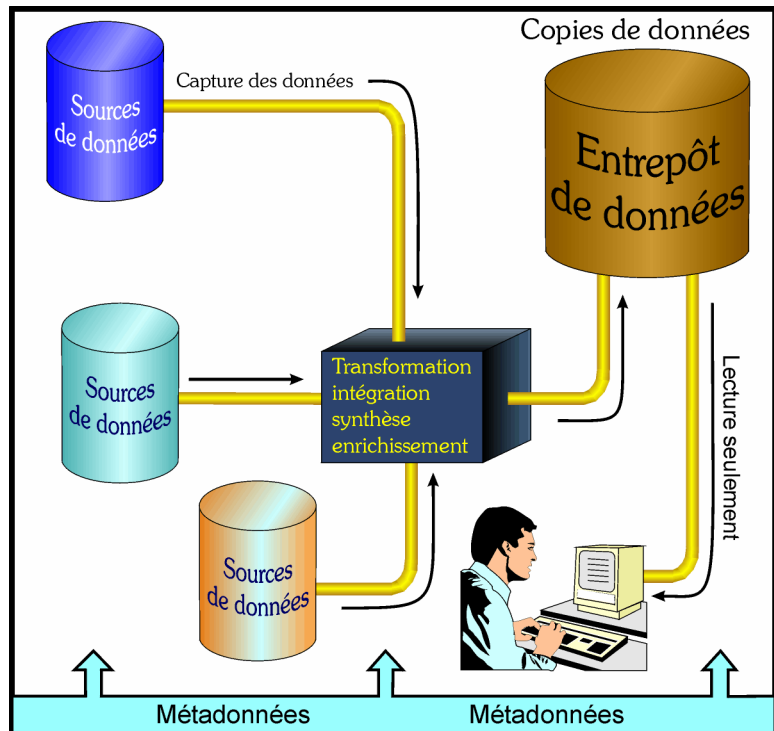
Le terme « entrepôt de données » est certainement l'un des plus galvaudés dans le monde de l'informatique, du moins, en cette fin de la décennie 90. Un grand nombre de définitions ont été proposées et leur application concrète varie grandement. De plus, plusieurs définitions sont biaisées par l'orientation qu'un auteur souhaite donner au concept de l'entrepôt de données afin de favoriser telle ou telle implantation d'entrepôt de données (il arrive même fréquemment qu'il travaille pour un fournisseur de solution). Des définitions sont présentées dans cette section. La première définition est celle de William Inmon. Selon Inmon [INM96], « *a data warehouse is a subject-oriented, integrated, time variant and non-volatile collection of data in support of management's decisions* ». C'est une définition classique que l'on rencontre fréquemment dans les articles et ouvrages traitant des entrepôts de données (cf. Moriarty et al., [MOR 96]; White [WHI 96] et Crandall [CRA 95]. De son côté, Poe [POE 95], souligne le rôle qu'ont les entrepôts de données pour gérer de grands volumes de données.

Devlin [DEV 97] définit l'entrepôt de données ainsi : « *a data warehouse is a single, complete, and consistent store of data obtained from a variety of sources and made available to end users in a way they can understand and use in a business context* ». Cette définition s'apparente à l'objectif qui est poursuivi pour ce projet, où nous cherchons à développer un guichet unique permettant à l'utilisateur de géorépertoires de trouver toute l'information dont il a besoin. C'est d'ailleurs la définition que nous avons retenue. Bien que ce projet ne vise que l'intégration de métadonnées spatiales et non pas l'intégration des données spatiales elles-mêmes, il est opportun de présenter une définition d'entrepôt de données spatiales. Un tel type d'entrepôt de données constitue une solution logique de ce projet. Rawling et al [RAW 97] proposent la définition suivante pour l'entrepôt de données spatiales : « *a geospatial data warehouse is a collection of subject oriented spatial data that is of known quality, is non-volatile, is time variant and includes the basic tools to access and extract information* ». Cette définition est en concordance avec plusieurs aspects de la définition de Inmon, sans que l'on fasse mention toutefois de l'aspect support aux gestionnaires.

Plusieurs auteurs insistent sur l'importance du mécanisme d'intégration des données d'un entrepôt de données. À juste titre d'ailleurs, car l'expérience nous a démontré que ce mécanisme constitue la pierre angulaire de tout entrepôt de données. Dans le cas qui nous intéresse, ce mécanisme nous fournit toutes les fonctionnalités dont nous avons besoin pour constituer la base uniforme et homogène constituée des données provenant des sources hétérogènes.

### 3.2. Architecture générique d'un entrepôt de données

Tout comme il y a un grand nombre de définitions pour l'entrepôt de données, les auteurs en proposent un grand nombre d'architectures. Parmi elles, il est quand même possible d'identifier une base commune. Cette base commune, c'est ce que l'on identifie comme l'architecture générique d'un entrepôt de données (voir figure 6). Celle-ci comprend cinq composants distincts, soit les sources de données, la transformation des données, l'entrepôt de données et l'accès aux données. Le cinquième



élément est constitué des métadonnées, lesquelles sont la fondation du système. Il faut bien noter ici que l'entrepôt de données ne constitue pas une base de données centralisée. L'architecture que nous proposons permet de laisser intactes les données sources, contrairement à ce qui arrive avec une base de données centralisée où tous les systèmes sources sont remplacés par un seul système avec toutes les données. Voyons rapidement en détail les caractéristiques de chacun des composants.

