

Interface cartographique pour l'exploration multidimensionnelle des indicateurs de santé environnementale sur le Worl Wide Web / Cartographic Interface for Multidimensional Exploration of Environmental Health Indicators on the World Wide Web

Dr Rachel Catelan et Dr Yvan Bédard

Université Laval, Centre de recherche en géomatique
Rachel.catelan@scg.ulaval.ca ; yvan.bedard@scg.ulaval.ca

Dr Pierre Gosselin

Centre hospitalier universitaire de Québec, Département de santé publique
pgosselin@cspq.qc.ca

Aperçu biographique / Biographical Summary

Dr Catelan is presently a post-doctoral researcher at Laval University's Centre for Research in Geomatics, with 7 years of experience in GIS. Her present work involves database management, data warehousing and Spatial Olap for Environmental Health applications running on the World-Wide-Web (intranet). Before joining the research team of Dr. Yvan Bedard, Dr Catelan acquired experience in both fundamental and applied research on government agencies and private industries projects in France and abroad (Canada, Costa Rica...). She acquired knowledge in geography, cartography, social science, environment, agriculture, regional planning public health, commercial databases and commercial spatial analysis tools. Results of her work has been published in her Ph.D theses, in national and international papers and presented at scientific and more general conferences.

Résumé / Abstract

To reduce and guard against health risks of environmental origin requires easy and rapid access to high quality statistics which must be analyzed in a useful manner to support decisions and interventions. In this aim and to attain their management objectives, several public health organizations wish to acquire a Geographic Information Systems (GIS) application.

In spite of this recent coordination of medicine, environment, and geomatics is of major importance, it remains that the sole use of a GIS is only at first step. It is widely acknowledged that traditional GIS are limited in multi-scale analysis (e.g. local vs regional), spatial statistics, handling temporal indices (historical and predictive) as they were not conceived initially for such operations. Such functions require considerable statistical manipulation and development of complex programming with a GIS. Thus, it is currently impossible for a non specialist in GIS to rapidly and easily analyze nor navigate through environmental health data, especially when such usage requires the comprehension of a phenomenon's development over a given period and to compare several themes of great complexity.

This project focuses on maximal exploitation of the Data Marts, OLAP and Data Mining tools in conjunction with GIS, will facilitate temporal and multi-level treatment of spatial statistics thanks to the multidimensional data base approach used in Data Marts. It will also allow rapid and simple data processing assisted by a SOLAP-type (Spatial OLAP) cartographic interface. Results obtained will lead to a solution closer to the desired one; that is, to have an efficient information system for the description and analysis of complex environmental health data. Finally, to make this tool available to several users located in various regions, at a reasonable cost, the use of an Intranet as a developmental support will be favoured. This applied research project includes the development of new knowledge introduced by the participation of the University of McMaster team that will permit access to high tech expertise and to Ontario data bases on atmospheric pollution, morbidity, and medicine usages related to asthma and other respiratory diseases, as well as expertise concerning environmental health indicators.

The focus of this paper is to describe:

- the problematic concerns of this project and our research objectives;
- the means used to reach in terms of human resources coming from three universities, the technologies applied, and the data collected from various different sources;
- the research method proposed; and
- a simulation of results expected.

Introduction

Réduire et prévenir les risques à la santé qui sont d'origine environnementale nécessite un accès simple et rapide à des données de bonne qualité, lesquelles doivent être analysées de façon utile pour supporter les décisions et interventions. Les problèmes auxquels font face les gestionnaires en environnement et en santé publique sont extrêmement variés. Les sources de contaminants peuvent être locales, régionales ou planétaires et se mesurer dans plusieurs vecteurs ou chez l'humain. Plusieurs constatent que l'accès aux données existantes doit être amélioré et que des mécanismes permettant une meilleure utilisation sur une base spatiale doivent être développés (Lepper 1995; WHO 1996). Partant de ces constats, notre équipe pluridisciplinaire composée de chercheurs spécialisés en santé environnementale (SE), en géomatique et dans le développement de méthodes orientées objet et agent logiciel, s'est engagée, dans le cadre d'un projet de recherche appliquée en géomatique de venir en aide aux intervenants en SE. Nous proposons de concentrer nos efforts à la réalisation d'un système d'analyse descriptive des données complexes de SE à l'aide d'une interface cartographique simple et rapide (inspiré des approches OLAP et hyperliens).

