



Le géorépertoire comme mode de recherche et d'accès direct aux données à référence spatiale

*F. Létourneau
RDDC Valcartier*

R&D pour la défense Canada – Valcartier

Mémoire technique

DRDC Valcartier TM 2004-376

Juin 2007

Canada

Le géorépertoire comme mode de recherche et d'accès direct aux données à référence spatiale

F. Létourneau
RDDC Valcartier

R et D pour la défense Canada – Valcartier

Mémoire technique
DRDC Valcartier TM 2004-376
Juin 2007

Auteur

François Létourneau

Approuvé par

Yves van Chestein
C/GIC

Publication approuvée par

Christian Carrier
SCh

© Sa Majesté la Reine, représentée par le ministre de la Défense nationale, 2007

© Her Majesty the Queen as represented by the Minister of National Defence, 2007

Abstract

The georeferenced digital library concept has considerably evolved during the last decade. From a simple repository of geospatial documents in the 1990s, this concept is now more in phase with a geospatial data access infrastructure, integrating various data discovery technologies and running in a Web environment. The integrated georeferenced digital library (IGDL) is an application working with the latest geospatial technologies like geospatial data warehouses, Internet map and imagery servers and automated geospatial data discovery Web services. The IGDL is no longer a unique application oriented solely towards geospatial data discovery.

An R&D project aiming at developing an IGDL has been conducted at DRDC Valcartier between 2000 and 2004. Through this project, developed IGDL prototypes feature innovative aspects, like the partial automation of geospatial data inventory and metadata capture, the definition of complex and irregularly shaped areas of interest and the provision of direct access to native or Web-based geospatial data.

This document introduces and describes the new concepts associated with the IGDL and positions it within a geospatial data infrastructure. It describes the project stages and emphasizes the guidelines that have oriented the development of the application. It concludes with descriptions of new development orientations for IGDL, including OLAP and peer-to-peer frameworks, which are used within new implementations of IGDLs for concept demonstration purposes.

Résumé

Le concept de géorépertoire a évolué considérablement depuis près d'une décennie. Du simple répertoire de documents à référence spatiale qu'il était dans les années 90, ce concept a évolué vers une infrastructure d'accès aux données géospatiales, intégrant de multiples technologies de découverte des données et fonctionnant dans un environnement Web. Le géorépertoire intégrateur est une application fonctionnant avec les toutes dernières technologies géomatiques, telles que les entrepôts de données, les serveurs Web de cartes et d'images et les services Web de recherche géodocumentaire automatisée. Le géorépertoire n'est donc plus uniquement une application de recherche géodocumentaire.

Un projet de recherche visant à concevoir un géorépertoire évolué a été mis de l'avant au début de l'an 2000 à RDDC – Valcartier et s'est poursuivi jusqu'en 2004. Les prototypes de géorépertoires découlant de ce projet présentent des caractéristiques novatrices, telles que la semi-automatisation de l'inventaire et de la saisie des données, la définition d'aires d'intérêt de forme irrégulière et l'accès direct aux données, sous forme native ou sous forme de pages Web.

Ce document vise à présenter les nouveaux concepts du géorépertoire et à le situer par rapport à une infrastructure de données à référence spatiale. Il fera donc principalement état de l'avancement de la recherche à ce sujet à RDDC – Valcartier en mettant l'accent sur les principes qui ont guidé son développement. Finalement, il sera fait mention des derniers développements intégrant les technologies OLAP et poste à poste dans des géorépertoires, à titre de démonstrateurs technologiques.

Executive summary

In a digital world where geospatial data is increasingly available, modern and efficient applications must be developed and used in order to facilitate geospatial data discovery and exploitation. While the frequency at which geospatial data is added to a geodata infrastructure might be considered low, it is a field where data may be oftentimes modified, following updates or in the case of creation of new products. Some datasets form the foundation of a repository of geospatial data and are not often modified, while others tend to be much more frequently updated. As a consequence, the content of geospatial data repositories faces a likelihood of frequent changes and thus Georeferenced Digital Libraries (GDLs) require a robust and modern technological environment that supports this dynamic, constantly evolving situation.

Georeferenced digital libraries are built to support the tasks of discovering geospatial data stored into various repositories or databases. Similar to bibliographical indexes found in public libraries, GDLs feature advanced descriptive and spatial analysis functions to help a user to find the proper datasets that answer his needs. Because many of the datasets referenced into GDLs are digital, it is logical and efficient to build within GDL interfaces functions that permits not only the discovery of the datasets, but also their access and exploitation. With the wide availability of Open Geospatial Consortium (OGC) interfaces implemented into many of the most popular commercial products, it is now easier to design GDLs that offer data access and visualization mechanisms. Moreover, implementing Open Geospatial Datastore Interface (OGDI) interfaces to GDLs allows an access not only to an image of the data, but to the data itself, which could potentially lead to the advanced exploitation of data, like spatial analysis, specialized or user-defined rendering and so on. In a nutshell, GDLs are no longer confined to geospatial data discovery, but now feature advanced integration with other systems that allow data retrieval, exploitation and analysis.

Létourneau, F. 2007. Le géorépertoire comme mode de recherche et d'accès direct aux données à référence spatiale. DRDC – Valcartier TM 2004-376. Defence R&D Canada – Valcartier

Sommaire

Dans un monde numérique où les données géospatiales sont de plus en plus accessibles, des applications modernes et efficaces doivent être développées afin de faciliter la découverte et l'exploitation de ces données. Bien que la fréquence à laquelle les données sont ajoutées dans une infrastructure de données géospatiales puisse être considérée faible, il s'agit en fait d'un domaine où les données sont susceptibles de changer très rapidement, soit à la suite de mises à jour, soit pour la production de nouveaux produits. Certains produits constituent les fondations relativement stables d'une base de données géospatiales, tandis que d'autres sont d'un naturel plus changeant. En conséquence, les répertoires de données géospatiales font face à une situation de mises à jour fréquentes de leur contenu et requièrent de ce fait un environnement technologique robuste et moderne qui supporte efficacement cette situation de changements continuels.

Les géorépertoires sont conçus pour supporter les tâches de découverte des données géospatiales disponibles dans les répertoires ou les bases de données. Similaires aux index bibliographiques des bibliothèques publiques, les géorépertoires offrent des services avancés d'analyse spatiale et descriptive permettant à un utilisateur de trouver les jeux de données correspondant à ses besoins. Puisque la majorité des jeux de données référencés par les géorépertoires sont numériques, il est logique et efficace d'offrir via les interfaces des géorépertoires des fonctions qui permettent non seulement la découverte des données, mais aussi leur accès ainsi que leur exploitation. En raison de la grande disponibilité d'interfaces OGC implantées dans de multiples systèmes d'information géographique commerciaux, il est maintenant de plus en plus facile de définir des géorépertoires qui offrent concurremment des mécanismes d'accès et de visualisation des données. D'autre part, en implantant des interfaces OGD dans les géorépertoires, il est désormais possible non seulement d'accéder à des images des données, mais aussi d'accéder directement à ces données afin d'effectuer des analyses spatiales complexes ou encore des rendus complexes des données effectués par les différents utilisateurs de ces dernières. En résumé, les géorépertoires ne sont plus uniquement confinés à la découverte des données, mais permettent maintenant l'exploitation avancée de celles-ci en se liant à des systèmes qui permettent l'accès et l'exploitation des données.

Létourneau, F. 2007. Le géorépertoire comme mode de recherche et d'accès direct aux données à référence spatiale. DRDC – Valcartier TM 2004-376. R et D pour la défense Canada – Valcartier

Table des matières

Abstract	i
Résumé	ii
Executive summary	iii
Sommaire.	iv
Liste des acronymes	viii
Remerciements	xi
1 Introduction	1
2 Mise en contexte.	3
2.1. Évolution du concept de géorépertoire.	3
2.2. Environnements technologique et géomatique du géorépertoire	6
3 Le géorépertoire intégrateur	7
3.1. Prise en charge des données sous forme numérique	7
3.2. Automatisation des fonctions d'inventaire.	8
3.3. Automatisation de la saisie des références aux données géospatiales	10
3.4. Génération d'une imagerie d'un document	10
3.5. Interface de saisie des métadonnées	11
3.6. Considérations technologiques pour l'automatisation de l'inventaire des données	12
3.7. Fonctions de cartographie dynamique	13
3.8. Définition de la zone d'intérêt	13
3.9. Requêtes spatio-temporelles et descriptives	15
3.10 Enregistrement des requêtes, récupération des requêtes	16
3.11 Accès direct aux données sous leur forme native ou par le Web.	16
4 Environnement de développement du géorépertoire	18
5 Nouvelles perspectives de développement des géorépertoires	19
5.1. L'intégration de la technologie OLAP dans les géorépertoires.	19
5.2. Le géorépertoire distribué	21
6 Conclusion	24
Bibliographie.	27
Liste de distribution	29

Liste des figures

Figure 1 : Carte-index, basée sur le Système national de référence cartographique	3
Figure 2 : MCD conçu pour le géorépertoire	8
Figure 3 : Exemple d'interface de saisie des métadonnées manquantes.	11
Figure 4 : Comparaison des modes de recherche de documents. En rouge, définition d'une aire de recherche et résultats obtenus (méthode traditionnelle) En violet, définition d'une polygone de recherche et résultats obtenus (méthode proposée pour le géorépertoire).	14
Figure 5 : Exemple de requête descriptive, interface Web	15
Figure 6 : Présentation des résultats d'une requête	16
Figure 7 : Accès direct aux données, version Web	17
Figure 8 : Module d'analyse de la couverture cartographique, maquette GEOLAP	21
Figure 9 : Architecture proposée pour le géorépertoire distribué	23

Liste des tableaux

Tableau 1 : Synthèse de l'évolution du géorépertoire	5
Tableau 2 : Exemple de résultat obtenu avec OGDI Crawler.	9

Liste des acronymes

BNDT	Banque nationale de données topographiques
DEM	Digital Elevation Model
DGN	Design File
DIGEST	Digital Geographical Information Exchange Standard
DTED	Digital Terrain Elevation Data
FGDC	Federal Geographic Data Community
GDL	Georeferenced Digital Library
GEOLAP	Geospatial On-Line Analytical Processing
GeoTiff	Geographic Tagged Image File Format
GLTP	Geographic Language Transfer Protocol
IGDL	Integrated Georeferenced Digital Library
ISO	International Organisation for Standardisation
iTiff	Internet Tagged Image File Format
HTTP	Hyper Text Transfer Protocol
J2 IM	Joint Information Management
MCD	Modèle conceptuel de données
MCE	Mapping and Charting Establishment
NGA	National Geo-Intelligence Agency
NIMA	National Imagery and Mapping Agency
OGC	Open GIS Consortium
OGDI	Open Geospatial Datastore Interface
OLAP	On-Line Analytical Processing
RPF	Raster Product Format
SIG	Système d'information géographique
SQL	Structured Query Language
SGBD	Système de gestion de bases de données
SVG	Scalable Vector Graphics
TC/211	Technical Committee 211

TIFF	Tagged Image File Format
UML	Universal Modelling Language
URL	Universal Resource Locator
VPF	Vector Product Format
WFS	Web Feature Server
WMS	Web Mapping Server
XML	Extensible Markup Language

Cette page est laissée intentionnellement en blanc

Remerciements

Nous désirons remercier les différentes personnes qui ont contribué au développement de ce projet. En premier lieu, M. Christian Martel, de RDDC Valcartier, qui a mis au point le prototype de géorépertoire intégrateur. M. Martel a effectué un travail remarquable pour modéliser les processus ainsi que pour créer le modèle de données du géorépertoire.