- 1- Les sources de données. Dans le contexte des entrepôts de données, les sources de données sont généralement hétérogènes et se retrouvent sur de multiples plates-formes. Elles sont généralement identifiées comme les *data sources*, ou les *legacy systems*. L'entrepôt de données contient des copies exactes ou modifiées des sources de données, ou des résultats de calculs.
- 2- Le deuxième élément de l'entrepôt de données (qui constitue en fait une étape charnière dans l'implantation d'un entrepôt de données) comprend toutes les opérations nécessaires à la saisie, la transformation, la synthèse et l'enrichissement des données, que l'on désigne généralement par « opérations de transfert de données ». Cet ensemble d'opérations est représenté par la boîte noire de la figure 6. Ces opérations sont certainement les plus importantes et les plus complexes à effectuer lors de l'implantation d'un entrepôt de données. À ce titre, Devlin [DEV 97] indique que « cette étape, identifiée comme le chargement des données dans l'entrepôt, est reconnue comme étant l'aspect technique le plus difficile à accomplir lors de l'implantation d'un entrepôt de données » (traduction libre). D'autre part, un sondage

paru dans l'édition de janvier 1997 de la revue Byte indique que 85% des répondants considèrent que le nettoyage des données est une tâche difficile. Sans l'ombre d'un doute, il convient d'accorder une attention particulière à cette étape de transformation des données, particulièrement pour les données spatiales.

- 3- Le troisième ensemble constitue le système de gestion de la base de données de l'entrepôt de données. C'est à cet endroit que résident physiquement les données copiées et transformées. Habituellement, le SGBD supportant l'entrepôt de données est optimisé pour traiter de grands volumes de données et en permettre l'accès à un grand nombre d'utilisateurs.
- 4- Le quatrième ensemble est constitué des outils de requête et de consultation des données. L'un des principaux buts d'un entrepôt de données est de permettre à une gamme variée d'utilisateurs d'obtenir rapidement de l'information qui est difficilement accessible de façon rapide dans les systèmes opérationnels. Pour ce faire, il existe un éventail d'outils permettant d'accéder aux données de l'entrepôt de données. Parmi ces outils, mentionnons le marché de données (*data mart*), les outils OLAP et finalement, les outils de requête et de génération de rapports.
- 5- Les métadonnées sont présentes à tous les niveaux de l'entrepôt de données, de la capture à la consultation des données. De manière générale, les métadonnées permettent de mieux connaître les données qui sont capturées, chargées, consultées, ainsi que les règles de transformation et de sélection. Selon Moriarty et Greenwood [MOR 97], « les métadonnées sont aussi essentielles aux usagers que ne le sont les données elles-mêmes ». Il faut mentionner toutefois que les métadonnées ne sont pas toujours disponibles et bien des usagers ne sont pas intéressés à consentir les efforts nécessaires à leur bonne utilisation (Cohn [COH 96]). L'usage des métadonnées est particulièrement important dans la phase de transfert des données. Selon Bédard et al. [BÉD 97], « plusieurs outils, tels que les outils de nettoyage, d'intégration, d'extraction ne peuvent fonctionner sans avoir recours aux métadonnées. »

Il est bien important de saisir l'importance que revêt l'outil de transformation des données pour l'entrepôt de données. C'est ce dernier qui est responsable de la qualité générale de l'entrepôt.

### **3.3. Extension du concept d'entrepôt de données pour les données spatiales**

Traditionnellement, l'intégration de données descriptives hétérogènes dans un entrepôt de données nécessite l'utilisation de méthodes particulières de transformation des données. Ainsi, selon les besoins, l'outil segmentera, fusionnera, synthétisera ou enrichira les données préalablement à leur intégration dans l'entrepôt. Ces opérations sont relativement complexes et nécessitent une planification très minutieuse des opérations de transformation.

L'ajout d'une dimension spatiale à l'intégration des données complexifie grandement l'opération. Il existe un intérêt manifeste à construire des entrepôts de données spatiales. En effet, il est généralement admis que 70 à 80 % des efforts de mise en place d'un SIG sont

reliés à l'acquisition et à l'intégration des données. Il y a souvent de nombreux problèmes d'intégration de données qui surviennent à cette étape : projections différentes, datum différents, format de données différents, normes de métadonnées différentes, etc. Ces problèmes sont de la même nature que ceux qui ont mené à l'émergence des entrepôts de données dans les applications de gestion et de finance. Il serait donc intéressant d'extensionner les concepts existants pour les adapter au contexte bien précis des données à référence spatiale, permettant ainsi de réduire les efforts consacrés à l'intégration des données dans un SIG.

Les problèmes d'intégration géométrique des données mentionnés précédemment sont bien connus et plusieurs solutions existent présentement pour y remédier partiellement. Les problèmes liés à l'intégration géométrique des données sont en voie d'être réglés. Outre les traducteurs de fichiers retrouvés normalement avec les SIG, des outils tels que FME (<http://www.safe.com>) ajoutent de nombreuses fonctions de manipulation de données et les nouvelles solutions telle que l'OGDI (<http://www.las.com>) et les GDO de Geomedia (<http://www.intergraph.com>) appliquent les concepts d'interopérabilité (voir <http://www.ogis.com> pour plus de détails). Ces solutions sont fort utiles pour la mise en place d'un entrepôt de données car elles permettent d'intégrer des données hétérogènes provenant de plusieurs sources différentes.

