

**Géomatique et santé environnementale:
innovations résultant du projet ICEM/SE**

Marie-Josée Proulx¹

Martin Nadeau¹, Sonia Rivest¹, Yvan Bédard¹
Pierre Gosselin² et Germain Lebel²

¹ Centre de recherche en géomatique
Université Laval
Sainte-Foy (Québec) Canada

² Centre de Santé Publique
2400, d'Estimauville
Beauport (Québec) Canada

[marie-josée.proulx, martin.nadeau, sonia.rivest, yvan.bedard]@scg.ulaval.ca
Pierre-L.Gosselin@crchul.ulaval.ca, germain.lebel@ssss.gouv.qc.ca

Biographies:

Marie-Josée Proulx détient un B.Sc. et une M.Sc. en Sciences Géomatiques à l'Université Laval. Elle travaille depuis 1995 dans l'équipe de bases de données spatiales et SIG du Centre de recherche en géomatique sous la direction du Dr Yvan Bédard. Elle possède une expérience à la fois en recherche fondamentale et en recherche appliquée, et a contribué à plusieurs conférences et articles scientifiques tant sur le plan national qu'international. Elle a également participé au montage de cours de formation continue au Canada, en Europe et en Afrique. Ses collègues du projet ICEM-SE proviennent de la même équipe au CRG ou de l'équipe du Centre de recherche du CHUQ-pavillon CHUL (Centre hospitalier de l'université Laval).

Résumé:

Le projet ICEM/SE (Interface Cartographique pour l'Exploration Multidimensionnelle des indicateurs de Santé Environnementale sur le World Wide Web) a mené au développement d'un outil d'analyse des données spatiales qui est innovateur. Pour notre partie de ce projet GEOIDE, nous visons à exploiter les données de façon rapide et simple en facilitant la gestion temporelle et multi-échelles des données par de nouvelles approches technologiques (i.e. outils SOLAP: Spatial On-Line Analytical Processing).

Cet outil a le potentiel de contribuer considérablement à l'amélioration de la surveillance en santé environnementale en donnant accès à une vingtaine d'indicateurs de santé et d'indicateurs environnementaux. Cet outil s'adresse à un usager non-initié à la géomatique par une approche totalement nouvelle. Il facilite l'accès aux données statistiques en permettant de créer facilement des cartes en quelques cliques de souris. La représentation cartographique a été automatisée pour que les cartes proposées soient élaborées en respectant les règles de la sémiologie graphique utilisées par les spécialistes en santé environnementale de l'équipe. L'interface permet d'explorer ces données à l'aide des outils de forage typiques du Geographic Knowledge Discovery. On peut donc visualiser la carte des régions sociosanitaires pour une maladie et forer ensuite en un ou deux cliques pour visualiser immédiatement les données détaillées des territoires de CLSC de cette région. On peut facilement naviguer dans les niveaux de détails et changer

les paramètres de la carte (ex. d'une carte de l'asthme, passer à la carte des maladies de l'appareil respiratoire). À tout moment, il est possible de basculer vers une autre représentation graphique (ex. histogramme, camembert, tableau) pour visualiser les données selon une autre perspective, comme pour suivre l'évolution temporelle d'une maladie via un histogramme. Par sa conception simple et conviviale, l'outil s'adresse à une clientèle d'utilisateurs moins techniques, il facilite la découverte de connaissances et améliore la prise de décision et les interventions en santé. De plus, par une "absence" de langage d'interrogation et des réponses instantanées, le prototype ICEM/SE vise à s'intégrer de façon transparente au flux de pensée de l'analyste humain. Ce projet, grâce à son aspect innovateur, a résulté en la première déclaration d'invention du réseau canadien d'excellence GEOIDE.

Abstract:

The ICEM/SE research project (Cartographic Interface for the Multidimensional Exploration of Environmental Health Indicators on the World Wide Web) led, after three years of research, to the development of a new type of spatial analysis tool that innovates in several aspects. Our contribution to this GEOIDE project aimed at exploiting the data in a very fast and simple way by facilitating the temporal and multi-scales management of the data with a new technological approach (SOLAP: Spatial On-Line Analytical Processing).