Nous désirons aussi remercier M. Gilles Clément, de la firme Global Geomatics, qui a en de nombreuses occasions adapté certaines fonctions de MapFusion afin de répondre à nos objectifs de développement et nous a fourni plusieurs conseils et excellentes idées. Finalement, nous désirons souligner l'apport de M. Paul Morin, de J2 IM, pour les nombreux commentaires constructifs et l'aide au développement du concept intégrateur de ce projet.

Cette page est laissée intentionnellement en blanc

1 Introduction

Trouver de l'information, peu importe sa forme, n'a jamais été une tâche facile. Ce problème devient de plus en plus aigu dans notre société, où la quantité d'information produite atteint des sommets inégalés. Heureusement, les technologies de l'information offrent des outils performants pour y parvenir. En géomatique, la découverte de l'information – les données à référence spatiale – a toujours représenté un défi important. En effet, différents paramètres entrent en ligne de compte dans les processus de recherche d'information à caractère spatial : couverture spatiale du document, échelle, date de production et de prise de données, formats utilisés pour représenter l'information, couverture nuageuse pour les images du territoire, etc. D'autre part, la majorité des utilisateurs de données à référence spatiale ne soupçonnent pas la richesse et la quantité de données disponibles et sont souvent démunis lorsqu'ils entreprennent des démarches en vue d'acquérir des données. Le géorépertoire est un outil qui peut être utilisé efficacement pour résoudre la majorité de ces problèmes ou à tout le moins pour les atténuer.

Au cours de plusieurs projets de recherche en développement d'applications de commandement et contrôle à RDDC Valcartier, nous avons été témoins de grandes difficultés de la part des analystes et des concepteurs des systèmes lors de leur recherche de données spatiales. La majorité d'entre eux se contentaient de réutiliser les mêmes jeux de données d'un projet à l'autre. Il faut indiquer toutefois que ces applications étaient utilisées principalement pour démontrer des principes et ne servaient pas dans les systèmes opérationnels. Pourtant, certains de ces projets auraient pu avoir un impact plus grand au niveau de l'interface cartographique s'ils avaient bénéficié de données géospatiales plus représentatives du territoire qu'elles représentaient ; malheureusement, ces données sont pour la plupart inconnues parce que mal ou non répertoriées. En offrant aux utilisateurs, analystes et concepteurs de systèmes des outils modernes et efficaces leur permettant de découvrir et d'exploiter les données géomatiques, nous offrons de nouvelles possibilités d'exploiter de façon optimale les données représentant le territoire dans les systèmes de commandement et contrôle et les applications militaires géomatiques.

C'est en gardant à l'esprit ces deux aspects que nous avons mis en branle, au cours de l'année 2000, un projet de développement d'un géorépertoire intégrateur. Le terme intégrateur représente une nuance importante au concept connu de géorépertoire, ce dernier n'étant plus conçu comme une application ayant sa propre finalité, mais plutôt comme une application pouvant être intégrée à une infrastructure de données géospatiales et offrant aussi un lien direct vers les données. Ainsi, le géorépertoire ne permet plus uniquement de découvrir les données disponibles au sein d'une organisation, il permet désormais à un utilisateur d'accéder et de visualiser directement celles-ci, en intégrant dans le géorépertoire les différents modes d'accès aux données à référence spatiale.

Les données géospatiales disponibles à RDDC Valcartier sont principalement des données de type militaire (VRF, RPF, DTED), mais une partie importante de ces données sont de format civil (Arc/Info, ShapeFile, DGN, DEM, GeoTIFF, iTIFF, etc.). Des données sont disponibles sous support numérique (CD-ROM) et aussi sur différents serveurs. Les serveurs de données géomatiques sont les suivants :

- Open Geospatial Datastore Interface (OGDI) / MapFusion : offre l'accès natif à plus de 80 formats de données, sans recourir à des conversions.
- Entrepôt de données : offre l'accès Web avec le protocole Web Mapping Service (WMS 1.x). Plusieurs autres systèmes offrent ce type d'interface aux données.
- Serveur d'images : offre l'accès aux images de type TIFF, GeoTIFF et iTIFF.

Le géorépertoire intégrateur agit un peu comme un courtier en données non affilié à un système particulier et offrant la découverte, l'accès et la visualisation de données, afin éventuellement d'en permettre leur utilisation dans différentes applications.

L'objectif de ce document technique est de faire état de l'avancement du projet de développement du géorépertoire, ainsi que de décrire les caractéristiques techniques du prototype développé au cours du projet. De plus, ce document permettra aux lecteurs de positionner ce projet par rapport au projet GeoGateway Services, un projet parainné par J2 Geomatics (maintenant J2 IM) et MCE et qui vise à développer une infrastructure offrant un accès efficace aux données géospatiales militaires.

Les chapitres suivants présenteront dans un premier temps l'évolution du concept de géorépertoire. Ceci permettra d'avoir une meilleure perspective concernant la nature des efforts entrepris pour la conception du géorépertoire intégrateur. Il s'ensuivra une description détaillée des différentes fonctions que nous avons implantées ainsi qu'une présentation de l'environnement technologique choisi.

2 Mise en contexte

Initialement, le géorépertoire était une application qui simplifiait la découverte des données géospatiales. D'une application spécialisée qu'il était à ses débuts, il fait maintenant partie intégrante de différents systèmes SIG. Les prochaines sections retracent l'évolution du concept du géorépertoire.

2.1 Évolution du concept de géorépertoire

Le géorépertoire est un concept qui existe depuis bien plus longtemps que le terme lui-même. En effet, depuis de nombreuses années, des cartes générales étaient utilisées pour représenter la couverture spatiale d'un ensemble de cartes, photos aériennes et images satellites. Cette méthode de représentation de l'information s'est longtemps révélée efficace et le demeure encore aujourd'hui, dans la mesure où les documents ou collections représentés ne changent à peu près jamais. La carte index peut représenter un mode d'indexation convenable pour des collections de documents telles que les feuillets 1 : 50 000 de la BNDT (voir figure 1).

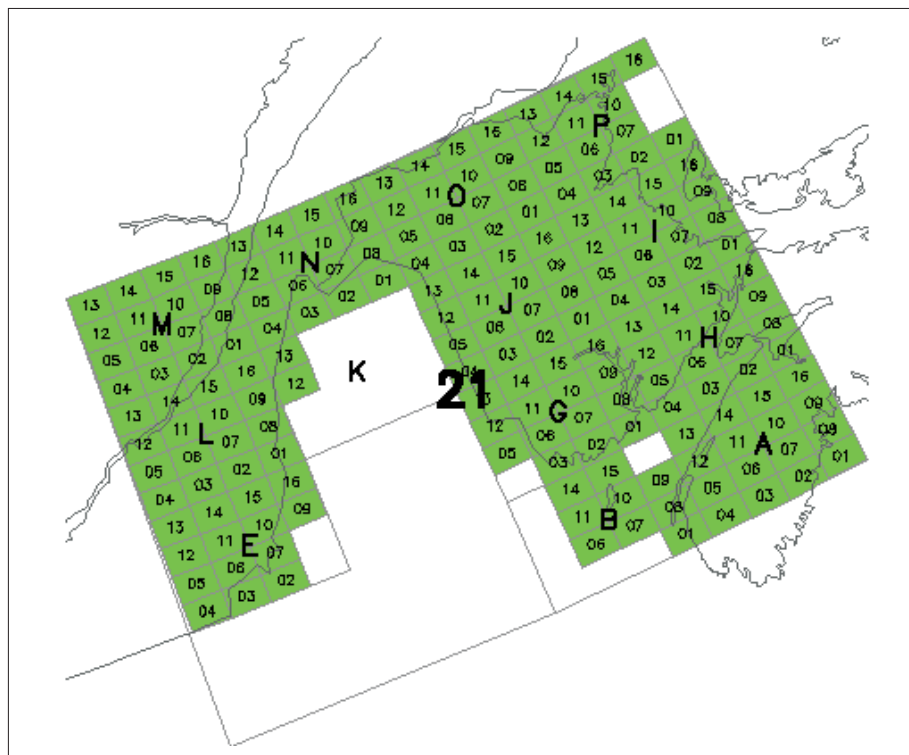


Figure 1 : Carte-index, basée sur le Système national de référence cartographique

Nombre d'organismes désirant déployer une solution plus flexible et plus moderne ont constaté graduellement que la solution de la carte index n'était plus adéquate. En effet, il devenait difficile d'y intégrer d'autres informations, ne serait-ce qu'un ensemble de métadonnées minimales. Les nombreux documents à référence spatiale disponibles rendaient cette solution de plus en plus difficile à supporter.

Les premiers géorépertoires ont vu le jour au début des années 90. Appelés alors atlas électroniques, ils utilisaient des systèmes d'information géographique (SIG), lesquels servaient à représenter les couvertures spatiales de cartes, photos et autres documents à référence spatiale et de coupler des bases de données permettant d'accéder aux métadonnées associées à ces documents. Cette approche requérait un inventaire manuel complet des données géoréférencées disponibles ainsi qu'un important effort pour effectuer la saisie des métadonnées. Cette approche a prévalu au cours de la première moitié de la décennie 90, avec quelques variantes. Bien qu'elle offre une plus grande flexibilité que la carte index, elle ne réglait pas les problèmes d'accès et de diffusion de l'information et s'avérait lourde pour gérer les mises à jour. D'autre part, l'accès à l'atlas électronique était limité en raison de l'obligation d'utiliser un SIG, dont le coût élevé limitait le déploiement à grande échelle. N'oublions pas que la technologie Web n'était pas encore disponible dans la première moitié de la décennie 90, ce qui limitait considérablement la diffusion de l'information.

L'arrivée d'Internet en 1995 a constitué une révolution importante pour les géorépertoires. En effet, on pouvait alors accéder directement au contenu de ceux-ci, via Internet, tant pour l'information à caractère spatial ou descriptive. Un utilisateur pouvait rechercher les documents disponibles non seulement au sein de son organisation, mais aussi chez d'autres producteurs de données s'affichant dans Internet.