Néanmoins, ces solutions sont encore incomplètes, car elles ne tiennent pas compte de nombreux paramètres définis dans les métadonnées. Par exemple on ne renseigne pas l'utilisateur sur « l'intégrabilité » des différents jeux de données. On peut avoir la possibilité technique d'intégrer des données provenant de sources différentes, mais il est parfois préférable de ne pas le faire. Par exemple, il est facile d'intégrer deux sources de données différentes avec les outils existants, mais si la précision des deux jeux de données sources est différente, la précision du jeu de données résultant ne pourra pas être quantifiée uniformément. Il n'apparaît pas logique d'intégrer des données produites à l'échelle 1 : 50 000 dans un jeu de données 1 : 1 000 000. Des problèmes de précision et d'ajustement se poseraient inévitablement. Des problèmes de généralisation des données sont aussi susceptibles de se manifester. En effet, comment gérer des données détaillées à des petites échelles sans modifier légèrement leur géométrie (par exemple, le cas d'une route longeant une rivière, que l'on doit déplacer aux petites échelles de manière à ce que ces deux entités ne se chevauchent pas). Il n'existe pour le moment aucun moyen efficace pour réaliser ces opérations automatiquement. Il devient donc avantageux d'utiliser un entrepôt de données pour solutionner ces problèmes comparativement à l'approche d'interopérabilité. En effet, les résultats de ces opérations manuelles non déductibles informatiquement sont emmagasinés dans l'entrepôt de données une fois pour toute et n'ont pas à être répétées. Par opposition, les solutions faisant appel à l'interopérabilité des systèmes et donc à la conversion à la volée des données ne permet pas de solutionner ce genre problème de manière efficace car la partie non déductible doit continuellement être réeffectuée par l'utilisateur. Toutefois, nous entrevoyons que la combinaison d'un entrepôt de données avec les outils d'interopérabilité (ainsi que des normes de métadonnées et d'échange de données) ouvre la voie à la meilleure solution.

La prochaine section mettra l'accent sur l'utilisation de ce concept dans le contexte des géorépertoires hétérogènes sur Internet. Pour ce faire, nous présenterons le prototype que

nous avons développé afin de tester et de valider cette approche.

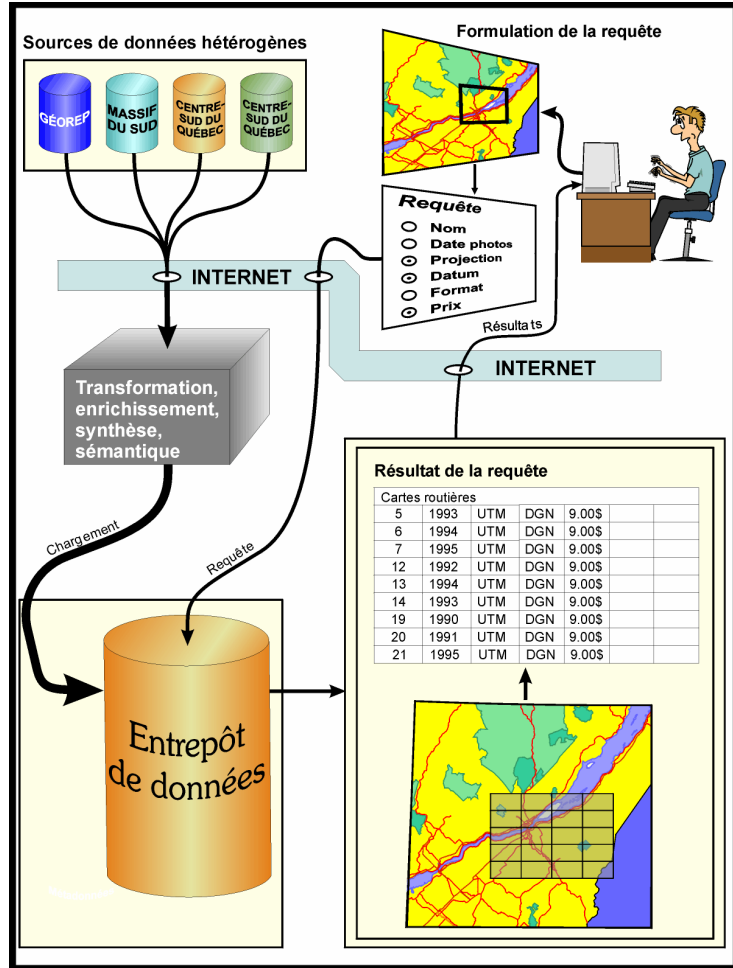
#### **4. Solution envisagée : Système optimal de sélection des données spatiales (SOSDS)**

L'un des objectifs du projet de recherche était de valider le concept d'entrepôt de données pour l'intégration des géorépertoires hétérogènes accessibles via Internet. Pour ce faire, nous avons développé le prototype SOSDS en 5 mois. Ce développement s'est fait avec des technologies simples à implanter. Ce prototype, comme nous le verrons au cours de cette section, nous a également permis d'expérimenter de nouvelles techniques de requêtes de données utilisant la logique floue. Cette possibilité est à la base même de notre mode de sélection des meilleurs jeux de données, *i.e.* ceux rencontrant le plus les besoins de l'utilisateur. Le prototype est accessible via Internet à l'adresse suivante : <http://sosds.scg.ulaval.ca>. Les prochaines pages en présentent les principaux éléments, soient son modèle général de fonctionnement, son lien avec les systèmes sources ainsi que sa structure de données interne, son interface cartographique et finalement son interface d'aide à la sélection des meilleurs jeux de données.

##### **4.1. Fonctionnement général du système SOSDS**

Afin de bien saisir le concept du prototype SOSDS, nous présenterons rapidement l'architecture choisie. La figure 7 contient un diagramme qui présente les principes généraux de l'architecture du prototype SOSDS. Nous retrouvons donc au coin supérieur gauche les sources de données, au nombre de quatre. Ces sources de données sont quatre géorépertoires que nous avons construits sur le Web et qui simulent l'hétérogénéité typiquement rencontrée (voir prochaine section). Par la suite, les données sont copiées des sources vers l'outil de transformation des données, où elles subissent toutes les transformations nécessaires à leur intégration et enrichissement dans l'entrepôt de données. Lorsque toutes les opérations de transformation des données sont complétées, les données produites sont intégrées dans la base de données de l'entrepôt.