Les objectifs, les moyens mis en oeuvre et les résultats attendus de ce projet de recherche appliquée sont décrits en première partie du présent document. La seconde partie servira, au contraire, à introduire l'outil logiciel final visé: présenter son interface graphique et l'essentiel des fonctionnalités.

Présentation du projet

Contexte

Pour atteindre leurs objectifs de gestion, plusieurs organisations en santé publique désirent acquérir un logiciel SIG (Système d'Information Géographique) leur permettant de géoréférencer leurs données (Gosselin P., Bédard Y., Elliott S., Jerrett Mike, R.Catelan, Poitras Ph., Gingras A., 2000). Il en résulte une capacité accrue d'analyse de ces observations, que ce soit par la simple production de cartes thématiques, pour mieux illustrer un phénomène, ou par l'utilisation de fonctions d'analyse spatiale (ex. opérations d'adjacence à un site polluant, distribution spatiale par bassin versant, corrélation spatiale). Quoique ce mariage entre médecine, environnement et géomatique soit récent, sa percée est d'une importance majeure pour l'identification des situations à risque et la communication de ces risques, ainsi que pour la protection de la santé publique.

Malgré cet avancement important, il n'en reste pas moins que la simple utilisation d'un SIG n'est qu'un premier pas (P.Gosselin, Y.Bédard, R.Catelan, 2000). En effet, il est fréquemment admis que les SIG sont limités dans les fonctions d'analyse multi-échelles (ex. analyse locale vs régionale), de statistique spatiale et de gestion des données temporelles (historiques et prédictives) car ils n'ont pas été conçus initialement pour de telles opérations. Lorsque l'on réussit à effectuer de telles opérations, ce n'est que suite à d'importantes manipulations de données et au développement de programmes complexes, ce qui exige beaucoup d'efforts et implique des temps de réponse très lents. De plus, les SIG demeurent inaccessibles aux non-spécialistes parce que trop complexes à utiliser (ex. requièrent SQL avec opérateurs spatiaux). Il en résulte que le spécialiste en santé environnementale doit consacrer d'importants efforts à apprendre les SIG ou doit s'adresser à son spécialiste des SIG, ce qui freine l'exploitation de la géomatique. Il est aujourd'hui impossible pour un non-spécialiste des SIG d'analyser *rapidement et facilement* les données de santé environnementale, surtout lorsqu'une telle analyse demande de comprendre l'évolution d'un phénomène sur une période donnée et de comparer plusieurs thèmes à différents niveaux de détail.

Objectifs

Ces constatations étant au cœur de notre problématique, nous proposons dans le cadre de ce projet de construire sur Intranet un système pour l'analyse descriptive des données complexes en santé environnementale à l'aide d'une interface cartographique simple et rapide basée sur l'approche multidimensionnelle des Data Marts et l'interface des outils OLAP.

Nous concentrons également nos efforts sur l'exploitation maximale des marchés de données (Data Marts) et des outils OLAP en conjonction avec les SIG. Ce projet, d'une durée de deux années, permettra ainsi de faciliter la gestion temporelle et la gestion multi-échelles des données spatiales grâce à l'approche *base de données multidimensionnelle* utilisée dans les Data Marts. Ce projet permettra également d'explorer ces données de façon très rapide et très simple grâce à une interface cartographique de type SOLAP (Spatial OLAP). Ce type d'interface nous apparaît très prometteuse puisque l'approche OLAP est déjà reconnue comme la solution la plus populaire pour les systèmes d'aide à la décision basés sur les données non-spatiales. Les résultats obtenus permettront donc de s'approcher de la solution désirée, soit: avoir un système d'information efficace pour la description et l'analyse des données complexes en santé environnementale. Evidemment, au cœur même de ce projet se retrouvent toutes les considérations relatives à l'adaptation de ces approches aux données spatiales.