This tool has the potential to considerably improve environmental health surveillance by facilitating the access to twenty health indicators. The tool is designed for users having no geomatics knowledge, thanks to a totally new approach. It is based on an innovative type of tool that can be seen as a « keyboardless GIS » given that analysis can be performed entirely with mouse clicks. The cartographic representation of data was automated so that the proposed maps follow the graphic communication practices that are used by our team specialists in environmental health. The tool allows to investigate these data by means of drilling functions typical of Geographic Knowledge Discovery. One can thus create a map of the socio-sanitary regions for a particular disease and then in one or two clicks drill down in a region to see the disease in a more detailed manner on its community health centers territories. One can easily navigate in the levels of details to change the map parameters (e.g. from a map of the lung cancer to a map of the cancers of the respiratory system). At any time, it is possible to visualize the data using different display methods (e.g. histogram, pie chart, table) to view the data in another perspective (e.g. visualize temporal trends of a disease on an histogram). With its simple and friendly design, the tool is addressed to non-technical users. The tool facilitates the discovery of new knowledge and improves the decision-making processes and the interventions. Furthermore, with its "absence" of query language and with its immediate answers, the prototype aims at becoming integrated, in a transparent way, into the stream of thought of a human analyst. This project, thanks to its innovative aspects, led to the first invention declaration of the Canadian network of excellence GEOIDE.

Introduction

Les logiciels SIG actuels et les systèmes de gestion de bases de données (SGBD) n'étant pas efficaces pour aider la prise de décision au niveau tactique et stratégique,

il fallait trouver une solution pour aider les professionnels en surveillance en santé à dresser facilement et rapidement le bilan de santé d'une population donnée.

Les SIG et SGBD sont en fait peu conviviaux pour interroger les données spatiales et les métadonnées et d'importants efforts doivent être faits pour compenser cette faiblesse (ex. formulaires et cartes préprogrammées, tableaux de bord, utilisation de requêteurs et d'outils OLAP). Sans ces efforts, l'utilisateur doit maîtriser un langage d'interrogation en plus de la structure des données qui est de plus en plus complexe avec la gestion des données spatio-temporelles. Cette aptitude n'est pas courante chez les professionnels de métiers autres que la géomatique ou l'informatique. Par conséquent, un expert du domaine, ici de santé environnementale, doit maîtriser l'informatique ou dépendre directement d'un expert technique pour l'exploitation des outils actuels.

D'abord, les logiciels SIG sont peu efficaces pour explorer et naviguer interactivement dans les données spatiales. En pratique, lorsqu'on utilise un SIG, la finalité est de produire une requête précise donnant le résultat espéré. Ce processus demande parfois de réaliser plusieurs cartes pour parvenir à nos fins et nécessite de multiples opérations qui nuisent au flux de pensée de l'utilisateur. En effet, celui-ci est concentré sur les processus à effectuer pour obtenir son résultat d'analyse plutôt que sur l'analyse elle-même et sur l'étendue des connaissances disponibles, i.e. les données. De plus, le contenu des bases de données supportées par les logiciels SIG est souvent de nature très détaillée. Par exemple un SIG peut répertorier les cas de cancer du poumon par adresse civique. La production d'information de synthèse demande un traitement complexe des données qui n'est pas facilement réalisable par tout type d'utilisateur. Par exemple, un information de synthèse serait la distribution des cancers de l'appareil respiratoire (i.e. regroupe des plusieurs cancer dont celui du poumon) par territoire de CLSC (i.e. regroupement d'adresses civiles).

Finalement, une requête spatio-temporelle est complexe à formuler et demande un temps de traitement très long dans les logiciels SIG, car la structure habituelle sous-jacente à ce type d'outil n'est pas adéquate pour gérer facilement les combinaisons de données spatiales et les données spatio-temporelles. Effectuer des comparaisons dans le temps et dans l'espace avec les outils actuels demeure de la haute voltige. Toutes ces limites sont typiques des logiciels développés à des fins dites "transactionnelles" (comme sont les SIG et SGBD), mais deviennent problématiques pour les applications décisionnelles.

OLAP versus SOLAP

Depuis 1996, notre équipe du a expérimenté une nouvelle technologie réservée jusqu'alors au domaine des technologies de l'information et très peu exploitée en géomatique. Il s'agit de la technologie OLAP (On Line Analytical Processing) dont le potentiel pour supporter l'analyse spatio-temporelle a été démontré pour la première fois à ce même congrès (Bédard 1997) et a clairement été démontré par la suite (Caron 1998). Cependant, sans un volet spatial adéquat permettant d'exploiter la composante géométrique des données spatiales, l'analyse demeure partielle. Notre objectif était donc de développer une nouvelle catégorie d'outils, les outils SOLAP (Spatial OLAP), qui tiendrait compte cette fois-ci de l'univers spatial dans lequel se situent nos analyses. Le projet ICEM-SE fut un des principaux projets grâce auxquels nous avons pu réaliser notre objectif.