Dès lors, l'utilisateur pourra interroger le contenu de l'entrepôt. Cette interrogation est faite en dessinant sur une carte de fond définie par l'utilisateur le territoire correspondant à l'aire de recherche. Ce territoire peut être un point, une ligne, une ellipse ou un polygone. Une fois ce territoire défini, l'utilisateur choisit les métadonnées qui feront partie de la requête et définit les paramètres de chacune de celles-ci. Par la suite, la requête est transmise à l'entrepôt, elle y est traitée par le système de gestion de base de données (SGBD), formatée par l'engin requêteurs et les résultats sont retransmis à l'utilisateur, sous forme de tableaux. Les résultats sont traités de manière à faciliter leur analyse par l'utilisateur. Finalement, ce dernier a le loisir de visualiser la couverture territoriale des



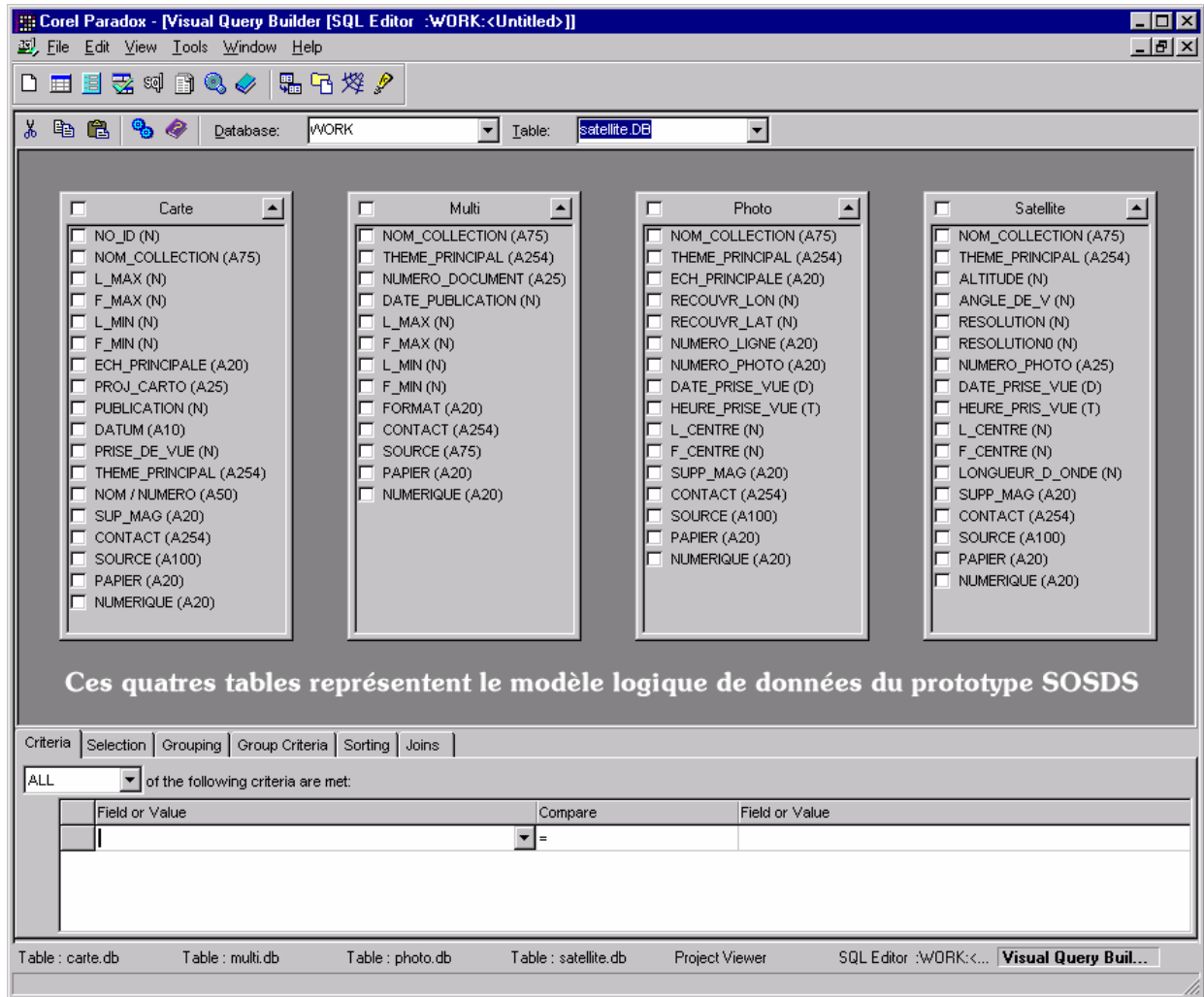
documents retournés par la requête, soit par feuillet ou par collection. Il faut noter que le système est conçu de manière à ce que l'utilisateur puisse retrouver la source initiale de données. Ainsi, s'il désire obtenir plus d'information sur un document particulier, il n'a qu'à suivre le lien vers la source de données initiale.

Voyons maintenant plus en détail certains des éléments que nous venons de présenter.

## 4.2. Sources de données et intégration

Les sources de données ont été à l'origine d'un des principaux problèmes auquel nous avons dû faire face. En effet, bien que les données des géorépertoires sur Internet soient disponibles pour consultation, il n'est pas nécessairement possible ou permis (en raison des droits d'auteur) de les copier dans un autre site. C'est pourquoi nous avons décidé de simuler les sources de données. Pour ce faire, nous avons conçu quatre géorépertoires représentatifs de la réalité. Chacun de ces géorépertoires offre une implantation distincte : le contenu, l'interface, le mode d'accès aux données sont tous différents. Le territoire couvert par ces géorépertoires s'étend de 45,75° N à 48,00° N et de 73,00° O à 69,50° O.