Toutefois, cette abondance subite de géorépertoires présentait aussi des problèmes, identifiés par [Létourneau, 1998], ainsi que [Létourneau, Bédard et Moulin, 1998]. L'hétérogénéité des informations transmises et le nombre élevé de sites de géorépertoires à consulter par l'utilisateur devenant alors un frein à leur pleine utilisation. L'approche "entrepôt de données" préconisée par ces auteurs représente une avenue de solution intéressante, ainsi que celle par agents intelligents, décrite par [Maamar, Moulin et Bédard, 1999]. En effet, ces solutions, qui visent à faire la synthèse des données disponibles dans de multiples géorépertoires et à les intégrer dans un référentiel de métadonnées commun, facilitent la recherche globale de données parce que l'utilisateur n'a plus à consulter de multiples géorépertoires pour trouver l'information dont il a besoin. Toutefois, ces systèmes n'ont pas abordé le problème d'accès direct aux données et ne permettent aucune transformation des différents référentiels spatiaux (systèmes de coordonnées, projection) ou formats de données, ne simplifiant donc pas la tâche de l'utilisateur qui cherche à exploiter des données hétérogènes provenant de différentes sources.

Les géorépertoires traditionnels ne permettent habituellement pas aux utilisateurs d'accéder directement aux données et de les visualiser. Habituellement, l'utilisateur n'obtient qu'une référence lui indiquant où se procurer les données. L'accès aux données peut se faire avec des bons de commandes, un système de requête sur le Web, l'installation de logiciels de visualisation des données, etc.

Une fois les données trouvées et récupérées, il est nécessaire de les intégrer dans l'application visée. L'intégration des données à référence spatiale dans un SIG, par exemple, n'est pas forcément une tâche facile, si l'on tient compte des formats de données disponibles et des problèmes inhérents de conversion qui peuvent se présenter. D'ailleurs, il est généralement admis que l'acquisition des données compte pour 60 à 85% des coûts d'implantation d'un SIG [Gouin, 97], dont une grande part peut être attribuée à la conversion des données.

Finalement, le temps et les ressources dévolues à l'inventaire des données géospatiales d'une organisation et à la saisie des métadonnées peuvent représenter un effort considérable pour une organisation. Il peut être difficile d'identifier et d'inventorier toutes les données disponibles et des erreurs de saisie sont toujours possibles. Une saisie automatisée des métadonnées pourrait accélérer le processus d'inventaire et permettrait de réduire les erreurs de saisie. Le tableau 1 synthétise l'évolution du géorépertoire des années 80 à aujourd'hui.

Période	Type de géorépertoire	Description sommaire
- à aujourd'hui	Carte-index	Carte papier représentant les couvertures géographiques des documents
1985 à 1990	Carte-index numérique	Numérisation de plusieurs carte index sur différents niveaux d'un fichier de CAO
1988 à 1992	Atlas électronique	Implantation de la carte-index numérique dans un SIG <ul style="list-style-type: none"> • Métadonnées minimales associées aux divers éléments • Premières fonctions d'analyse spatiale • Un seul utilisateur par système
1993 à 1995	De l'atlas électronique au géorépertoire	Couplage de l'atlas électronique à une base de données relationnelle <ul style="list-style-type: none"> • Premières normes de métadonnées • Requêtes spatio-temporelles et descriptives plus évoluées • Un seul utilisateur par système
1995 à 1996	Géorépertoire sur Internet	Base de données du géorépertoire accessible par interface Web <ul style="list-style-type: none"> • Premier géorépertoire client-serveur • Interface limitée pour les requêtes • Faible support pour l'analyse spatiale
1997 à 2000	Géorépertoire graphique sur le Web	Premiers couplages avec des SIG sur le Web <ul style="list-style-type: none"> • Approche client-serveur • Interface Java / Active-X pour la définition des requêtes • Interface cartographique pour la définition de l'aire d'intérêt • Basé sur des entrepôts de données
2000 / 2001	Géorépertoire intégrateur	Évolution du concept vers le portail géomatique <ul style="list-style-type: none"> • Interface Web, utilisateurs multiples • Inventaire automatisé des données numériques • Support de la norme ISO TC/211 • Visualisation et accès direct aux données
2001 à 2004	Composante de SIG	Implantation de géorépertoire comme composante d'un SIG <ul style="list-style-type: none"> • Recherche automatisée dans MapFusion, ArcGIS et autres • Indexation des résultats • Accès direct aux données

Tableau 1 : Synthèse de l'évolution du géorépertoire

2.2 Environnements technologique et géomatique du géorépertoire

Parallèlement à l'évolution des géorépertoires, les systèmes géomatiques ont eux aussi beaucoup évolué ces dernières années. Les logiciels de géomatique empruntent de plus en plus la voie tracée par les suites graphiques et bureautiques où l'on trouve une très grande intégration des différentes fonctions dans une suite logicielle complète. Ainsi, les logiciels SIG sont-ils maintenant capables de traiter des données d'imagerie et de photogrammétrie, alors que les logiciels d'imagerie intègrent désormais des couches vectorielles. Les frontières entre les données vectorielles et matricielles s'estompent, les métadonnées sont de plus en plus présentes, tandis que la publication sur le Web de cartes, photos et images composites devient la norme. D'autre part, plusieurs des problèmes d'accès aux données sont en voie d'être résolus, en employant les technologies d'entrepôt de données à référence spatiale, ainsi que des mécanismes d'accès direct aux données tels que l'OGDI. Les interfaces WMS et WFS offrent aussi un bon potentiel pour l'intégration des données, que ce soit au moyen des serveurs en cascade pour le protocole WMS ou par la transformation des données pour le protocole WFS.

Notons aussi que les métadonnées sont de plus en plus imbriquées dans les structures de données auxquelles elles sont rattachées. Cette intégration permet donc une automatisation de la cueillette des métadonnées, au moyen de mécanismes d'extraction de ces données. Dans le cas où les données sont placées sur un serveur Web, il est donc possible à la fois d'extraire les métadonnées d'un jeu de données ainsi que de visualiser les données, dans une même séquence d'opérations.

En utilisant un mécanisme approprié pour stocker l'extraction des métadonnées, dans une structure XML par exemple, l'intégration efficace et élégante de cette information dans une base de données s'en trouve grandement facilitée. Des mécanismes, développés sous forme de services Web, peuvent être mis en place afin de détecter tout nouveau jeu de données disponible et, subséquemment, extraire automatiquement les métadonnées associées à ces jeux de données.

C'est dans ce nouveau contexte que se situe le géorépertoire du futur. Il devra permettre à une organisation de référencer l'ensemble de ses données géomatiques disponibles, tout en offrant des accès à des serveurs externes. Il devra aussi offrir des mécanismes améliorés pour l'inventaire et la saisie des données. Finalement, en exploitant la technologie Internet, il devra offrir, dans la mesure du possible, l'accès direct aux données recherchées dans leur mode natif ou sous forme de pages Web.

3 Le géorépertoire intégrateur

Au début de l'an 2000, un projet de recherche a été lancé à RDDC Valcartier pour concevoir une nouvelle génération de géorépertoires. Le but principal du projet était de développer un géorépertoire répondant aux besoins d'une organisation cherchant à rendre disponibles ses données géospatiales en mode natif ou à l'aide de pages Web, sous forme d'images ou de fichiers SVG. Ce type de géorépertoire devait permettre de référencer à la fois des données stockées sur des serveurs internes et externes à une organisation. Son mode de fonctionnement convient plus à un intranet, bien qu'il offre certaines fonctions permettant une recherche de données externes à l'organisation ainsi qu'un accès externe aux données (données disponibles sur un site Internet ou par un serveur OGDl externe par exemple). Deux versions de l'application géorépertoire ont été développées, la première sous la forme d'une application native Windows et la seconde sous la forme d'une application Web fonctionnant sur un intranet. Les deux versions de l'application utilisent le même SGBD, les mêmes bases de données, ce qui résulte en une synchronisation totale entre le contenu des deux versions de l'application.

Les caractéristiques du système sont les suivantes :

- Prise en charge des données sous forme numérique seulement
- Automatisation des fonctions d'inventaire
- Automatisation de la saisie des données
- Semi-automatisation de la saisie des métadonnées
- Génération d'images représentant les documents
- Fonctions de cartographie dynamique
- Fonctions avancées de définition de la zone d'intérêt (point, ligne, polygone)
- Requêtes spatio-temporelles et descriptives
- Enregistrement des requêtes, ouverture de requêtes enregistrées
- Lorsque possible, accès direct aux données sous leur forme native ou Web

Les sections suivantes décrivent en détail chacun des éléments présentés ci-haut.

3.1 Prise en charge des données sous forme numérique

Bien que les documents papier soient encore très nombreux, ils ne représentent plus la source de données privilégiée si l'on cherche à intégrer de l'information à caractère spatial dans une application de commandement et contrôle ou de géomatique militaire. À moins de disposer d'un scanner ainsi que d'un mécanisme pour rétablir la référence

géographique, ces documents n'ont plus beaucoup d'intérêt dans le monde des applications numériques. C'est pourquoi ce projet n'aura pour but que de traiter des données numériques. Le modèle conceptuel de données (MCD) du système reflète d'ailleurs cette orientation. Basé sur un profil de la norme ISO 19115, ce modèle est optimisé pour gérer efficacement l'hétérogénéité de l'environnement géographique militaire avec lequel le géorépertoire doit composer. La figure 2 présente le MCD conçu pour le prototype du géorépertoire intégrateur. Ce MCD, en format UML, permet de gérer à la fois des collections de documents (par exemple, l'ensemble des feuillets 1 : 250 000 de la BNDT), des documents (un feuillet de cette collection) et des couvertures parmi un feuillet (exemple, l'hydrographie d'un feuillet). Ce modèle de données a été entièrement implanté dans le SGBD Corel Paradox.

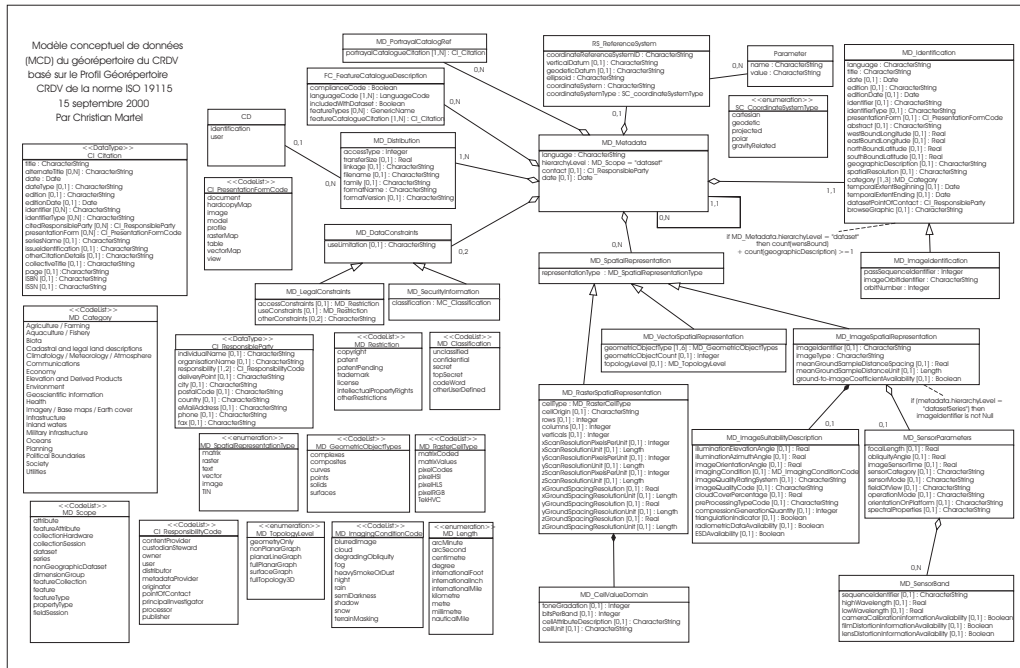


Figure 2 : MCD conçu pour le géorépertoire

3.2 Automatisation des fonctions d'inventaire

L'inventaire des données est la pierre angulaire d'un géorépertoire au contenu riche et diversifié. En effet, cette opération est essentielle à la mise en place d'un géorépertoire qui offre une référence complète aux données disponibles pour une organisation. Étant donné que nous avons choisi de ne travailler qu'avec des données numériques, nous avons la possibilité d'automatiser l'inventaire des données. Les données numériques se retrouvent habituellement sur des disques rigides, des cédéroms, des bases de données ou des entrepôts de données. Afin d'automatiser l'inventaire des données, il s'agit, dans le cas de données stockées sur des disques, de simplement parcourir ces supports de données et de répertorier les fichiers qui correspondent aux critères de recherche. Par exemple, on pourra trouver tous les fichiers ArcView ShapeFile, en recherchant les fichiers dont l'extension est .shp. Toutefois, cela ne garantit pas que le fichier soit véritablement un ShapeFile, car aucune validation n'est effectuée.