Au départ, l'analyse OLAP a été définie pour la prise de décision puisqu'elle est conçue pour être d'utilisation facile et rapide. D'abord, l'aisance vient de la capacité de conduire l'analyse sans devoir surmonter un langage d'interrogation ou comprendre la structure sous-jacente de la base de données (Marchand et *al.* 2002), qui peut être très complexe dans le cas particulier de bases de données spatio-temporelles (Bédard et van Chestein, 1995).

Dans l'outil OLAP, les données sont importées des systèmes transactionnels et restructurées dans un cube, à travers des dimensions. Par exemple, les maladies, les territoires de CLSC et le temps sont les dimensions du cube Incidence. Plusieurs cubes peuvent être produits à partir des mêmes données transactionnelles détaillées stockées dans les SIG et SGBD. Les valeurs sont calculées pour chaque combinaisons de dimensions à l'intérieur du cube, et ceci tant pour les niveaux agrégés que les niveaux détaillés. Par exemple, les taux d'incidence de maladies sont calculées pour chaque territoire de CLSC, chaque maladie pour une année (figure 1 à gauche).

L'analyse OLAP permet aussi à l'utilisateur d'explorer et de naviguer très facilement à travers les différents niveaux de détail des dimensions à l'aide d'outils de forage (ex. figure 1, Visualiser l'incidence de chaque maladie pour chaque territoire de CLSC pour un temps précis pour ensuite forer dans les maladies afin de visualiser l'incidence d'une seule maladie (ex. Asthme) pour chaque territoire de CLSC à ce même temps précis. En fait, l'analyste agit directement avec les données et se concentre sur les résultats de l'analyse plutôt que sur la procédure exigée par l'outil pour exécuter le processus.

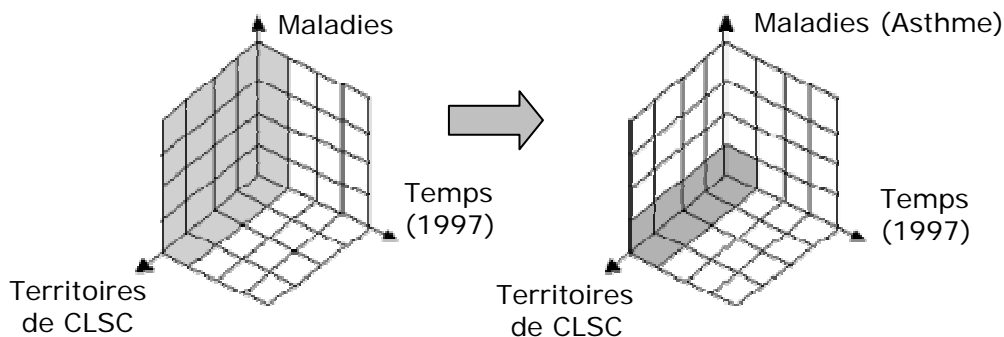


Figure 1: Visualiser l'incidence de chaque maladie pour chaque territoire de CLSC pour un temps précis pour ensuite forer pour visualiser l'incidence d'une seule maladie (ex. Asthme) à un temps précis (ex. 1997).

Deuxièmement, l'approche OLAP est rapide parce que les données sont précalculées par niveau de détails (ex. les données sont calculées par territoires de CLSC (niveau plus détaillé), mais sont aussi agrégées par régions sociosanitaires et par province (niveaux plus généraux)). Le temps de traitement de l'analyse est alors réduit (OLAP Council 1995; Pendse 2000) et répondre très vite aux questions complexes est chose possible. Cela permet à l'utilisateur de maintenir sa pensée, son attention, sans être distrait par des temps de réponse lents. Cela permet l'accès à toutes les vues possibles ou les combinaisons des données, facilitant l'apparition de nouvelles hypothèses et encourageant la découverte de connaissance. (Rivest, 2001).