La constitution de l'entrepôt de données s'est faite en transformant ces données provenant des géorépertoires sources dans la base que nous avons définie pour l'entrepôt de données. Cette base comprend la trentaine de métadonnées que nous jugeons les plus utiles pour un usager non-spécialiste (voir figure 8). La transformation des données s'est faite



manuellement pour des raisons académiques et des contraintes de temps et de logiciels.

Cette approche nous a permis d'identifier les principaux problèmes susceptibles de se produire lors de la transformation des données. Parmi ceux-ci, notons :

- 1- Transformation des coordonnées géographiques (ex. : degré-minute-décimale en degré-décimal ou coordonnées UTM en longitude/latitude);
- 2- Fusion ou segmentation de champs de tables de bases de données (ex. : le numéro d'une photo aérienne dans l'entrepôt qui est formé du numéro de ligne de vol et du numéro de la photo provenant du système source);



- 3- Enrichissement des données : création de nouvelles données à partir des données existantes (ex. : position d'un feuillet générée à partir du code d'identification du feuillet);
- 4- Recodage des attributs, uniformisation (ex. : modifier le contenu des champs pour les rendre uniformes dans l'entrepôt de données);
- 5- Sélection des métadonnées (ex. : ne sélectionner que les métadonnées qui doivent être intégrées dans l'entrepôt).

Concrètement, la transformation s'est faite en intégrant dans le tableur Quattro Pro 8 l'ensemble des données disponibles dans nos géorépertoires sources. Nous y avons fait les opérations de transformation, d'enrichissement et de synthèse des données (en utilisant parfois des formules contenues dans le tableur). Il en est résulté une base de données de plus de 800 enregistrements qui furent ensuite intégrés dans le SGBD Corel Paradox 8. Nous avons opté pour ce SGBD en raison de sa simplicité (les tables se manipulent aisément dans un tableur), de son support pour l'ODBC et de ses capacités à être utilisé comme serveur de données dans une application Web. La figure 8 présente le modèle logique de données de l'entrepôt.

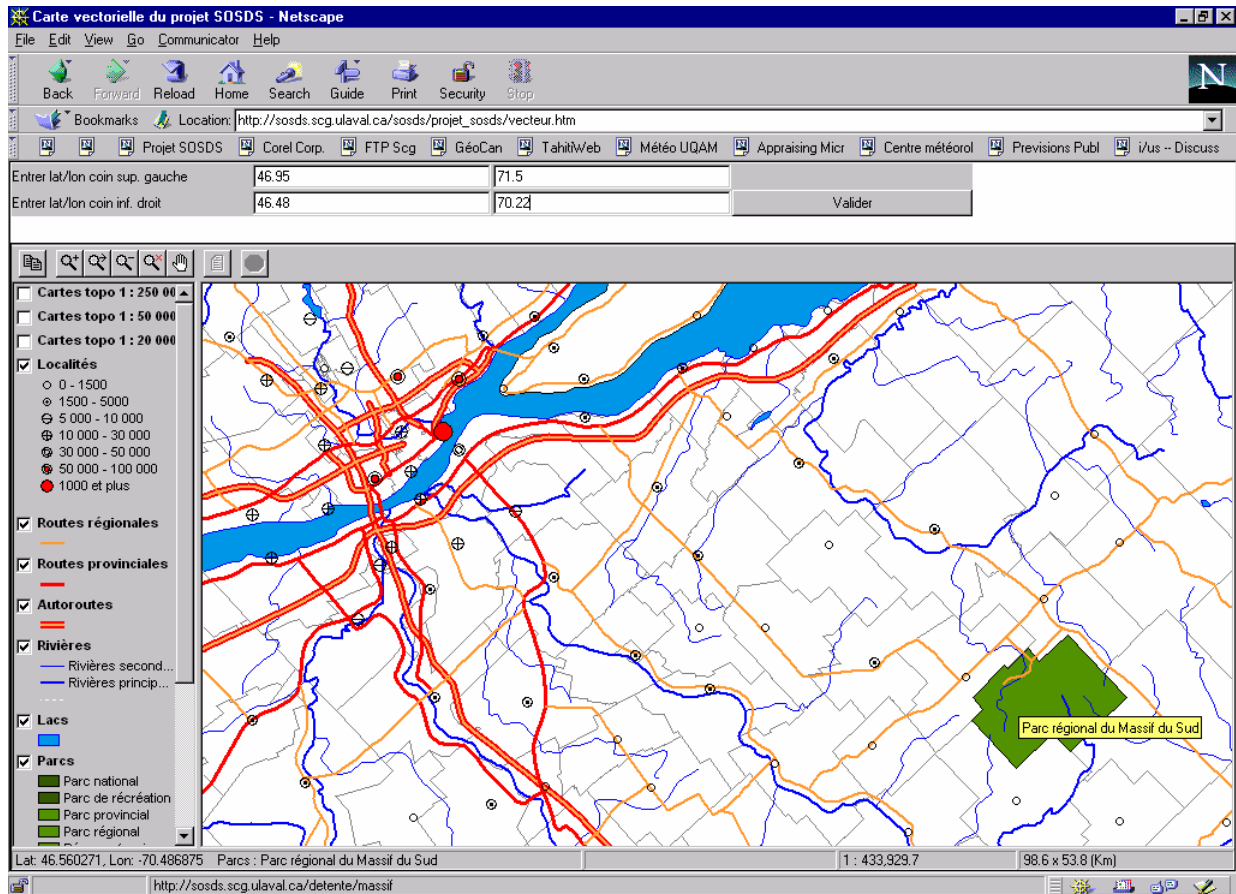
Il faut noter ici que les tables ne sont pas reliées, ce qui correspond à la gestion de fichiers plats indépendants plutôt que d'une vraie base de données. Cette situation s'explique par le choix de notre outil d'interrogation des données qui ne supporte pas les joints relationnels. Nous y reviendrons à la section 5.4. Cette conception de l'entrepôt de données nous a permis de construire l'interface d'interrogation des données. Cette dernière est formée de deux parties, soit l'engin cartographique dynamique et l'engin de requête en logique floue. La prochaine sous-section traitera de l'interface graphique du système, particulièrement de l'engin cartographique dynamique.