Finalement, considérant le désir de rendre ce genre d'outil disponible à plusieurs utilisateurs situés dans différentes régions, et ceci à un coût acceptable, nous privilégions l'utilisation d'un Intranet comme support de développement.

Choix technologiques

L'architecture technologique définie de manière à parvenir à la réalisation d'un système permettant aux utilisateurs les moins expérimentés d'avoir accès aux données existantes est présentée ci-dessous.

Les données de nature très diverses (santé, environnementale, cartes, documents médiatiques, etc.) colligées de différentes sources (rapports, documents cartographiques, reportages, sites Web, etc.) seront regroupées dans un entrepôt de données (Data Warehouse). Ces données disponibles depuis le Web seront exploitées à partir de l'interface SOLAP (OLAP spatial) utilisée pour naviguer tant à travers les données géospatiales que pour la navigation dans le géorépertoire multimédia, entreprendre des requêtes spatiales et accéder aux méta-données. La présence du géorépertoire est fondamentale. Ce dernier sert à effectuer des requêtes spatiales, pour déterminer quelles données sont disponibles pour une portion de territoire déterminée. Des algorithmes utilisant la reconnaissance de texte et de parole sont développés pour la géoréférence automatique des données multimédia.

Finalement, il est important de noter que l'ensemble est donc une interface unique facile et rapide permettant simultanément de découvrir les données existantes et de naviguer dans celles-ci, si elles sont en format numérique.

Les choix technologiques qui seront utilisés pour la réalisation du prototype final visé ne sont pas actuellement totalement arrêtés. Toutefois, afin de satisfaire les besoins des deux extrémités du spectre des utilisateurs, nous considérons indispensable le développement des deux solutions technologiques suivantes:

- L'une visant à la réalisation d'un outil économique disposant d'un nombre restreint de fonctionnalités satisfaisant les attentes de la majorité des utilisateurs. L'exploration et les premiers tests réalisés à partir des outils logiciels Géomédia d'Intergraph, Powerplay de Cognos et de manière plus approfondie les produits de la société Softmap offrent des résultats satisfaisants qui nous amènent à croire en la réussite d'un tel outil ;
- L'autre solution, davantage orientée vers le besoin des utilisateurs les plus expérimentés, nécessitera le développement d'un outil plus lourd intégrant une multitude de fonctions d'analyse spatiale et de visualisation des données, dont la visualisation 3D. Les coûts de déploiement seront, évidemment, sans commune mesure avec la précédente solution. Le choix des outils logiciels qui serviront à la réalisation de ce prototype, au terme d'exploration des outils proposés sur le marché nous amène à considérer la technologie Oracle (Oracle 8i, Express Oracle) et l'outil de visualisation AVS proposé par la société de même nom parmi les outils les mieux adaptés à nos attentes.

Evidemment l'accessibilité depuis le web des prototypes résultants importe quelque soit l'orientation technologique considérée.