Cependant, dans un outil OLAP, la dimension spatiale est traitée comme une autre dimension descriptive (ex. Territoires de CLSC, figure 1). Sans prendre en considération la composante cartographique des données, l'outil OLAP présente des limitations sérieuses dans l'appui à l'analyse spatio-temporelle (aucune visualisation spatiale (pour apprécier la proximité et la localisation), pratiquement aucune analyse spatiale, aucune exploration basée sur des cartes, etc.).

Afin d'exploiter les capacités d'analyse spatio-temporelle d'un outil OLAP conventionnel, il est essentiel d'ajouter une composante spatiale offrant au minimum la visualisation spatiale des éléments géométriques de l'ensemble de données.

Les systèmes OLAP commerciaux offrant une composante spatiale présentent beaucoup de limitations (Business Objects 1998; Cognos 2000; Knosys 2000). Actuellement, ces outils offrent une certaine composante de visualisation des données qui permet d'apprécier la proximité des régions et leur localisation, mais il est souvent impossible de manipuler et de naviguer dans les données cartographiques spatio-temporelles comme il est facilement et rapidement possible de le faire avec les données tabulaires et les diagrammes. Les outils de forage de données spatiales sont souvent inexistantes. Ce concept est pourtant fondamental. Le développement de solutions pour faciliter la visualisation et la navigation dans les données spatio-temporelles fait partie de notre programme de recherche sur la conception et l'analyse de bases de données spatiales et de notre participation dans trois projets GEOIDE.

Le projet de recherche ICEM/SE

Le projet de recherche ICEM/SE, avait un double objectif. Il fallait répondre à des besoins du domaine de la santé environnementale en essayant de relever des défis de nature géomatique.

Au départ, le projet devait permettre d'améliorer la prise de décision et les interventions dans le domaine de la santé environnementale par la découverte de nouvelles connaissances en facilitant l'élaboration et la validation de nouvelles hypothèses de recherche. Pour ce faire, il fallait améliorer les opérations de surveillance en santé environnementale par l'accès à des données de haute qualité. Un grand travail de cueillette de données et d'intégration a alors été fait afin de définir près de 20 indicateurs de santé environnementale permettant de produire des analyses (cf. tableau 1).

Ces indicateurs avaient la particularité d'être très hétérogènes en termes de résolution spatiale et de couverture temporelle. Par exemple, les données de l'indicateur de la qualité de l'air sont disponibles à partir de 1975 tandis que les données des substances rejetées le sont à partir de 1989. Ces deux indicateurs sont différents d'autant plus qu'ils possèdent une géométrie ponctuelle définie différemment (i.e. stations d'échantillonnage versus industries polluantes). La même chose se présente dans les indicateurs de santé où les taux de mortalité diffèrent en terme de couverture temporelle des taux d'incidence. Cette caractéristique propre à notre jeu de données spatio-temporelles a mené au choix d'une architecture multidimensionnelle qui permet de stocker les valeurs calculées seulement pour les combinaisons possibles dans notre jeu de données.

Indicateurs environnementaux	Résolution spatiale i.e. plus petite région couverte	Couverture temporelle par années
1- Substances altérant la couche d'ozone (par concentration)	Pays	1979- ...
2- Substances altérant la couche d'ozone (par type de substances)	Province	1993- ...
3- Gaz à effet de serre	Province	1990-1994
4- Ventes de crème solaire	Province	1994-1995
5- Ventes d'engrais	Pays	1992-1995
6- Exposition au soleil	Province	1996
7- Contamination	Terrain contaminé	1993-...
8- Collecte sélective	Municipalité	1995-1997
9- Ventes de pesticides	Subdivision de recensement agricole	1981, 1886, 1991, 1996
10- Perception de la qualité des eaux	Région sociosanitaire	1996
11- Qualité de l'air	Station d'échantillonnage	1975-...
12- Substances rejetées (INRP)	Industrie	1993-...
13- Fertilisants	Bassin versant	1995
14- Qualité de l'eau	Station d'approvisionnement	1989-1993
Indicateurs de santé		
15- Taux de mortalité des maladies de l'appareil respiratoire	Territoire de CLSC	1984-...
16- Taux d'hospitalisation des maladies de l'appareil respiratoire et de l'asthme	Territoire de CLSC	1991-1996
17- Taux de visite à l'urgence pour l'asthme	Territoire de CLSC	1992, 1994, 1996
18- Taux d'incidence des cancers	Territoire de CLSC	1989-1995
19- Taux de mortalité des cancers	Territoire de CLSC	1984-1996
20- Taux de prévalence des rhinites allergiques	Région sociosanitaire	1998
21- Cas d'intoxications et appels d'intoxications	Région sociosanitaire	1987-...