### **4.3. Interface avec un engin cartographique dynamique**

L'un des buts premiers du système SOSDS est d'être accessible au plus grand nombre d'utilisateurs. Le système doit être en mesure de répondre aux besoins d'une gamme variée d'utilisateurs allant de l'utilisateur recherchant une carte topographique avoisinant son chalet près du lac Pas d'Eau au gestionnaire de projet qui recherche un ensemble de cartes et de documents de nature spatiale pour un territoire particulier. Pour ce faire, le système doit donc être en mesure de permettre à l'utilisateur de se situer sur le territoire, parfois dans un contexte bien particulier. Ainsi, l'utilisateur peut construire lui-même sa carte de fond, avec les éléments cartographiques qui lui sont familiers (limites territoriales, hydrographie, zones forestières, géologiques, etc.).

L'engin cartographique utilisé transmet via Internet les cartes de fond sous forme vectorielle. Ces cartes sont accompagnées d'une légende les décrivant et d'un *tooltip* (petite barre jaune avec du texte) identifiant chacun des éléments sur lequel l'utilisateur pointe sa souris. L'engin

cartographique dynamique utilisé par SOSDS est MapGuide 2.5 d'Autodesk. La figure 9



présente l'interface cartographique avec la carte de fond.

Puisque les données sont sous forme vectorielle, il est possible de les utiliser dans un contexte dynamique. Ainsi, l'utilisateur peut effectuer des changements d'échelle, visualiser en temps réel les coordonnées liées au pointeur de la souris, calculer des distances, etc. La symbologie et l'affichage des éléments se fait en fonction de l'échelle. Parmi toutes ces fonctions, la plus importante demeure celle permettant à l'utilisateur de dessiner à l'écran la zone d'intérêt. Cette zone peut être un point, une ligne ou un polygone, dans leur forme régulière ou irrégulière. C'est à partir de cette zone de recherche bien définie que la requête sera effectuée. Voyons maintenant comment l'utilisateur interroge le contenu de l'entrepôt de données et comment les résultats lui sont présentés.

#### 4.4. Aide à la sélection des documents

L'approche traditionnelle lors de l'interrogation des données d'un entrepôt consiste à consulter le contenu d'une base de données pour vérifier si les documents recherchés s'y trouvent. Cette requête peut être faite manuellement si le volume de données est faible ou de façon automatique si le volume de données est grand. L'interrogation des bases de données

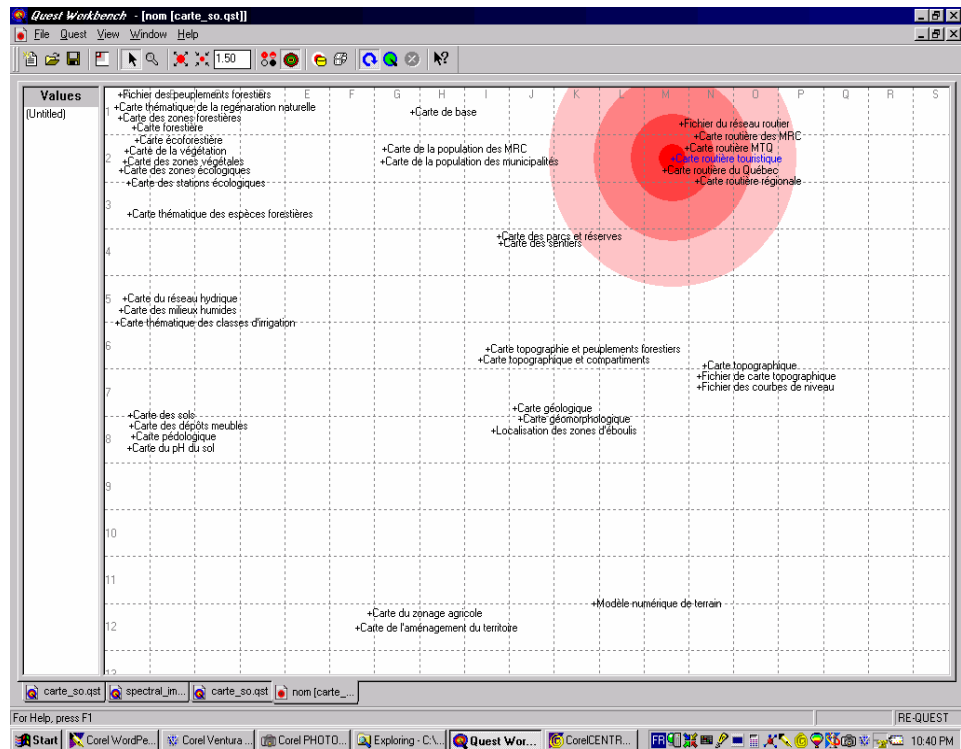
se fait généralement à l'aide du langage SQL, mais sans que l'utilisateur en soit conscient. Il suffit, par exemple, d'indiquer la valeur que l'on recherche pour chaque métadonnée en l'inscrivant dans les champs respectifs et l'engin de requête formule la requête à la base de données selon ces paramètres définis par l'utilisateur. Bref, l'interrogation des données est transparente pour l'utilisateur et il n'a pas à se soucier de la syntaxe du langage SQL. L'important pour ce dernier est d'obtenir les meilleurs documents possibles.