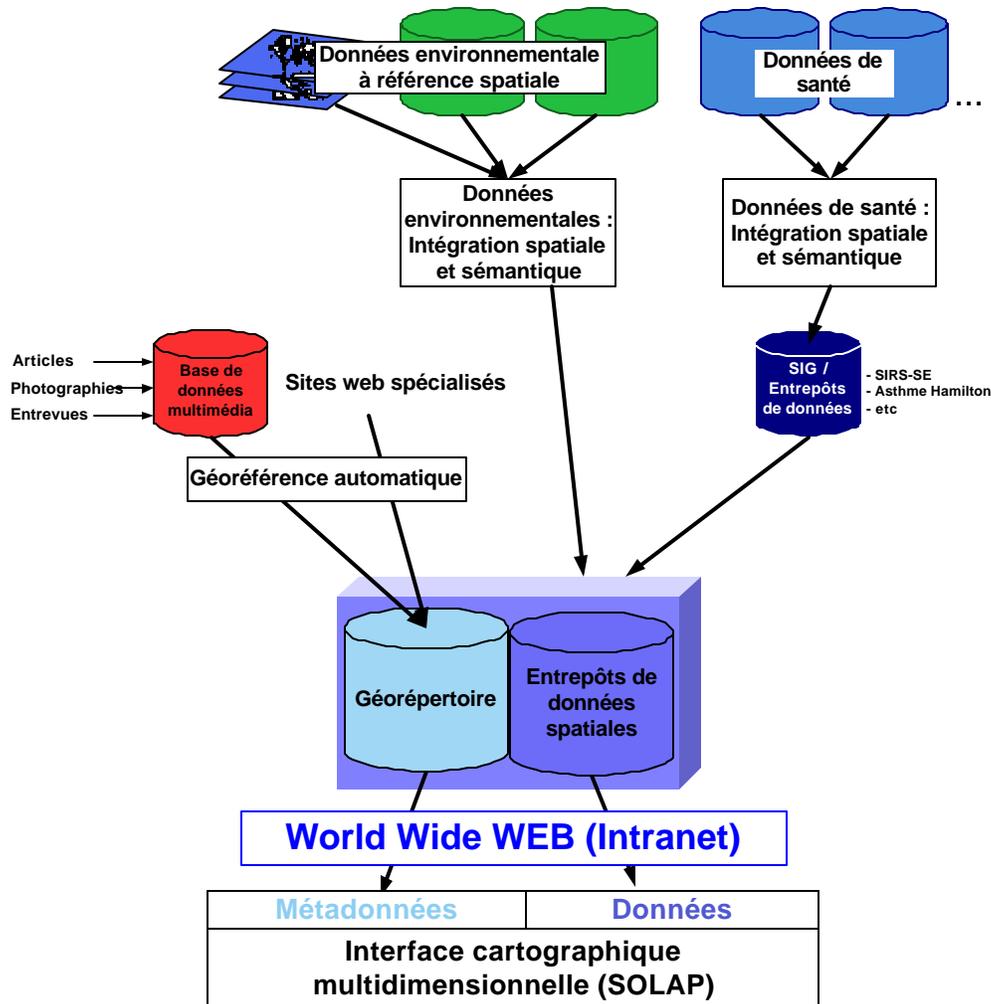


Figure 1: Architecture technologie d'ICEM/SE

Résultats et bénéfices attendus

Les résultats de cette recherche se traduiront donc par la mise en place d'un prototype fonctionnel de système de description et d'analyse des données complexes en santé environnementale (l'architecture graphique et l'essentiel des fonctionnalités sont présentés ci-dessous), basé sur le projet de Système d'information à référence spatiale en santé environnementale (SIRS-SE) du Comité de santé environnementale, ainsi que le développement de nouvelles connaissances grâce à l'expérimentation. Dans la mesure des ressources disponibles, cette expérimentation priorisera notamment certaines données générées dans le cadre du projet *Étude des variations spatio-temporelles de l'asthme au Québec* (Drs. A.Laberge, P.Lajoie) en fonction d'ententes réalisées entre Merck Frosst Canada et le CHUQ; certaines données ontariennes dont la base ISAAC (Dr M.Sears) sur la santé respiratoire des enfants du Sud de l'Ontario et les données environnementales et médicales pertinentes (Drs D.Pengelly, S.Elliott, D.Cole, P.Kanaroglou).