Cette tâche est cependant beaucoup plus complexe dans le cas de données structurées en base de données, telles que les données VRF (vector relational format) dont les données ne sont pas contenues dans un seul fichier mais dans un ensemble complexe de fichiers formant une base de données. Des solutions technologiques performantes seront utilisées pour résoudre ce problème.

Pour automatiser l'inventaire des données, trois mécanismes distincts sont proposés. Le premier mécanisme fait appel à l'OGDI crawler. Cette application a pour but de répertorier sur un serveur OGD I toutes les sources de données pouvant être accessibles par l'OGDI. La recherche produit un fichier texte structuré (disponible ultérieurement sous la forme d'une structure XML) pouvant être intégré à la base de données du géorépertoire. Le tableau 2 présente un exemple de résultat de recherche de données. En plus d'inscrire les références aux données (c'est-à-dire., chemins d'accès au serveur OGD I), cette application extrait un ensemble minimal de métadonnées pour tous les jeux de données inventoriés. Ainsi, la couverture spatiale des données y est extraite, ainsi que d'autres métadonnées en fonction de leur disponibilité. Les données DIGEST, de par leur nature, comportent un bon ensemble de métadonnées de base qu'il est possible d'exploiter. Finalement, il est possible de créer une imagerie de chaque jeu de données avec cette application .

```

BEGIN_GROUP = STATUS ProtocolVersion = 1.1 BoundingWENS = Y TimePeriod = N
FreeText = Y NumberOfProducts = 20 END_GROUP = STATUS

BEGIN_GROUP = TARGET Supplier = «a» DataSet = «b» AuxInfo = «c» END_GROUP = TARGET

BEGIN_GROUP = PRODUCT ProductID = img1 BoundingWENS =
(-71.250,-71.145,46.731,46.658) Time = 1999-01-01T00:00:00 Name =
«geotiff/e:/msidata/crdv/94813_26.TIF» Browse = Full =
«gltp://huahine.drev.dnd.ca/geotiff/e:/msidata/crdv/94813_26.TIF» DetailedMetadata =
«gltp://huahine.drev.dnd.ca/geotiff/e:/msidata/crdv/94813_26.TIF» coverages = « {
«gltp://huahine.drev.dnd.ca/geotiff/e:/msidata/crdv/94813_26.TIF»
Matrix { 94813_26 } } » FreeText = «Matrix» END_GROUP = PRODUCT

BEGIN_GROUP = PRODUCT ProductID = img2 BoundingWENS =
(-71.347,-71.242,46.729,46.656) Time = 1999-01-01T00:00:00 Name =
«geotiff/e:/msidata/crdv/94813_28.TIF» Browse = Full =
«gltp://huahine.drev.dnd.ca/geotiff/e:/msidata/crdv/94813_28.TIF» DetailedMetadata =
«gltp://huahine.drev.dnd.ca/geotiff/e:/msidata/crdv/94813_28.TIF» coverages = « {
«gltp://huahine.drev.dnd.ca/geotiff/e:/msidata/crdv/94813_28.TIF»
Matrix { 94813_28 } } » FreeText = «Matrix» END_GROUP = PRODUCT

BEGIN_GROUP = PRODUCT ProductID = img3 BoundingWENS =
(-71.445,-71.339,46.731,46.658) Time = 1999-01-01T00:00:00 Name =
«geotiff/e:/msidata/crdv/94813_30.TIF» Browse = Full =
«gltp://huahine.drev.dnd.ca/geotiff/e:/msidata/crdv/94813_30.TIF» DetailedMetadata =
«gltp://huahine.drev.dnd.ca/geotiff/e:/msidata/crdv/94813_30.TIF» coverages = « {
«gltp://huahine.drev.dnd.ca/geotiff/e:/msidata/crdv/94813_30.TIF»
Matrix { 94813_30 } } » FreeText = «Matrix» END_GROUP = PRODUCT

BEGIN_GROUP = PRODUCT ProductID = img4 BoundingWENS =
(-71.462,-71.356,46.788,46.715) Time = 1999-01-01T00:00:00 Name =
«geotiff/e:/msidata/crdv/94813_47.TIF» Browse = Full =
«gltp://huahine.drev.dnd.ca/geotiff/e:/msidata/crdv/94813_47.TIF» DetailedMetadata =
«gltp://huahine.drev.dnd.ca/geotiff/e:/msidata/crdv/94813_47.TIF» coverages = « {
«gltp://huahine.drev.dnd.ca/geotiff/e:/msidata/crdv/94813_47.TIF»
Matrix { 94813_47 } } » FreeText = «Matrix» END_GROUP = PRODUCT

BEGIN_GROUP = PRODUCT ProductID = img12 BoundingWENS =
(-71.472,-71.365,46.900,46.827) Time = 1999-01-01T00:00:00 Name =
«geotiff/e:/msidata/crdv/94813_85.TIF» Browse = Full =
«gltp://huahine.drev.dnd.ca/geotiff/e:/msidata/crdv/94813_85.TIF» DetailedMetadata =
«gltp://huahine.drev.dnd.ca/geotiff/e:/msidata/crdv/94813_85.TIF» coverages = « {
«gltp://huahine.drev.dnd.ca/geotiff/e:/msidata/crdv/94813_85.TIF»
Matrix { 94813_85 } } » FreeText = «Matrix» END_GROUP = PRODUCT

BEGIN_GROUP = PRODUCT ProductID = img13 BoundingWENS =
(-71.370,-71.264,46.900,46.828) Time = 1999-01-01T00:00:00 Name =
«geotiff/e:/msidata/crdv/94813_87.TIF» Browse = Full =
«gltp://huahine.drev.dnd.ca/geotiff/e:/msidata/crdv/94813_87.TIF» DetailedMetadata =
«gltp://huahine.drev.dnd.ca/geotiff/e:/msidata/crdv/94813_87.TIF»

```

Tableau 2 : Exemple de résultat obtenu avec OGD I Crawler

Le deuxième mécanisme fait appel à la méthode *getcapabilities* du protocole WMS. En invoquant cette méthode au moyen d'un appel HTTP, tout serveur de données supportant ce protocole produira un fichier XML indiquant les données disponibles sur ce serveur. De plus, s'il implante le mécanisme Cascading Map Server, le serveur retournera non seulement la liste de ses données lors de la requête *getcapabilities*, mais aussi la liste des données de tous les serveurs internes et externes qui sont référencés par le protocole Cascading Map Server. Ainsi, l'utilisateur du géorépertoire aura accès, sous forme de pages Web, non seulement aux données du serveur mais aussi à celle d'une fédération de serveurs référencés par le Cascading Map Server. La méthode *getcapabilities* retourne également un ensemble minimal de métadonnées qui seront intégrées au géorépertoire, mais n'offre pas de fonctions permettant de générer des imagerie. L'information retournée par une requête *getcapabilities* est en format XML. On trouvera la définition (DTD) de la structure XML dans le document disponible à l'adresse suivante : <http://www.opengis.org/techno/RFC13.pdf>.

Finalement, le troisième mécanisme vise à référencer les images pouvant être fournies par le serveur d'images au moyen d'une requête locale visant à inventorier l'ensemble de ces images. Cet inventaire permet de produire des imagerie, mais ne produit aucune métadonnée autre que la couverture spatiale de chaque fichier. Cette automatisation de l'inventaire offre deux avantages importants : premièrement, le temps consacré aux recherches menant à l'inventaire des données est considérablement réduit ; et deuxièmement, il est possible de programmer les recherches de sorte que le géorépertoire contienne toujours une liste à jour des documents disponibles. Notons finalement que les erreurs de saisie se trouvent ainsi éliminées.

3.3 Automatisation de la saisie des références aux données géospatiales

Les fichiers produits par l'une des trois méthodes d'inventaire automatique contiennent des références pour chaque jeu de données disponible sous la forme d'un URL. Ainsi, lors de l'intégration de l'information contenue dans ces fichiers d'inventaire dans la base de données du géorépertoire, toutes les références (URL) permettant d'accéder aux jeux de données seront automatiquement chargées dans cette base de données. La saisie des références (URL) aux données n'est donc plus une préoccupation pour l'administrateur du géorépertoire.

3.4 Génération d'une imagerie d'un document

Bien que l'un des objectifs premiers d'un ensemble de métadonnées soit de documenter les données d'un document et par ricochet d'informer un utilisateur potentiel de leur nature, elles sont parfois insuffisantes pour avoir une idée juste d'un document que l'on veut utiliser dans une application géomatique. Par exemple, une

image satellite ayant une couverture nuageuse de 20 % peut ne représenter aucun intérêt pour un utilisateur si la région d'intérêt est complètement cachée par les nuages, bien qu'ils ne représentent que 20 % de la superficie de l'image. L'imagerie est donc un moyen particulièrement efficace et performant qui permet de visionner un document avant son utilisation. La génération automatique d'imagerie d'un jeu de données, peu importe le format de la donnée, représente une solution intéressante pour compléter les métadonnées disponibles. L'application OGDI crawler et celle servant à référencer les images du serveur d'images offrent cette possibilité. Pour la recherche avec l'entrepôt de données, nous devons développer des techniques pour générer les imagerie, par exemple générer des images de 100 x 100 pixels pour chacun des jeux de données et les intégrer à la base de données du géorépertoire. Les imagerie sont efficaces pour représenter des photos aériennes ou des images satellite. Par contre, pour les fichiers vectoriels très denses, la taille réduite de l'imagerie ne permet pas toujours de visualiser correctement les données qu'elle représente.