Tableau 1. Résolution spatiale et couverture temporelle des indicateurs environnementaux et de santé du projet ICEM/SE.

Pour répondre adéquatement à notre objectif, un système d'analyse devait être mis en place de façon à doter les intervenants professionnels du réseau de santé publique d'un moyen efficace pour décrire et analyser les données complexes en santé environnementale.

Pour y parvenir, il était nécessaire d'exploiter directement la technologie SOLAP dans ce système. Ainsi, l'équipe de géomatique devait démontrer qu'il était possible, en conservant une interface cartographique de navigation simple, d'accroître la rapidité des résultats d'analyse spatiale (figure 2).

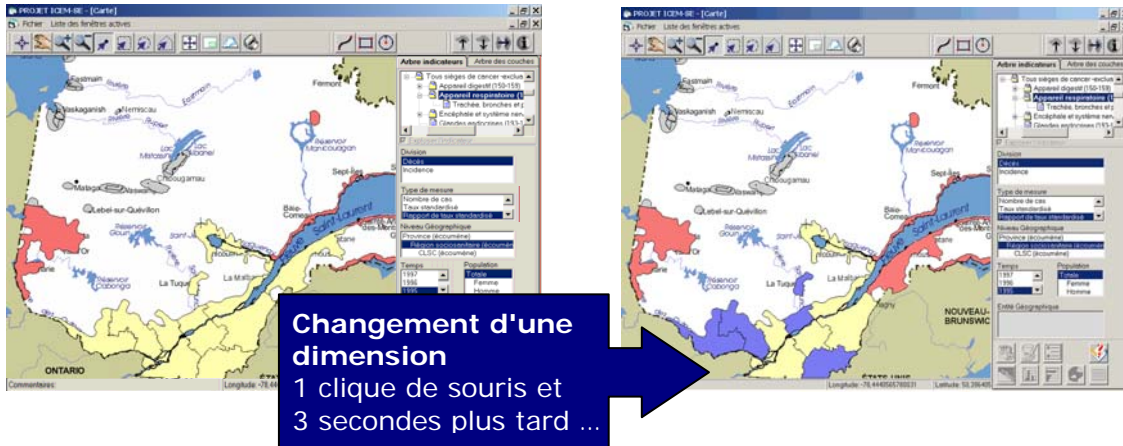
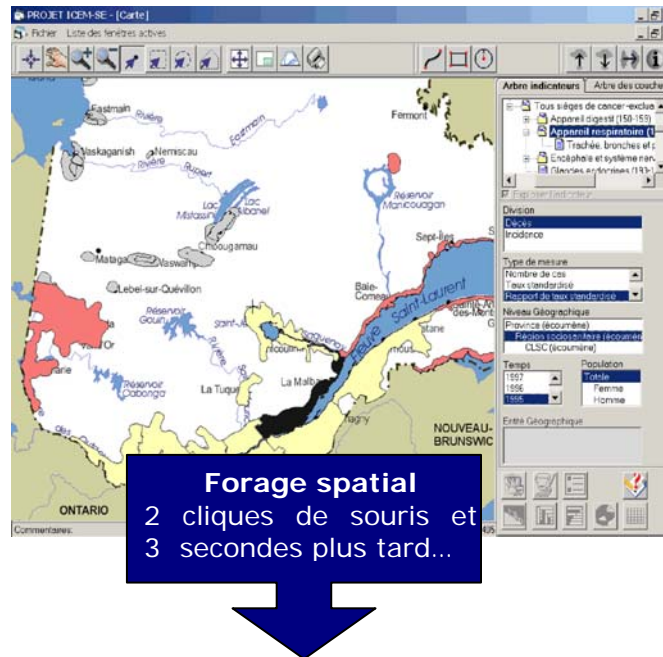


Figure 2. (à gauche)- Création d'une carte de l'incidence du cancer du foie chez les femmes en 1997 pour toutes les régions. (à droite)- Changement à l'intérieur d'une dimension pour obtenir la carte de l'incidence du cancer du foie chez les hommes en 1997 pour toutes les régions.

L'équipe devait aussi démontrer qu'il serait possible de faciliter l'analyse temporelle et multi-échelles des données spatiales. Par conséquent, l'outil devait inclure des fonctionnalités de navigation spatiale (i.e. forage spatial) sur les données cartographiques (figure 3).