L'ennui avec cette méthode générale d'interrogation c'est que la recherche de documents se fait généralement de façon rigide, avec la logique booléenne (vrai ou faux). On trouve le document seulement si ce dernier répond parfaitement aux critères de sélection. Ainsi, si l'utilisateur recherche des photographies aériennes d'échelle 1:15 000 mais que la base de données ne contient que des photographies d'échelle 1:15 840 (ancienne échelle de 20 chaînes au pouce), la requête ne retournera pas ces documents qui pourtant pourraient très bien convenir à l'utilisateur. L'utilisateur ne peut connaître *a priori* le contenu d'une base de données (et ainsi indiquer l'échelle de 1:15 840), sauf si on lui présente la liste des valeurs, ce qui alourdirait indûment le système.

Face à cette situation, nous avons décidé d'utiliser un engin de requête utilisant la logique floue. Il s'agit de Quest 2.5, conçu par Inso. Ce logiciel permet d'interroger une base de données en fonction de paramètres flous définis par l'administrateur de la base de données. Reprenons notre exemple de photographie aérienne. Si l'administrateur définit un paramètre sur l'échelle d'un document, indiquant que les échelles comprises entre  $\pm 1 : 5\,000$  de la valeur cible recherchée sont considérées lors d'une requête, l'utilisateur obtiendra une réponse à sa requête s'il recherche une échelle 1 : 15 000 mais que la base de données ne contient que des échelles 1:15 840 (recherche dans l'intervalle de 1 : 10 000 à 1 : 20 000 pour une échelle cible de 1 : 15 000).

Il est possible de faire le même genre d'opération avec des données textuelles. Par exemple, nous

pouvons définir dans un espace bidimensionnel des relations de proximité entre les différents documents contenus dans l'entrepôt. Nous regroupons ensemble tous les documents appartenant à un même



thème (topographie, foresterie, cartes administratives, routières, etc.), comme le présente la figure 10. Ainsi, en définissant ces relations de proximité, une plus grande flexibilité est permise à l'engin de requête pour trouver des documents répondant le mieux aux besoins exprimés par l'utilisateur. Par exemple, si l'utilisateur recherche une carte géologique avec des critères bien précis (date de publication, échelle, datum, projection, etc.) mais que le système ne contient pas de cartes géologiques répondant parfaitement à ses besoins, il est possible, selon les documents disponibles dans l'entrepôt de données que le système suggère à l'utilisateur une carte géomorphologique dont les caractéristiques correspondent largement aux besoins de l'utilisateur. Cette sélection d'un autre type de document est possible en raison de la relation de proximité définie entre la carte géologique et la carte géomorphologique sur la carte des valeurs (voir figure 10).

L'interface d'interrogation est très simple (voir figure 11). L'utilisateur choisit des valeurs dans des listes déroulantes ou indique les valeurs qu'il recherche. Par la suite, il indique l'importance de chaque métadonnée dans sa requête, ce qui permet une pondération des requêtes. Cette pondération permet aussi de classer les documents en fonction de leur correspondance aux besoins de l'utilisateur. La figure 12 présente le résultat d'une requête. Il faut noter ici que de nombreux documents contenus dans l'entrepôt de données répondaient aux critères de l'utilisateur. Si l'on continuait dans la liste des résultats, des documents répondant moins aux critères de l'utilisateur apparaîtraient. Nous invitons le lecteur à visiter le site Internet du prototype SOSDS pour expérimenter cet engin de requête.

Rang / Numéros	Nom de la collection	lmax	fmax	lmin	fmin	Échelle princ.	Projection	Date public.	Datum	Date prise vue
99 1	Carte topographique	72.75	45.875	72.5	46	1:20 000	MTM	1995	Nad 83	1993
99 2	Carte topographique	72.5	45.875	72.25	46	1:20 000	MTM	1995	Nad 83	1993
99 3	Carte topographique	71	47.625	70.75	47.75	1:20 000	MTM	1995	Nad 83	1992
99 4	Carte topographique	70.75	47.25	70.5	47.375	1:20 000	MTM	1994	Nad 83	1992
99 5	Carte topographique	70.75	47.375	70.5	47.5	1:20 000	MTM	1994	Nad 83	1992
99 6	Carte topographique	70.5	47.5	70.25	47.625	1:20 000	MTM	1995	Nad 83	1992
99 7	Carte topographique	70.25	47.5	70	47.625	1:20 000	MTM	1995	Nad 83	1992
99 8	Carte topographique	70.5	47.625	70.25	47.75	1:20 000	MTM	1995	Nad 83	1992
99 9	Carte topographique	70.25	47.625	70	47.75	1:20 000	MTM	1995	Nad 83	1992
99 10	Carte topographique	72.75	46.75	72.5	46.875	1:20 000	MTM	1994	Nad 83	1992
99 11	Carte topographique	72	46.75	71.75	46.875	1:20 000	MTM	1993	Nad 83	1991
99 12	Carte topographique	71.75	46.75	71.5	46.875	1:20 000	MTM	1993	Nad 83	1991
99 13	Carte topographique	72	46.875	71.75	47	1:20 000	MTM	1993	Nad 83	1991
99 14	Carte topographique	71.75	46.875	71.5	47	1:20 000	MTM	1993	Nad 83	1991
99 15	Carte topographique	70.5	47	70.25	47.125	1:20 000	MTM	1994	Nad 83	1991

En résumé, l'utilisation d'un engin de requête avec logique floue offre beaucoup plus de souplesse que les requête SQL et le logiciel utilisé ne requiert aucunement de la part de l'utilisateur qu'il compose ses requêtes. Tout ce qui est requis de la part de l'utilisateur est que ce dernier choisisse les métadonnées de sa requête, indique les valeurs recherchées et spécifie l'importance de chaque métadonnée dans sa requête.