3.5 Interface de saisie des métadonnées

Lorsque l'administrateur intègre le résultat de l'inventaire automatique des jeux de données au géorépertoire, toutes les informations déjà recueillies sont automatiquement placées dans les champs correspondants (voir figure 3). Toutefois,

Figure 3 : Exemple d'interface de saisie des métadonnées manquantes

pour que le géorépertoire soit fonctionnel, certaines métadonnées doivent être obligatoirement inscrites dans la base de données. Certaines de ces métadonnées ne sont pas disponibles aux outils d'inventaire automatique. Par conséquent, avant qu'une référence à un jeu de données ne soit chargée dans la base de données du géorépertoire, les métadonnées obligatoires devront être saisies par l'administrateur du

système. L'interface de saisie est conçue de manière à guider l'administrateur dans sa saisie de données. Il est aussi possible d'effacer des références à des jeux de données avec cette interface.

3.6 Considérations technologiques pour l'automatisation de l'inventaire des données

Peu de technologies géomatiques existantes permettent d'effectuer efficacement les opérations décrites précédemment. En effet, une connaissance préalable de chaque format de fichier et de sa structure est requise afin de rendre possibles les fonctions d'inventaire automatisé, d'extraction des métadonnées et de génération automatique des imageries. De plus, l'administrateur qui lance une recherche de documents ne connaît probablement pas *a priori* les données disponibles et, en ce sens, ne peut guider le module de recherche automatisée des données. Ce module doit donc être autonome pour la recherche d'information ; tout au plus, l'administrateur y spécifie ses paramètres de recherche de données (par exemple, le répertoire où sont situées les données).

L'environnement d'applications géomatiques à RDDC Valcartier est composé de logiciels OGDI développés par Global Geomatics, d'un entrepôt de données à référence spatiale développé par CubeWerx inc., d'un serveur d'images à haute résolution développé par HMR inc., de logiciels de modélisation et de simulation d'environnements 3D ainsi que de logiciels géomatiques courants (ArcView, MapInfo, GeoMedia, entre autres). Les données utilisées sont principalement de type militaire (DTED, VRF, RPF), mais comprennent aussi une large part de données provenant d'organismes civils. L'application de recherche de documents géospatiaux doit donc être en mesure d'inventorier l'ensemble des données utilisées dans ces systèmes.

Notre choix technologique a été fortement influencé par cet environnement lors de la phase initiale du développement du géorépertoire. Notre choix s'est porté sur l'application OGDI Crawler de Global Geomatics. Cette application permet d'automatiser la découverte de documents en recherchant localement les données disponibles sur les disques de l'ordinateur hôte. Basé sur l'OGDI, cet engin de recherche fonctionne avec un très grand nombre de formats de données, extrait automatiquement quelques métadonnées et génère automatiquement des imageries de chaque document retrouvé (en option). Les résultats retournés en format XML intègrent élégamment les métadonnées en s'appuyant sur le profil de métadonnées FGDC.

Cette solution s'avère efficace pour la majorité des documents exploités par les applications géomatiques à RDDC Valcartier. Toutefois, elle n'est pas conçue pour exploiter les données contenues sur un serveur WMS. Une solution exploitant la méthode *getcapabilities* permet d'automatiser l'extraction de l'information et permet aussi de référencer des données provenant de serveurs externes, si ces derniers

respectent le protocole Web Mapping Service. Finalement, l'accès aux données relatives aux images et aux photos fournies par le serveur d'images de HMR, bien que possible par l'OGDI Crawler, sera assuré par un module du serveur d'images. En effet, ce module est optimisé pour créer très rapidement des imagerie pour l'ensemble des données disponibles, bien que la taille des images originales puisse dépasser 1 Go.

Ce sont donc les résultats provenant de l'OGDI Crawler ainsi que de l'extraction des données du serveur d'images et de l'entrepôt de données qui seront intégrées à la base de données du géorépertoire. Au besoin, l'administrateur du système peut ajouter lui-même les métadonnées manquantes à un document ou à une collection. Les prochaines sections décrivent les principales fonctions pour les clients du géorépertoire.

3.7 Fonctions de cartographie dynamique

Un géorépertoire ne saurait s'appeler ainsi s'il n'était pas appuyé par une interface cartographique permettant de représenter les couvertures spatiales des documents et collections, ainsi que la définition des zones d'intérêt pour les recherches de documents par l'utilisateur. L'interface cartographique doit donc offrir à la fois une représentation exacte du territoire ainsi que la capacité de définir graphiquement une zone d'intérêt sur ce territoire. Un lien direct doit pouvoir s'établir aussi entre cette interface cartographique et l'interface de recherche des métadonnées, de manière à construire la requête globale qui sera envoyée à la base de données du géorépertoire. Finalement, ce module cartographique doit se présenter sous la forme d'un composant s'intégrant à un environnement Web. Pour ce faire, notre choix s'est arrêté sur le composant Active-X MapFusion de Global Geomatics (basé sur l'OGDI), lequel répond à l'ensemble des exigences formulées. Un client Web, basé sur des composantes SVG, permettra aussi d'implanter ce type de fonctionnalités.

Cette composante offre des fonctions d'édition graphique avancée permettant entre autres, le dessin de polygones, de rectangles et de formes irrégulières. Il est également possible au moyen de cette composante de récupérer directement le contenu cartographique d'une source de données de type OGDI référencée dans le géorépertoire.

3.8 Définition de la zone d'intérêt

À notre connaissance, la très grande majorité des géorépertoires existants basent leur mode de définition de la zone d'intérêt sur le rectangle englobant minimal (*minimum bounding rectangle*). Principe fort connu et simple à implanter, la zone d'intérêt est

définie par les coordonnées des points supérieurs gauche et inférieur droit. Cette solution s'avère efficace pour la recherche de documents traditionnels, comme les cartes topographiques gouvernementales. Toutefois, si l'on désire retrouver des

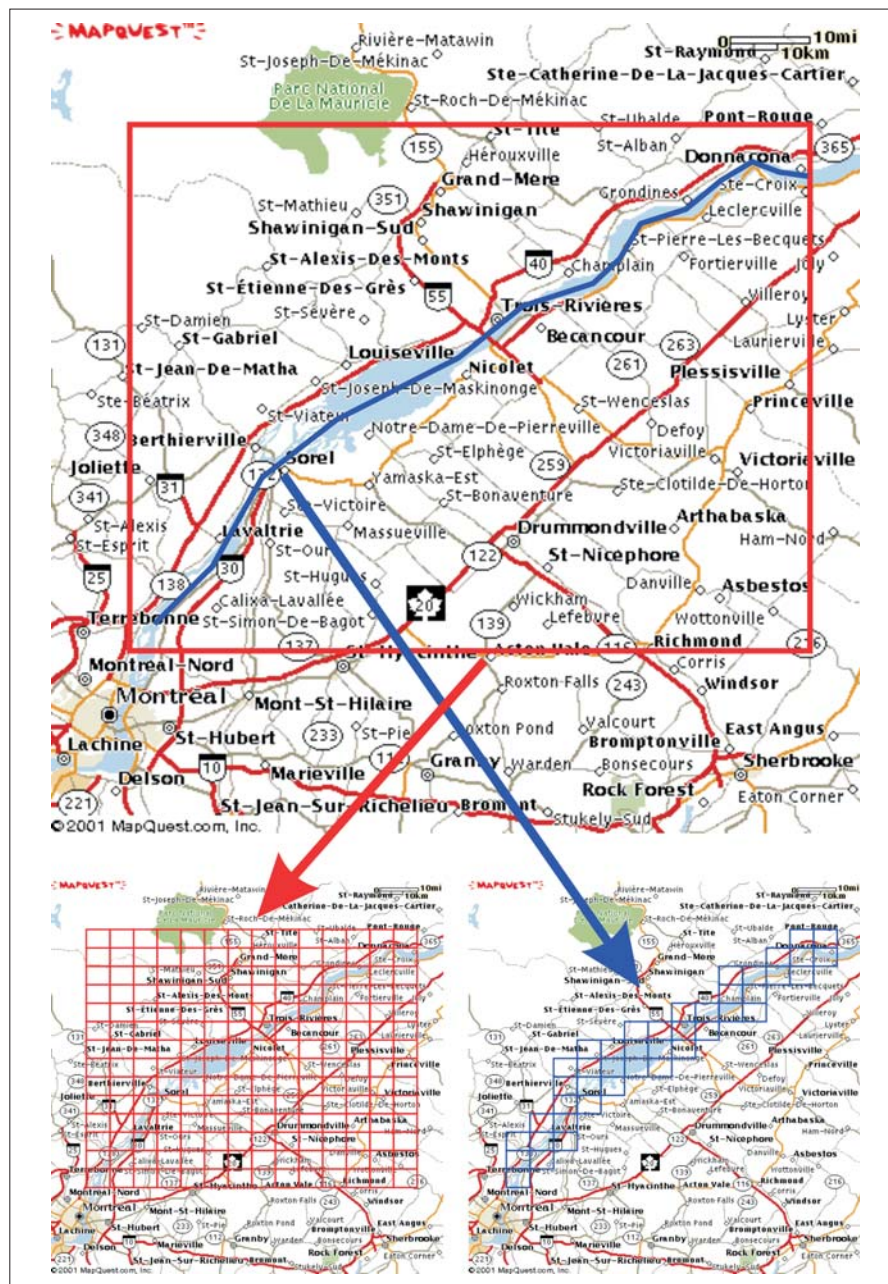


Figure 4 : Comparaison des modes de recherche de documents. En rouge, définition d'une aire de recherche et résultats obtenus (méthode traditionnelle) En violet, définition d'une polygone de recherche et résultats obtenus (méthode proposée pour le géorépertoire)

documents suivant une diagonale ou une ligne irrégulière, par exemple pour trouver des orthophotos suivant le parcours d'une rivière, il est fort probable que le nombre de

documents retrouvés par la méthode de recherche conventionnelle sera beaucoup trop élevé (voir figure 4). En développant une méthode de recherche basée sur un polygone irrégulier ou une polyligne, les résultats obtenus seront plus conformes à la requête de l'utilisateur. L'utilisation du composant MapFusion nous permettra de définir la zone d'intérêt, sous forme linéaire ou polygonale. Toutefois, un mécanisme d'analyse spatiale supplémentaire devra être développé directement dans la base de données afin d'obtenir les résultats souhaités lors de requêtes polygonales complexes. Dans le but de simplifier les calculs, seule l'analyse spatiale comportant des lignes diagonales, un point ou un rectangle, est implantée dans la première version du géorépertoire.

3.9 Requêtes spatio-temporelles et descriptives

L'une des fonctions majeures d'un géorépertoire consiste à offrir une interface de requête où l'on indique l'intervalle temporel recherché et où l'on choisit certaines caractéristiques associées aux métadonnées, comme l'échelle de prise de données, la projection, le format du fichier, etc. Cette interface entre l'utilisateur et la base de données est assurée par une interface de définition des requêtes. L'utilisateur n'a donc pas à apprendre le langage d'interrogation de la base de données pour effectuer des recherches. L'interface de l'application Windows offre beaucoup plus de souplesse pour la définition des requêtes descriptives et spatiales que l'interface de l'application Web (voir figure 5), en raison notamment des capacités offertes par les formulaires du SGBD Corel Paradox.