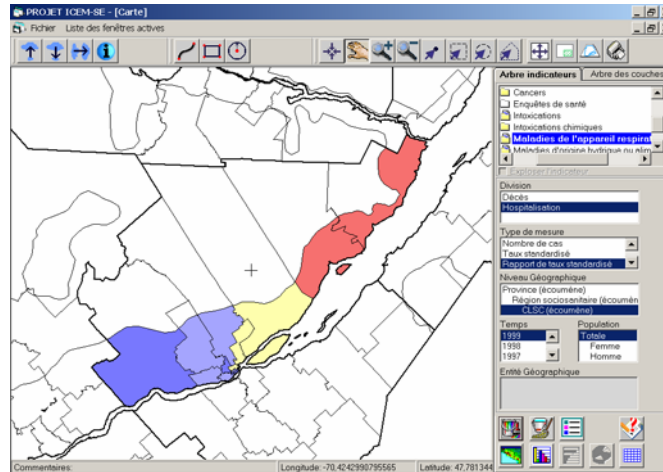


Figure 3: Région sociosanitaire sélectionnée (en haut) pour faire un forage sur une région particulière et obtenir le taux d'incidence d'une maladie recalculé par territoires de CLSC (en bas).

Par conséquent, l'outil SOLAP développé au cours de ce projet permet de répondre aux attentes tout en facilitant l'analyse par des fonctionnalités propres. D'abord, par son interface cartographique simple d'utilisation, l'outil permet à l'analyste de créer facilement des cartes répondant aux règles de l'art de la communication graphique. Des composantes de sémiologie graphique ont été intégrées à l'outil afin de permettre à des néophytes en géomatique de créer des cartes cohérentes avec la bonne sémiologie graphique et une légende adéquate (non montrée). Cette fonctionnalité rend l'utilisateur autonome vis-à-vis les experts en cartographie du domaine.

L'outil permet aussi à l'utilisateur d'afficher ses données sous forme de cartes, tableaux, histogrammes et camemberts en plus de permettre la comparaison entre eux. Par conséquent, certains résultats deviennent plus facilement assimilables pour l'utilisateur. Par exemple, l'appréciation de tendances temporelles existant dans les données est plus claire sous forme d'histogramme que de tableau.

Finalement, l'outil synchronise les résultats de la navigation par forage dans les différentes dimensions et types d'affichages. Ainsi, les mêmes niveaux de détails des données apparaissent à l'utilisateur sous différentes formes (cf. figure 4).

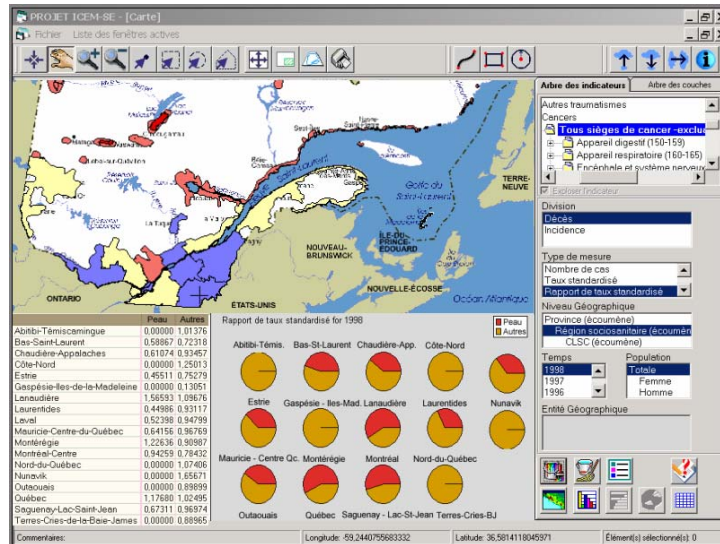


Figure 4. L'outil permet la comparaison entre les différents types d'affichage et la navigation par forage dans les dimensions est synchronisée.

Conclusion

Ce projet a fourni une solution unique aux professionnels de la santé environnementale tout en apportant plusieurs contributions dans le monde de la géomatique tant du point de vue scientifique que pratique.