Les bénéfices d'un tel outil sont multiples. En effet, même si une grande part des données intégrées au prototype sont disponibles, elles ne sont cependant pas toutes accessibles (ou intégrées de manière uniforme) dans les Directions de Santé Publique. De plus, plusieurs de ces informations comportent une dimension spatiale qu'il est important de visualiser. Ainsi, à moyen et long terme, l'intégration des données utilisées dans le domaine de la santé environnementale dans un système d'information à référence spatiale contribuera notamment à :

- Assurer un accès convivial et opérationnel aux principales banques de données (sanitaires, socio-économiques, sociales, environnementales et administratives) aux intervenants en SE ;
- Uniformiser le traitement des données (validation des données de base, périodes, méthodes d'analyses statistiques) et de maintenir un haut niveau de qualité méthodologique et de crédibilité ;
- Produire une représentation cartographique des données concernant la mortalité et la morbidité, pour certains problèmes de santé potentiellement reliés à des problématiques environnementales ;
- Rendre accessible diverses informations sur les activités menées par les équipes en santé environnementale, et d'en suivre l'évolution dans le temps ;
- Réaliser, de manière économique, l'analyse régionale des caractéristiques environnementales, sociales et sanitaires ;
- La réalisation de bilans de l'état de santé de la population en rapport avec les risques biologiques, chimiques et physiques ;
- Répondre rapidement à des demandes courantes ;
 - Vérification sommaire d'agrégats géographiques (clusters) suspectés dans une région donnée ;
 - Étude des associations pour d'autres problèmes de santé qui présentent des déterminants ou des vecteurs communs ;
 - Réalisation de bilans thématiques.

Prototype : interface et fonctionnalités de l'outil logiciel

Pour les raisons évoquées ci-dessus, à savoir la multiplication des problèmes auxquels doivent faire face les gestionnaires en SE donc des facteurs et des sources de données de toutes natures à intégrer au système, la gestion des données à l'aide d'un géorépertoire est apparue très vite indispensable. L'intérêt de ce catalogue de documents géoréférencés (cartes, plans, photographies aériennes, images satellitaires, documents multimédias) est de permettre de :

- retrouver plus facilement les collections, documents et données qui existent ;
- mieux connaître leur nature (ex: source, formats, normes) ;
- leur qualité (ex: précision, accessibilité, légitimité) ;
- leur valeur (ex: utilisation, marché, bénéfices) ;
- de tels choix technologiques nous ont conduit, lors de la réalisation de la maquette du prototype, à proposer deux interfaces graphiques distinctes selon la nature des opérations engagées.

La première interface graphique proposée à l'ouverture d'ICEM/SE sert à la sélection et l'extraction de documents du géorépertoire (Voir figure 2). Comme on peut le lire sur la figure, l'utilisateur est invité à préciser, pour la sélection des documents recherchés, les quatre variables suivantes :

- la variable spatiale. L'utilisateur doit définir la zone géographique pour laquelle il souhaite réunir de l'information soit de manière interactive en sélectionnant sur la carte le territoire correspondant soit en entrant les coordonnées géographiques s'il détient une telle information ;
- la variable thématique. L'utilisateur doit déterminer le sujet de sa recherche (ex: le cancer, les maladies infectieuses...);
- le type de document. Cette information importe du fait de la multiplication des types de documents réunis dans le géorépertoire ;
- la variable temporelle. Cette information facultative offre, à l'utilisateur la possibilité de définir les dates limites au delà desquelles les documents n'auront plus d'intérêt par rapport à sa recherche.

Dès lors que ces dernières ont été définies, l'utilisateur peut exécuter la requête. Les résultats viennent s'afficher immédiatement sous forme de tableau en précisant pour chaque document extrait, comme on peut le voir sur la figure 2, son nom, le type du document, le territoire couvert et la date de publication du document ou la période de validité s'il s'agit de bases de données. Du fait de la diversité des types de documents réunis au sein du

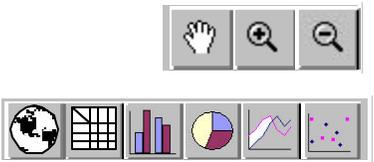
géorépertoire, l'utilisateur peut être amené, pour la consultation de ces documents, à manipuler divers exécutoires. Si pour l'interrogation de certains documents tels les films, les documents cartographiques... l'outil qui sera proposé n'est pas encore clairement défini voire sera laissé au libre choix de l'utilisateur, au contraire, pour le traitement des bases de données spatiales l'utilisateur sera invité à poursuivre son exploration de la base de données sélectionnée avec ICEM/SE et sera dirigé vers une nouvelle interface.