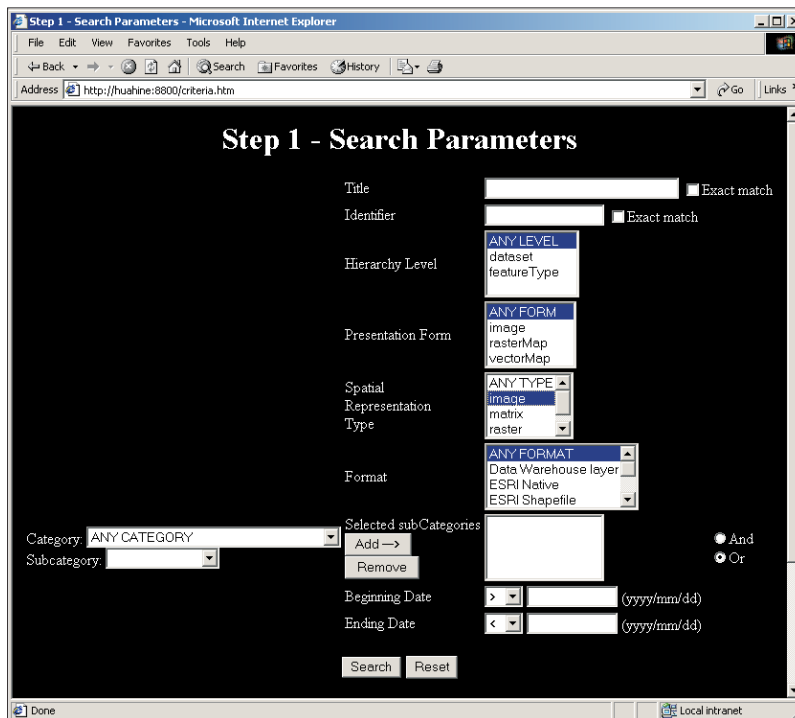


Figure 5 : Exemple de requête descriptive, interface Web

3.10 Enregistrement des requêtes, récupération des requêtes

Le système permet d'enregistrer les requêtes. L'utilisateur aura aussi la possibilité de récupérer des requêtes enregistrées et de les exécuter à nouveau. Cette fonction peut se révéler particulièrement utile dans un environnement de production dynamique de données, où il y a une haute fréquence de production de nouveaux documents, évitant ainsi aux utilisateurs de reconstruire leurs requêtes.

3.11 Accès direct aux données sous leur forme native ou par le Web

Dans les géorépertoires traditionnels, la recherche de l'information représente la finalité de l'application. Lorsque l'utilisateur obtient les résultats de sa requête (voir figure 6), il doit faire les démarches pour se procurer les données. Dans notre cas, le géorépertoire peut servir à cette fin mais offre aussi la possibilité d'aller plus loin en construisant automatiquement les codes d'accès aux données (par exemple, URL d'une couverture OGD). L'utilisateur peut utiliser ces codes pour accéder directement aux

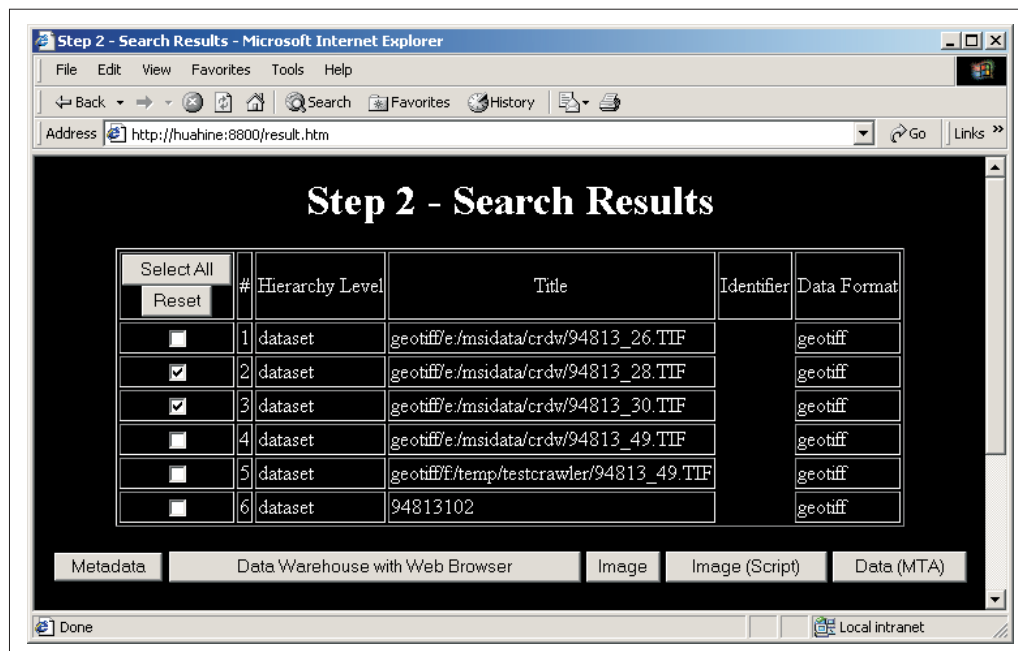


Figure 6 : Présentation des résultats d'une requête

données si ces dernières sont disponibles, soit par un serveur OGD soit par un autre mécanisme d'accès aux données tel que le *simple vector feature* ou en utilisant le protocole WFS, deux modes d'accès développés par l'OGC. Les méthodes d'accès aux données via le géorépertoire sont contextuelles, c'est-à-dire que, selon le type de données que l'utilisateur souhaitera consulter, il aura accès à toutes les méthodes

possibles pour visualiser ses données et y accéder (par exemple, lancer un SIG et visualiser les documents choisis lors de la recherche).

L'autre option offerte à un utilisateur consiste à coder dynamiquement l'URL permettant d'afficher dans un navigateur Web les données servies par un entrepôt de données ou un serveur d'images (voir figure 7), ou encore par un serveur OGD utilisant l'interface MapFusion. Cette approche a le mérite de faciliter l'accès aux données, représente une façon élégante d'y parvenir et ne nécessite pas l'emploi d'un logiciel autre qu'un navigateur Web. Il serait également possible de créer dynamiquement une requête *getmap* supportée par des serveurs se conformant à la spécification technique WMS du consortium OGC. Ainsi, nous pourrions combiner des données provenant d'un entrepôt de données, d'un serveur d'images et d'autres applications supportant cette interface. La spécification de la requête *getmap* peut être trouvée dans le document disponible à l'adresse <http://www.opengis.org/techno/RFC13.pdf>.

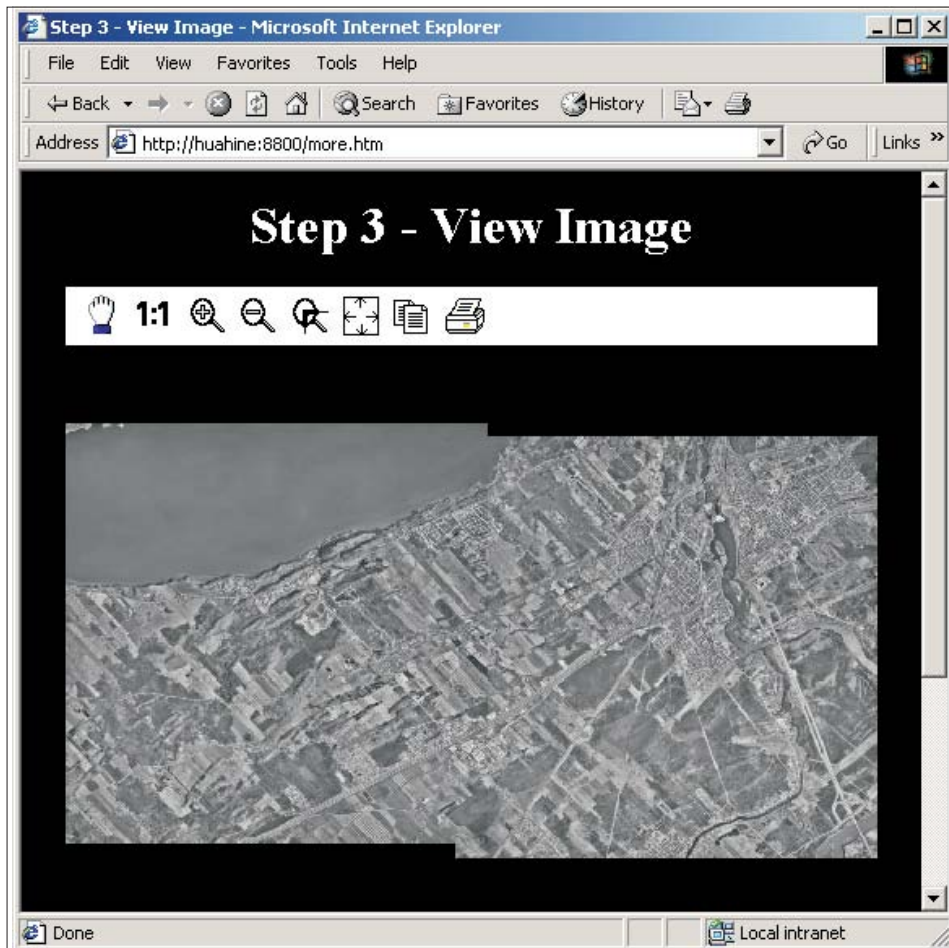


Figure 7 : Accès direct aux données, version Web

4 Environnement de développement du géorépertoire

Ce géorépertoire est réalisé avec un objectif bien précis, qui a guidé les choix technologiques. Notre ligne directrice a été de choisir des environnements de développement permettant à la fois de consulter le géorépertoire en utilisant soit un navigateur Web (mais avec des fonctionnalités réduites) soit une application légère, fonctionnant dans l'environnement Windows. Le géorépertoire doit supporter plusieurs utilisateurs (au minimum 10 usagers concourants), pouvoir fonctionner en réseau ou fonctionner sous un environnement PC mobile (sur un ordinateur portable, par exemple), offrir une interface Web fonctionnant avec les principaux navigateurs et être distribuable gratuitement ou sans devoir acheter une licence du SGBD pour l'exploiter. Notre choix s'est donc porté sur le SGBD Corel Paradox Developer's Edition 9 (qui permet de créer des applications distribuables gratuitement) et l'environnement de développement Borland Delphi 5, lequel est aussi un environnement de développement utilisé dans plusieurs projets à RRDC Valcartier. Ces deux produits interagissent très bien entre eux, partagent le même engin de base de données (le BDE de Borland) et permettent de développer une gamme complète de géorépertoires, en offrant notamment la possibilité d'inclure des composants Active-X ainsi que de publier le contenu de base de données sur Internet. L'utilisation de Paradox comme serveur Web a permis une très grande souplesse dans la définition des requêtes ainsi que de bonnes performances, compte tenu de la complexité du modèle de données implanté. De plus, Paradox possède, via son langage ObjectPAL, toutes les méthodes requises pour construire un site dynamique intégrant élégamment les modules Active-X, JavaScript et HTML permettant d'accéder aux données. Les formulaires de saisie des données ont aussi été créés en utilisant Paradox.