Le présent projet a permis de consacrer d'importantes ressources pour concevoir et développer une application dans un domaine non-traditionnel, le tout faisant appel à une nouvelle catégorie d'outils permettant d'explorer rapidement et facilement les données géospatiales à différents niveaux de granularité. La facilité d'utilisation et la flexibilité du système SOLAP en sont deux caractéristiques importantes qui, lorsque combinées avec la rapidité des réponses, surpassent de façon importante les capacités jusqu'ici offertes par les systèmes d'information géographique. L'utilisation d'une interface de type hyperlien, i.e. par simple sélection avec la souris de l'information désirée et des opérations de navigation, constitue une innovation importante qui nous permet de qualifier le résultat de "keyboardless GIS" appliqué à la santé environnementale.

Ce projet innove également par la combinaison des structures de bases de données multidimensionnelles avec les fonctions SIG, ces dernières étant typiquement transactionnelles. Ce n'est qu'après plusieurs expérimentations technologiques effectuées dans plusieurs autres projets que nous avons réussi à trouver les meilleures solutions et ainsi à innover en terme de combinaison de solutions. Nous avons de plus défini les caractéristiques idéales d'un SOLAP (Rivest et al 2001) et travaillons présentement à les développer. Il en résulte des possibilités d'exploration des données géospatiales jamais offertes sur le marché et qui supportent la découverte de nouvelles connaissances géographiques en santé environnementale.

Depuis, le prototype ICEMSE a donné lieu à la première déclaration d'invention du réseau canadien d'excellence GEOIDE. Le transfert technologique vers l'industrie se

fait présentement avec la firme Les Conseillers KHEOPS Technologies pour la mise en marché d'un produit commercial. Enfin, d'autres projets ont également contribué au développement des concepts théoriques et aux expérimentations SOLAP depuis 1996, et nous prévoyons continuer nos travaux sur ce sujet pour les cinq prochaines années (ex. enrichissement du SOLAP 2D, SOLAP 3D, SOLAP temps réel).

Remerciements

L'équipe tient à remercier le réseau canadien de centres d'excellence GEOIDE et les partenaires du projet ICEM-SE pour leur support financier durant les différentes parties du projet, soit le Ministère de la Santé et des Services Sociaux du Québec, SOGIQUE, l'Institut national de santé publique du Québec, le Plan géomatique du gouvernement du Québec, les Conseillers Khéops Technologies.

Bibliographie

- Bédard, Y. and Y. van Chestein (1995) La gestion du temps avec les systèmes de gestion de données localisées : État actuel et avenues futures. Actes du Colloque Géomatique V, La route de l'innovation, ACSG, Montréal*
- Bédard, Y. (1997). Spatial OLAP, Vidéoconférence, 2ème Forum annuel sur la R-D, Colloque Géomatique VI: Un monde accessible, 13-14 novembre, ACSG, Montréal.*
- Business Objects (1998) Business Objects and ESRI Announce Strategic Partnership, Business Objects Press Releases. <http://www.businessobjects.com/news/>.*
- Caron, P.Y. (1998) Étude du potentiel OLAP pour supporter l'analyse spatio-temporelle. Mémoire de MSc, Dép. Sciences géomatiques et Centre de recherche en géomatique, Université Laval.*
- Cognos (2000) Cognos Visualizer Product Information <http://www.cognos.com/products/visualizer>.*
- Gosselin, P., Bédard, Y (2002) Final Report, GEOIDE Soc#1 Project, ICEM/SE: Interface Cartographique pour l'Exploration Multidimensionnelle des indicateurs de Santé Environnementale sur le World Wide Web, Université Laval, mars 2002, 20 p.*
- Knosys (2000) Geo Spatial Mapping: MapInfo MapX Plug-In for Knosys ProClarity 3.0. <http://www.knosysinc.com/productextent1.html>.*
- Marchand, P., Y. Bédard and G. Edwards. (2002). A hypercube-based method for spatio-temporal exploration and analysis. GeoInformatica, Submitted May 2002.*
- OLAP Council (1995) OLAP and OLAP Server Definitions. <http://www.olapcouncil.org/research/glossaryfy.htm>.*
- Pendse, N. (2000) What is OLAP? The OLAP report. <http://www.olapreport.com/fasmi.htm>.*
- Rivest, S., Bédard, Y. and Marchand P. (2001), Towards better support for spatial decision-making: Defining the characteristics of Spatial On-Line Analytical Processing (SOLAP), Geomatica, the journal of the Canadian Institute of Geomatics, Vol. 55, No. 4, 2001, pp. 539 to 555*