La figure 3 donne un aperçu de l'interface graphique et des fonctionnalités proposés pour l'exploration des données. Pour répondre aux attentes des intervenants en santé environnementale (leur attente d'outils efficaces, simples d'utilisation et rapides), nous avons considéré indispensable que les utilisateurs aient une représentation graphique de l'ensemble des données observées et puissent avoir une vision suffisamment générale de l'ensemble des données sémantiques se rattachant à la thématique étudiée. D'où l'importance de scinder l'écran en trois entités (voir figure 3) permettant d'avoir simultanément un affichage graphique des données explorées et de naviguer dans l'arbre hiérarchique des données sémantiques tout en gardant un oeil informatif sur les normes habituellement observées (fenêtre située dans l'angle droit au bas de l'écran).

Pour l'interrogation et l'exploration des données nous avons retenu essentiellement trois types d'outils:

- les outils de navigation dans les données sémantiques ;
- les outils de navigation spatiale ;
- les outils graphiques.

Le rôle joué par chacun d'eux est résumé dans le tableau ci-dessous.

<p>- les outils de navigation dans les données sémantiques</p> 	<p>Ils permettent d'effectuer des "drill up", "drill down" ou "drill across" thématiques. Ils servent pour naviguer au sein des données sémantiques. Exemple de "drill up" thématique: passage d'une consultation des données du cancer du sein chez les femmes âgées de plus de 15 ans dans la région de Québec à une consultation des données concernant l'ensemble des femmes pour un même territoire. (changement du niveau d'agrégation de l'information).</p>
<p>- les outils de navigation spatiale</p> 	<p>De même que précédemment les outils de navigation spatiale facilitent le changement de l'échelle géographique, tel que le passage d'un mode de représentation provincial à une représentation des données à l'échelle régionale (drill down spatial). Ils sont indispensables pour la navigation au sein des données spatiales.</p>
<p>- les outils graphiques</p> 	<p>Les outils de dessin servent comme dans la majorité des logiciels à gérer les données graphiques: déplacer l'image sur l'écran, agrandir/réduire l'affichage graphique à l'écran... Le second ensemble d'icônes jouent un rôle déterminant dans notre prototype. Ils facilitent la modification du mode de représentation graphique. Ils permettent de passer d'une représentation cartographique à une visualisation des données sous forme d'histogramme ou dans un tableau de données.</p>

L'utilisateur peut, à partir de ces quelques outils entreprendre une exploration approfondie des données. Il peut par exemple, après avoir extrait du géorépertoire une base de données (ex: la base de données du cancer de la région du Québec) et sélectionné le sous-ensemble de données qu'il souhaite explorer (ex: les données sur le cancer du sein), en donner une représentation graphique, faire varier l'échelle géographique (ex: passer de données régionales aux données de CLSC¹), modifier le mode d'affichage graphique (ex: passer d'une représenta-

¹ Centre Local de Service Communautaire

ICEM-SE project (Interface cartographique pour l'exploration multidimensionnelle des indicateurs de sante environnementale sur W3)

File Edit View Explore Format Window Help



Sélection de documents

Tapez votre recherche

Cancer

Types de données

- Tous types de données
- Base de données
- Base de données géospatiales
- Cartes
- Site Web
- Films
- Article scientifique
- Autres

Temporalité des données

Date limite inférieure:

Date limite supérieure:

Coordonnées géographiques

Longitude Min. Longitude Max.

Latitude Max. Latitude Min.