En utilisant ces deux environnements de développement, nous sommes en mesure d'offrir à l'utilisateur un géorépertoire complet sous la forme d'une application autonome Windows et d'un géorépertoire sur le Web basés sur une seule base de données. Nous offrons donc la souplesse et la richesse des fonctions avec l'application Windows et l'aspect interfonctionnement avec la version Web du géorépertoire.

5 Nouvelles perspectives de développement des géorépertoires

Les travaux décrits précédemment ont fait référence à des implantations classiques de géorépertoires, conceptuellement semblables aux cartes-index utilisées depuis plusieurs années, mais avec l'interactivité qu'apporte un système géomatique dynamique et couplé à une base de données. Depuis 2002, de nouvelles approches ont été explorées pour le développement de géorépertoires. Ces approches s'inspirent parfois d'outils de partage de données tels que Napster ou encore, intègrent de nouveaux concepts de traitement et d'analyse de l'information tels que les technologies OLAP géospatiales (GEOLAP). RDDC Valcartier a exploré quelques-uns de ces concepts au cours des derniers mois. Ces projets ne sont pas encore complétés et feront l'objet de mémos techniques qui leur seront propres. Nous jugions toutefois opportun d'intégrer des informations se rattachant à ces projets au présent mémo technique afin de brosser un tableau plus complet de la situation.

5.1 L'intégration de la technologie OLAP dans les géorépertoires

Le géorépertoire est un outil approprié pour effectuer une recherche afin de trouver des documents cartographiques répondant à certains critères. Les requêtes et analyses sont habituellement supportées par des requêtes SQL, qui ont prouvé leur efficacité dans l'exécution de requêtes typiquement transactionnelles. Toutefois, le langage SQL n'est pas approprié pour effectuer des analyses complexes plus orientées vers l'aide à la décision. Les technologies OLAP, avec leur structure multidimensionnelle et leur capacité d'agréger l'information, sont beaucoup plus souples et performantes. Le grand dictionnaire terminologique (<http://www.granddictionnaire.com>) définit ainsi la technologie OLAP : *“Technique informatique d'analyse multidimensionnelle, qui permet aux décideurs, en entreprise, d'avoir accès rapidement et de manière interactive à une information pertinente présentée sous des angles divers et multiples, selon leurs besoins particuliers”*. Bien que cette technologie ait été développée principalement pour soutenir les utilisateurs d'applications analytiques d'entreprises, les concepts qu'elle sous-tend sont adaptables au monde de la géomatique, et ce avec les mêmes bénéfices. Il faut toutefois consacrer un effort important pour intégrer les données sources aux bases de données multidimensionnelles.

L'intérêt d'intégrer les technologies OLAP aux géorépertoires, outre la réalisation de requêtes et analyses plus complexes, est de développer un volet décisionnel complémentaire aux fonctions de base d'un géorépertoire. Le géorépertoire peut devenir, outre un outil de recherche de l'information, un outil de gestion et d'analyse

d'inventaire cartographique en intégrant des fonctions, telles que l'analyse du volume des demandes par région géographique, ou encore la gestion des non-conformités des produits.

La grande force des systèmes OLAP est leur capacité à synthétiser et à agréger de l'information, à présenter simultanément différentes vues sur les données, à effectuer du forage sur les données et dans le cas de la technologie GEOLAP, à ajouter un contexte cartographique qui est en phase avec le contexte de l'analyse. Ainsi, un phénomène représenté sur une carte (par exemple, le volume de données demandées dans un espace géographique) peut aussi être représenté sous forme de tableau ou de graphique. Si la région géographique sur laquelle porte l'analyse change, automatiquement les données sont ajustées dans les tableaux et graphiques, et vice versa. Les analyses que l'on peut en tirer se trouvent ainsi à être renforcées par les multiples moyens utilisés pour représenter l'information.

Le projet GEOLAP vise donc à explorer certains aspects de l'intégration des fonctions décisionnelles dans un géorépertoire, le tout alimenté par un moteur OLAP d'analyse des données. En collaboration avec le Service cartographique des Forces (S Carto) et le Centre d'information topographique - Sherbrooke, RDDC est à mettre au point une première version de l'application. Ces deux organismes agissent à titre d'experts du domaine, RDDC Valcartier étant le maître d'oeuvre des travaux, en collaboration avec la firme Syntell de Québec. Les résultats des travaux de consultation et de définition des besoins ont permis de construire une maquette de l'application. Cette maquette (voir figure 8) a été produite afin d'illustrer les concepts et de les valider avant de procéder au développement du système.

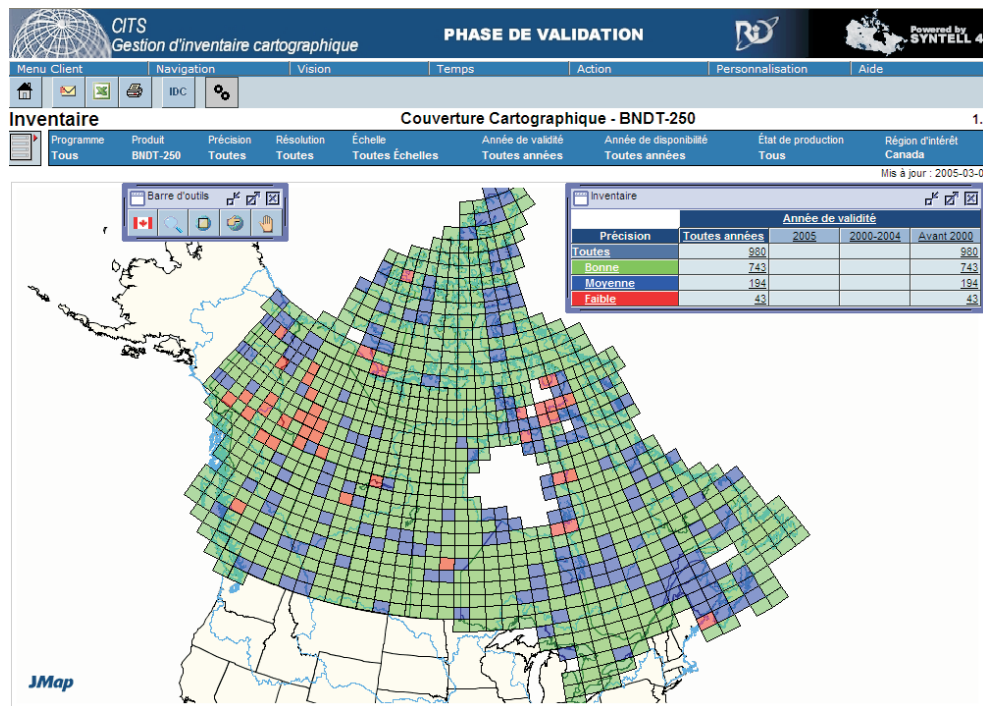


Figure 8 : Module d'analyse de la couverture cartographique, prototype GEOLAP

5.2 Le géorépertoire distribué

L'implantation traditionnelle d'un géorépertoire consiste à créer une base de données reliée à une organisation particulière et à permettre à différents utilisateurs de se connecter à cette base de données afin d'effectuer des recherches ou des analyses. Il est toutefois possible d'entrevoir des implantations différentes, répondant à des besoins précis. Par exemple, lors d'une crise environnementale ou d'une catastrophe naturelle (par exemple, la crise du verglas), un grand nombre d'intervenants, appartenant à différentes organisations, sont appelés à interagir. De nombreuses cartes, photos et images satellites seront produites pour gérer l'événement. Toutefois, en raison des différents systèmes utilisés, formats de données et autres considérations de partage de l'information, il est très difficile de mettre sur pied un géorépertoire intégrant les métadonnées produites par ces différentes organisations. Le partage des données et, par conséquent, l'accès aux meilleures données se trouve grandement complexifié. La solution que RDDC Valcartier propose à ce problème est la mise sur pied d'un géorépertoire décentralisé, implanté selon une architecture poste à poste. En utilisant la technologie de l'OGDI - qui offre un accès transparent à une multitude de données militaires et civiles, peu importe le système de coordonnées et la projection - et en la couplant à un serveur de données WMS, il devient alors possible d'offrir, à une multitude d'utilisateurs appartenant à différentes organisations, un accès direct aux données partagées.

Le concept de ce géorépertoire distribué est relativement simple. Il s'agit tout d'abord d'offrir à chaque participant un serveur OGDI couplé à une interface WMS. Le serveur OGDI détecte automatiquement toutes les données géospatiales qu'un utilisateur souhaiterait partager. Les métadonnées sont extraites des données détectées et sont ensuite publiées automatiquement, sans intervention de la part de l'utilisateur, dans une base de données centrale. L'utilisateur doit ensuite se connecter à la base de données pour spécifier certains paramètres (nom du fichier, description, quel groupe a les droits d'accès), ce qui rend alors les données accessibles aux autres utilisateurs. Ces opérations concernent les utilisateurs qui désirent publier leurs données dans la communauté. Comme certaines données peuvent être de nature sensible et ne pas être destinées à n'importe qui, il importe alors de gérer les droits d'accès, soit sur une base individuelle, soit sur la base de groupes qui détiennent des droits spécifiques concernant l'accès aux données.

Les utilisateurs qui souhaitent accéder aux données n'ont qu'à effectuer des requêtes sur la base de données centrale, laquelle permet de structurer les résultats d'une recherche de manière à fournir des liens directs aux données qui résident sur les postes des utilisateurs partageant leurs données. Un mécanisme d'alerte, utilisant des requêtes enregistrées sur le serveur, est en mesure d'informer un utilisateur que de nouvelles données répondant aux critères de sa requête sont disponibles dans la communauté. Encore une fois, le message d'alerte est structuré de façon à permettre l'accès direct aux données. L'accès aux données se fait de deux façons, soit directement en utilisant le protocole GLTP supporté par l'OGDI, soit en utilisant le serveur WMS. Le protocole GLTP permet de se connecter aux données en utilisant une application Windows (de type "map viewer"), tandis que le serveur WMS convient aux applications Web et aux différents clients négociant avec ce protocole d'accès. La

figure 9 illustre l'architecture de la solution proposée. Il est intéressant ici de constater que certaines des fonctionnalités de ce système (détection automatique des données, accès direct aux données) ont été développées dans le cadre des premiers projets de géorépertoire à RDDC Valcartier.