## 5. Conclusion

L'utilisation d'un entrepôt de données permettant de faire l'intégration de métadonnées provenant de géorépertoires hétérogènes sur Internet offre un potentiel très intéressant d'applicabilité. En effet, notre recherche a démontré que ce concept se prête particulièrement bien à la création d'un guichet unique camouflant l'hétérogénéité des géorépertoires. En effet, il contient toute l'information de base dont est susceptible d'avoir besoin une large gamme d'utilisateurs lors de la recherche de documents à référence spatiale. L'utilisateur n'aurait donc plus besoin de consulter plusieurs sites différents pour obtenir l'information dont il a besoin pour prendre une décision, elle se trouverait à sa portée dans le système SOSDS. Finalement, l'utilisation d'un engin cartographique dynamique et d'un engin de requête utilisant la logique floue permet de fournir toutes les fonctionnalités nécessaires à la formulation de requêtes efficaces permettant d'exploiter à fond tout le contenu de l'entrepôt de données. Nous avons donc pu, grâce à ce projet, évaluer pour la première fois l'efficacité des entrepôts de données et les requêteurs flous pour les géorépertoires sur le Web.

## 6. Bibliographie

[ANZLIC]

Norme ANZLIC : <http://www.anzlic.org.au/>

Bédard Y., Larrivée S., Proulx M.J., Caron P.-Y., Létourneau F., Lam S. [BÉD 97]  
1997 Geospatial Data Warehousing : positionnement technologique et stratégique. Centre de recherche en géomatique, Université Laval et Centre de recherches pour la Défense de Valcartier (CRDV). 79 p.

Bédard, Y., Proulx M.-J., Létourneau F., Moulin B., Maamar Z. 1997. [BÉD 97a]  
1997 Géorépertoires sur Internet: Outils puissants de démocratisation des données à référence spatiale. Géomatique VI: Un monde accessible. Association canadienne des sciences géomatiques, Montréal, 13 et 14 novembre 1997.

[CGSB]

Norme CGSB : [http://www.unb.ca/GGE/Research/GEG\\_GI\\_Standards/cgsb/](http://www.unb.ca/GGE/Research/GEG_GI_Standards/cgsb/)

Cohn M. [COH 96]

1996 One Pitfall of Data Mining : Data "mine"ing. Computer World, février 1996, vol. 30, number 9 p. 37

Crandall R. [CRA 95]

1995 OLAP in the Data Warehouse. Data Base Advisor, août 1995, pp 76 - 80

Devlin, B. [DEV 97]

1997 Data Warehouse from Architecture to Implementation. Addison Wesley Longman inc.432 p.

[DIGEST]

Norme d'échange de données DIGEST :

<http://www.j2geo.ndhq.dnd.ca/digest/html/DIGEST.HTM>

[FGDC]

Federal Geographic Data Committee. <http://www.fgdc.gov/>

Gouin D., Morin P., Clément G., Larouche C. [GOI 97]

1997 Solving the Geospatial Data Barrier. *Geomatica*, Vol. 51, num. 3, 1997, pp 278-287.

Inmon, W.H. [INM 96]

1996 Building the Data Warehouse. John Wiley & Sons, 410 p

Létourneau F. [LET 98]

1998 Analyse du potentiel de l'approche entrepôt de données pour l'intégration des métadonnées provenant d'un ensemble de géorépertoires disponibles sur Internet. Mémoire de maîtrise pour l'obtention du grade de maître ès science (M. Sc.). Département des sciences géomatiques, Université Laval, Québec, Canada, 188 pp

Maamar Z., Moulin B. Bédard Y., [MAA 98]

1998 Software Agent-Oriented Frameworks for the Interoperability of Georeferenced Digital Libraries on the World Wide Web: the SIGAL Project. Dans *Interoperating Geographic Information Systems*, édité par M. Goodchild, M. Egenhofer, R. Fegeas et C. Kottman, publié par Kluwer (soumis).

Maamar Z. [MAA 97]

1997 Contribution à la résolution des problèmes d'interopérabilité des systèmes, une méthode de conception par frameworks orientés-agents logiciels. Thèse de PhD, Département d'informatique, Université Laval. Québec, Canada.

Moriarty T., Greenwood R.P. [MOR 96]

1996 Data's Quest From Source to Query. *Database Programming & Design*, octobre 1996, pp 78-81

Poe V. [POE 95]

1995 Building a Data Warehouse for Decision Support. Practice Hall, 210 p

Proulx, M.J., Bédard Y., Létourneau F., Martel C. [PRO 97]

1997 Catalogage des données spatiales sur le World Wide Web : concept, analyses des sites et présentation du géorépertoire personnalisable GEOREP. *Revue Internationale de Géomatique*, Paris, Édition Hermès, vol. 7, no 1, mars 1997.

Rawling J., Kucera H. [RAW 97]

1997 The Trials and Tribulation of Implementing a Spatial Data Warehouse. Eleventh Annual Symposium on Geographic Information Systems, GIS 97, Vancouver, février 1997, p. 510

[SAIF]

Norme d'échange des données SAIF : <http://www.elp.gov.bc.ca/gdbc/fmebc/>

White C. [WHI 96]

1996 Data Warehouse : What's in a Name? Database Programming & Design, mars 1996, pp 53 - 57