Documents trouvés

	Nom du document	Types de données	Territoire couvert	Date des données	Voir les données
1	Le cancer au Québec	Films	Québec	15 Déc 1998	
2	Base de données des cancers	Base de données géospatiales	Canada	1993-1997	
3	Les effets des BPC sur la santé	Article scientifique	Québec	Déc. 1995	
4	Cancer et autres maladies	Sites Web	Monde		
5	Carte des cancers au Québec	Carte	Région de Montréal	1996	
6					
7					

Figure 2 : Interface graphique d'ICEM/SE pour l'extraction des données du géorépertoire

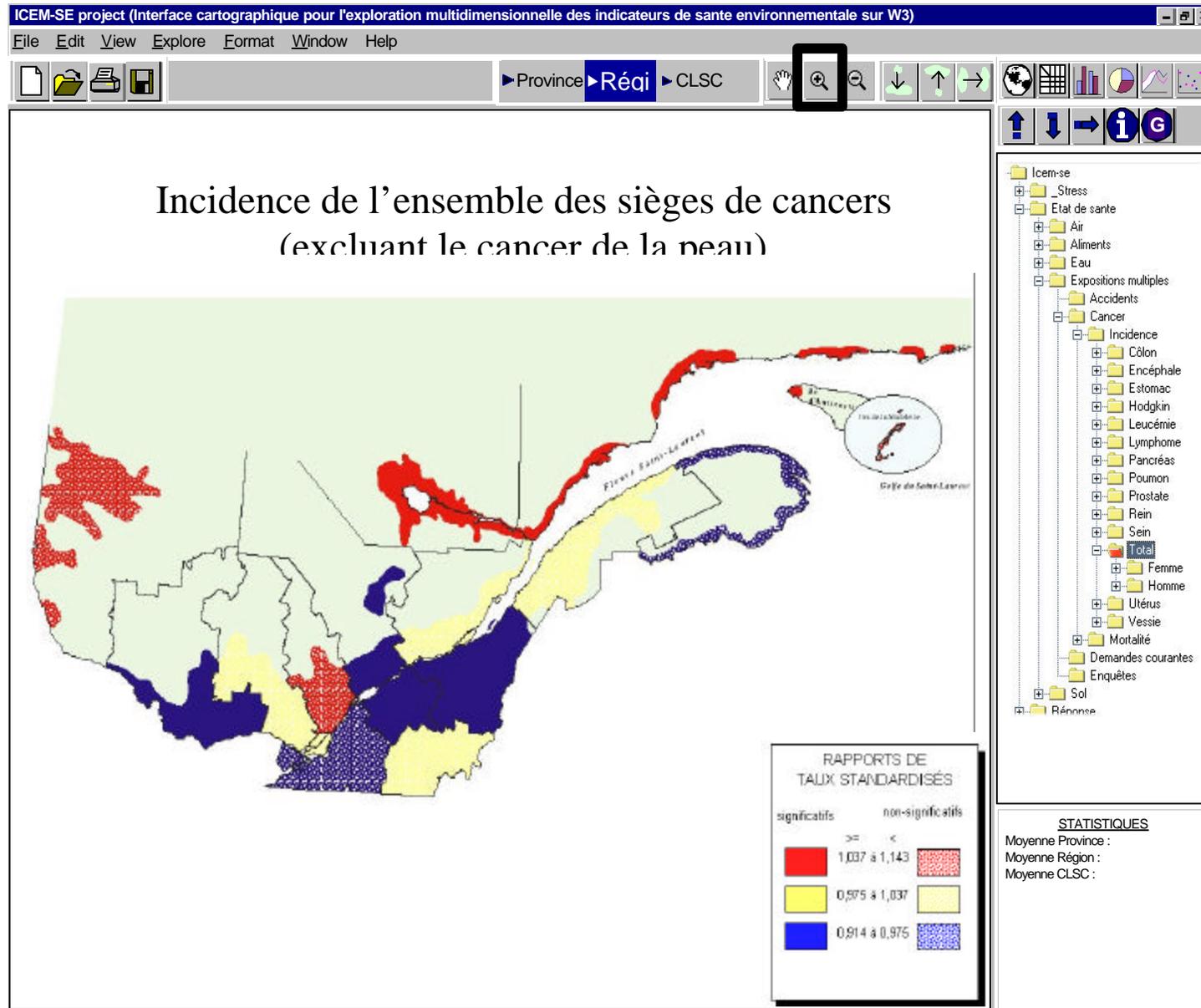


Figure 3 : Interface graphique d'ICEM/SE pour l'exploration et la navigation au sein des données