Architecture GARNET

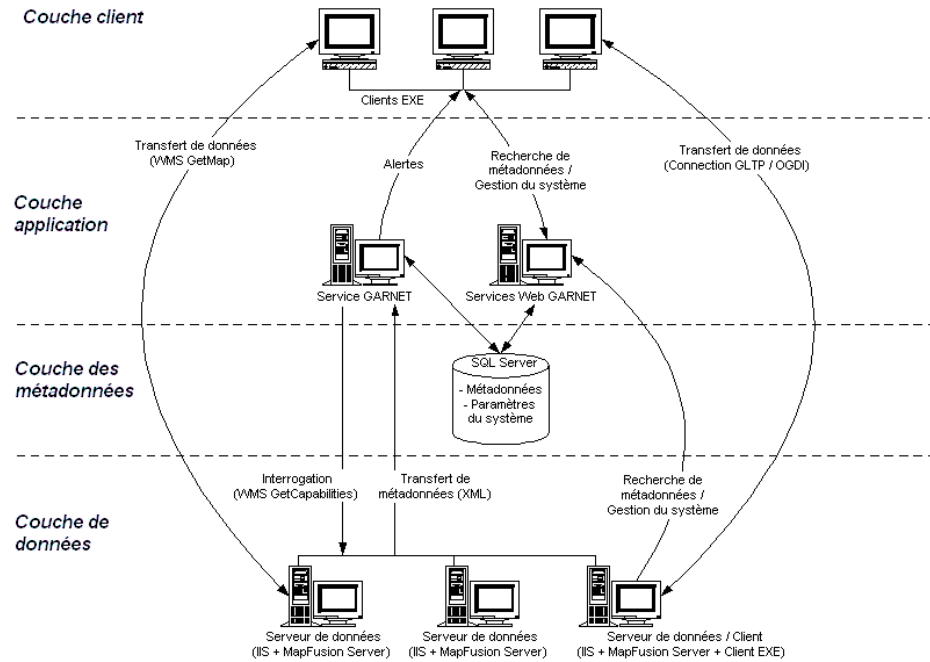


Figure 9 : Architecture proposée pour le géorépertoire distribué

6 Conclusion

Le géorépertoire ne doit plus être uniquement considéré comme une application permettant seulement de trouver des documents ou des collections de données à caractère spatial. Trois améliorations significatives au concept de géorépertoire sont présentées dans ce document. Il s'agit de l'automatisation de l'inventaire et de la saisie des métadonnées, de la définition de zones d'intérêt complexes (point, polyligne, polygone irrégulier) et l'accès direct aux données en mode natif ou sous forme de pages Web.

Ces améliorations ont été finalement démontrées dans des applications évoluées de géorépertoires qui sont en cours de développement à RDDC Valcartier. L'intégration des technologies OLAP dans les géorépertoires permettra à des gestionnaires d'inventaires cartographiques de mieux analyser l'utilisation qui en est faite, ainsi que d'identifier des priorités de mise à jour ou de développement de nouveaux produits. D'autre part, la mise sur pied du géorépertoire distribué, basé sur une architecture poste à poste, permet de constituer un environnement de géorépertoire en quelques minutes et offre des capacités d'intégration très rapide de données hétérogènes.

Le géorépertoire est de moins en moins considéré comme une application ayant sa propre finalité. Il s'inscrit dans le grand courant d'intégration des applications géomatiques et présente les fonctionnalités d'un portail vers un ensemble de données à caractère spatial. Pierre angulaire de toute infrastructure de données à référence spatiale, il offre l'occasion d'améliorer le processus de découverte et d'accès aux données géospatiales, et représente pour les utilisateurs un mécanisme efficace qui réduit grandement la difficulté associée à la recherche et à l'accès aux données.

Cette page est laissée intentionnellement en blanc

Bibliographie

- 1 Ressources Naturelles Canada (2000), Site Web CEONet, <http://ceonet.ccrs.nrcan.gc.ca>
- 2 Open GIS consortium (OGC) : <http://www.opengis.org>
- 3 GlobalGeomatics, site web : <http://www.globalgeo.com>
- 4 CubeWerx, site web : <http://www.cubewerx.com>
- 5 HMR / Bentley : site web : <http://imaging.bentley.com>
- 6 Gouin Denis., P. Morin, G. Clément., C. Larouche. (1997). Solving the Geospatial Data Barrier. *Geomatica*, Vol. 51, num. 3, 1997, pp 278-287.
- 7 Létourneau François, (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
- 8 Létourneau François., Y. Bédard, B. Moulin (1998). Perspectives d'utilisation du concept d'entrepôt de données pour les géorépertoires sur Internet. *Geomatica*, Association canadienne des sciences géomatiques, vol. 52, no. 2, pp. 145-163
- 9 Maamar Zakaria, B. Moulin, Y. Bédard (1999). Software agent_oriented frameworks for the interoperability of georeferenced digital libraries on the World Wide Web The SIGAL project. Book chapter in Goodchild, M.F., M.J. Egenhofer, R. Fegeas, and C.A.Kottman, editors *Interoperating Geographic Information Systems*. Boston Kluwer Academic Publishers. pp. 335-355.

Cette page est laissée intentionnellement en blanc

Liste de distribution

Interne

- 1 Directeur général
- 3 Bibliothèque des documents
- 1 François Létourneau
- 1 Yves van Chestein
- 1 Denis Gouin
- 1 Christian Martel
- 1 Marc-André Morin
- 1 Réjean Lebrun

Externe

- 1 DSIGRD
- 1 Paul Morin, J2 IM, 615 Booth St., Ottawa, ON, K1A OE9
- 2 Pierre Simard, MCE, 615 Booth St, Ottawa, ON, K1A OE9
- 1 Dr Yvan Bédard, Centre de recherche en géomatique, Pavillon L.-J. Casault, Université Laval, Québec, QC, G1K 7P4

SANS CLASSIFICATION
COTE DE SÉCURITÉ DE LA FORMULE
(plus haut niveau du titre, du résumé ou des mots-clefs)

FICHE DE CONTRÔLE DU DOCUMENT

1. PROVENANCE (le nom et l'adresse) RDDC Valcartier, section GIC 2459 boul Pie-XI Nord, Québec (QC) G3J1X5		2. COTE DE SÉCURITÉ (y compris les notices d'avertissement, s'il y a lieu) Sans classification	
3. TITRE (Indiquer la cote de sécurité au moyen de l'abréviation (S, C, R ou U) mise entre parenthèses, immédiatement après le titre.) Le géorépertoire comme mode de recherche et d'accès direct aux données à référence spatiale (U)			
4. AUTEURS (Nom de famille, prénom et initiales. Indiquer les grades militaires, ex.: Bleau, Maj. Louis E.) LÉTOURNEAU, François			
5. DATE DE PUBLICATION DU DOCUMENT (mois et année) 2007		6a. NOMBRE DE PAGES 27	6b. NOMBRE DE REFERENCES 9
7. DESCRIPTION DU DOCUMENT (La catégorie du document, par exemple rapport, note technique ou mémorandum. Indiquer les dates lorsque le rapport couvre une période définie.) Mémorandum technique			
8. PARRAIN (le nom et l'adresse)			
9a. NUMÉRO DU PROJET OU DE LA SUBVENTION (Spécifier si c'est un projet ou une subvention) 15AL01		9b. NUMÉRO DE CONTRAT	
10a. NUMÉRO DU DOCUMENT DE L'ORGANISME EXPÉDITEUR TM-2004-376		10b. AUTRES NUMÉROS DU DOCUMENT N/A	
11. ACCÈS AU DOCUMENT (Toutes les restrictions concernant une diffusion plus ample du document, autres que celles inhérentes à la cote de sécurité.) <input checked="" type="checkbox"/> Diffusion illimitée <input type="checkbox"/> Diffusion limitée aux entrepreneurs des pays suivants (spécifier) <input type="checkbox"/> Diffusion limitée aux entrepreneurs canadiens (avec une justification) <input type="checkbox"/> Diffusion limitée aux organismes gouvernementaux (avec une justification) <input type="checkbox"/> Diffusion limitée aux ministères de la Défense <input type="checkbox"/> Autres			
12. ANNONCE DU DOCUMENT (Toutes les restrictions à l'annonce bibliographique de ce document. Cela correspond, en principe, aux données d'accès au document (11). Lorsqu'une diffusion supplémentaire (à d'autres organismes que ceux précisés à la case 11) est possible, on pourra élargir le cercle de diffusion de l'annonce.)			

SANS CLASSIFICATION
COTE DE LA SÉCURITÉ DE LA FORMULE
(plus haut niveau du titre, du résumé ou des mots-clefs)

SANS CLASSIFICATION

COTE DE LA SÉCURITÉ DE LA FORMULE
(plus haut niveau du titre, du résumé ou des mots-clefs)

13. SOMMAIRE (Un résumé clair et concis du document. Les renseignements peuvent aussi figurer ailleurs dans le document. Il est souhaitable que le sommaire des documents classifiés soit non classifié. Il faut inscrire au commencement de chaque paragraphe du sommaire la cote de sécurité applicable aux renseignements qui s'y trouvent, à moins que le document lui-même soit non classifié. Se servir des lettres suivantes: (S), (C), (R) ou (U). Il n'est pas nécessaire de fournir ici des sommaires dans les deux langues officielles à moins que le document soit bilingue.)

La découverte de l'information géospatiale numérique représente encore de nos jours une tâche qui requiert des efforts importants. Avec l'avènement des données numériques, une partie de cette tâche se trouve à être facilitée, notamment en raison de l'automatisation de l'inventaire des données et de la semi-automatisation de la saisie des métadonnées. Le développement des interfaces OGD, WMS et WFS permet désormais un accès direct aux données. Les géorépertoires se trouvent donc à être enrichis de ces fonctionnalités et offrent désormais un inventaire plus exhaustif de données, de meilleures métadonnées et un accès direct aux données, pour fins de visualisation et d'exploitation. De nouvelles technologies, dont OLAP géospatial et peer-to-peer permettent d'envisager de nouvelles fonctionnalités pour les géorépertoires.

Ce document présente en premier lieu un historique des géorépertoires, souligne leur importance dans le processus de découverte de données et fait état des fonctionnalités du géorépertoire intégrateur développé dans la période 2000 - 2003 à RDDC Valcartier et enrichi en 2004.

14. MOTS-CLÉS, DESCRIPTEURS OU RENSEIGNEMENTS SPÉCIAUX (Expressions ou mots significatifs du point de vue technique, qui caractérisent un document et peuvent aider à le cataloguer. Il faut choisir des termes qui n'exigent pas de cote de sécurité. Des renseignements tels que le modèle de l'équipement, la marque de fabrique, le nom de code du projet militaire, la situation géographique, peuvent servir de mots-clés. Si possible, on doit choisir des mots-clés d'un thésaurus, par exemple le "Thesaurus of Engineering and Scientific Terms (TESTS)". Nommer ce thésaurus. Si l'on ne peut pas trouver de termes non classifiés, il faut indiquer la classification de chaque terme comme on le fait avec le titre.)

Géorépertoire

Découverte des données géospatiales

Inventaire cartographique automatisé

OGD

Spécifications WMS, WFS

SANS CLASSIFICATION

COTE DE SÉCURITÉ DE LA FORMULE
(plus haut niveau du titre, du résumé ou des mots-clefs)

Defence R&D Canada

Canada's Leader in Defence
and National Security
Science and Technology

R & D pour la défense Canada

Chef de file au Canada en matière
de science et de technologie pour
la défense et la sécurité nationale



WWW.drdc-rddc.gc.ca