tion cartographique à une représentation sous forme d'histogramme), sélectionner un nouveau sous ensemble de données (ex: les données sur le cancer de la vessie), établir des comparaisons...
L'intérêt d'une telle interface est de proposer un outil facile d'utilisation permettant de naviguer dans les données sans que le cours de notre réflexion soit freiné par la formulation de complexes requêtes.

Conclusion

Comme nous avons pu le montrer dans le présent article, le gestionnaire des risques à la santé d'origine environnementale fait actuellement face à un ensemble de problèmes très vaste, pour lesquels l'information n'est pas facilement accessible. En effet, des milliers de produits chimiques et contaminants peuvent être présents à des concentrations variables dans l'air, l'eau, le sol, les aliments, en transit ou stockés dans la cours des usines; les sources peuvent être locales ou globales, tout comme les solutions, et les problèmes de santé très subtils à cerner.

Partant de ces constats, nous avons entrepris dans le cadre de ce projet de recherche appliquée, la réalisation sur Intranet un système pour l'analyse descriptive des données complexes en santé environnementale à l'aide d'une interface cartographique simple et rapide basée sur l'approche multidimensionnelle des Data Marts et l'interface des outils OLAP.

Une retombée importante d'un projet est de rendre disponible des informations existantes de SE et d'établir des mécanismes adéquats de traitement de ces données pour la prise de décision en gestion du risque réel ou perçu par la population. Un tel SIG accessible par le W3 peut servir à la diffusion de l'information auprès de la population, à rendre l'information environnementale et sanitaire pertinente plus accessible pour les entreprises (études d'impact, études de marché), en plus de contribuer à une planification utile des services de santé et de prévention.

L'intérêt d'un tel système d'information est donc primordial pour les intervenants en SE. Et pourtant, de tels systèmes ne sont pas, à notre connaissance, en fonction nulle part à l'échelle d'une juridiction administrative d'importance dans le monde, puisque les systèmes existants (Lepper, 1995; Risk, 1997) demeurent à visée de recherche.

Bibliographie

De Lepper M.J.C. et al.(eds.), (1995), *The Added Value of Geographical Information Systems in Public and Environmental Health*, Kluwer Academic Publishers, London (U.K.), on behalf of WHO Regional Office for Europe, 355 p.

Elliott P. et al. (eds.), (1992), *Geographical and Environmental Epidemiology: Methods for Small-Area Studies*, Oxford University Press, Oxford (U.K.), on behalf of WHO Regional Office for Europe, 382 p.

Eyles J., D.Cole and B.Gibson, (1996), *Human Health in Ecosystem Health, issues of meaning and measurement*, International Joint Commission, Ottawa and Washington, 73 p.

Fotheringham S. and Rogerson P.(eds.), (1994), *Spatial analysis and GIS*, Taylor and Francis, London (U.K.), 281 p.

Gosselin P., Bédard Y., Catelan R., (2000), Literature review of international Data Warehouse, GIS, WWW and OLAP health applications (Canda), Health Canada, 27p.

Gosselin P., Bédard Y., Elliott S., Jerrett Mike, R.Catelan, Poitras Ph., Gingras A., (2000) *GIS and OLAP in Health Surveillance : Needs Analysis for Successful Integration*, (Canda), Health Canada, , 53p.

Risk, (1997), International Conference: *Mapping Environmental Risks and Risk Comparison, Book of Papers*, Amsterdam, 22-24 october 1997, organized by RIVM, EPA, EU, OECD, TNO and VROM, 372 p.

World Health Organization (WHO), (1996), *Climate change and human health*, eds McMichael A.J. et al., Geneva, p.219-223.