

**Technologies de pointe de l'information appliquées à la surveillance de la santé
environnementale :
Aperçu et analyse**

Yvan Bédard, PhD ¹
William D. Henriques, PhD, MSPH ²

¹ Centre de recherche en géomatique, Université Laval, Ville de Québec, Canada G1K7P4,
yvan.bedard@scg.ulaval.ca

² Coordonnateur SIG, Agency for Toxic Substances and Disease Registry, Division of Health Assessment and
Consultation, 1600 Clifton Road NE, MS E56, Atlanta GA 30333, whenriques@cdc.gov

Introduction

Au cours des dernières années, nous avons assisté à l'introduction massive de nouvelles technologies de l'information qui ont complètement changé le visage de notre société. Bon nombre d'auteurs parlent en fait d'une révolution des communications, comme celles survenues à la suite de l'invention de la presse à imprimer, de la radio et de la télévision. Avec cette révolution technologique sont apparues de nouvelles expressions comme «village planétaire», «société de l'information» et «planète virtuelle» pour décrire les nouvelles réalités de la connectivité et la révolution financière du commerce électronique. Bien que leur implantation soit davantage concentrée dans les pays développés, ces nouvelles technologies ont aussi fait apparition dans des groupes isolés de nations en voie de développement. Elles dépendent de la convergence de plusieurs technologies : ordinateurs puissants et abordables, appareils électroniques de mesure et de surveillance en temps réel, importante production d'information numérique en différents formats, moyens de communication sans fil plus rapides. Maintenant, ces technologies viennent toucher à tous les domaines d'application, y compris la surveillance de la santé environnementale. Le présent article donne un aperçu des technologies qui ont ou qui auront le plus répercussions sur la santé environnementale. La première section fait un survol des technologies de l'information dont l'influence est incontestable, notamment le World Wide

Bédard, Y. & W. Henriques, 2003, *Technologies de pointe de l'information pour la surveillance de la santé environnementale, Environnement et santé publique, Edisem, Vol. Chapitre 36, 27p.*

Web, les systèmes de gestion de base de données (SGBD) et les serveurs universels, les entrepôts de données et les outils d'aide à la décision, et finalement les systèmes d'information géographique et les technologies associées. La deuxième section définit la surveillance de la santé environnementale et les besoins d'information dans ce domaine. La troisième section donne des exemples pratiques de l'utilisation des technologies de l'information en surveillance de la santé environnementale et décrit les avantages et les ressources nécessaires. La quatrième section porte sur les défis que représente une mise en oeuvre réussie de ces technologies et propose certaines suggestions quant aux orientations futures. Une conclusion vient terminer l'article.

Aperçu des technologies de pointe de l'information dans la surveillance de la santé environnementale

Le World Wide Web

La pénétration massive d'Internet et du **World Wide Web** (W3) dans la société d'aujourd'hui a ouvert la voie à de nouvelles façons de fournir et d'accéder à de l'information et à des services. Jamais auparavant n'avait-il été possible d'accéder à d'importantes quantités d'information et de services partout dans le monde, 24 heures sur 24. De plus, sur le plan technologique, les technologies indépendantes ont suivi le courant des technologies principales et y adhèrent aujourd'hui beaucoup plus facilement. Voici les plus importantes technologies de l'information offertes actuellement :

- a) Le **courriel** (ou courrier électronique) est l'application la plus utilisée dans Internet. On s'en sert pour envoyer/recevoir des messages de n'importe où, et transmettre des

documents électroniques (p. ex. rapports écrits, photos d'accident, cartes géographiques, vidéos, fichiers de musique, présentations).

- b) Les **sites Web** sont un endroit où une organisation ou un individu peut offrir aux internautes de l'information statique et limitée, ou bien de l'information dynamique et massive. Sur les sites Web, les meilleures interfaces-utilisateurs sont celles conçues de sorte à faciliter la navigation du site avec des hyperliens (liens sur lesquels on clique pour aller vers d'autres pages du site, ou vers d'autres sites Web) ou des formulaires à remplir. Lorsqu'une organisation décide de restreindre une partie de son site Web aux employés à l'interne, on parle alors d'un **intranet**, et lorsqu'on donne accès à certains clients externes ou à des partenaires, on parle d'un **extranet**.
- c) Les sites Web qui offrent un vaste éventail d'information organisée et indexée, des moteurs de recherche, des services personnalisables (p. ex. bulletins météo où l'utilisateur sélectionne sa région, nouvelles du sport, rendements boursiers, événements thématiques, rapport de circulation locale, etc.), ainsi que des forums de discussion («chat rooms») sont désignés sous le nom de **portails** (p. ex. Yahoo). Lorsqu'un portail se consacre à un domaine en particulier (p. ex. la santé environnementale), on parle de **portail vertical**. Le portail s'adresse davantage au grand public, alors que le portail vertical est destiné aux spécialistes d'un domaine en particulier.
- d) Certains sites Web offrent des fonctionnalités de commerce électronique (ou **cybercommerce**) qui permettent à une organisation tel un magasin de vente au détail de vendre ses produits directement sur le Web. Ces sites comprennent un catalogue de produits, une liste de prix, les conditions d'achat et les procédures de facturation.

Certains sites sont des **bibliothèques électroniques** qui fournissent à un groupe d'utilisateurs bien défini un catalogue de documents électroniques d'intérêt (p. ex. rapports, fichiers de données, cartes et images satellites) à rechercher par mots-clés ou synonymes et à télécharger par Internet. Compte tenu de la nature confidentielle des renseignements qu'on y soumet dans le cadre des transactions (numéros de compte ou de carte de crédit du client), les sites de commerce électronique sont habituellement dotés de systèmes de protection de la confidentialité.

- e) Le Web offre également des technologies pour **l'apprentissage en ligne et les groupes de travail**. Elles sont basées sur des capacités de communication multi-point statiques ou interactives : communications écrites, groupes de discussion, tableau blanc virtuel, gestion d'agendas partagés, manipulation et réplique de logiciels et enregistrements audios et vidéos en temps réel point-à-point. L'organisation hôte doit disposer d'un logiciel spécialisé, mais pas l'utilisateur final (sauf dans le cas d'une webcam en ligne facultative).

Au milieu de l'année 2000, on estimait à 1,5 milliards le nombre de pages d'information résidant sur le Web, sans compter les centaines de milliers de bases de données spécialisées auxquelles on peut accéder. Vu cette abondance de sites et de bases de données, trouver l'information la plus appropriée et la plus à jour demeure un défi de taille. C'est pourquoi il existe, en plus des portails et des portails verticaux, des **sites de recherche** (quelques sites généraux et des milliers de sites spécialisés) destinés à faciliter la recherche sur le Web. Ces sites utilisent diverses approches d'exclusion : recherche avec au moins un des mots proposés, recherche de tous les mots proposés, recherche d'une expression telle quelle, recherche avec variations grammaticales (p. ex. cheval et chevaux), ajout d'opérateurs booléens (ET, OU, etc.), exclusion de certains mots,

Bédard, Y. & W. Henriques, 2003, Technologies de pointe de l'information pour la surveillance de la santé environnementale, Environnement et santé publique, Edisem, Vol. Chapitre 36, 27p.

interrogation en langage naturel (p. ex. langage clair), filtrage (p. ex. recherche dans une langue précise, dans un domaine d'application, ou seulement dans les sites récemment mis à jour) et recherche dans les résultats. Les **sites de métarecherche** sont des sites qui permettent d'interroger simultanément plusieurs sites de recherche et qui présentent les données de façon ordonnée. Malgré toutes ces options, il n'est pas rare que l'on obtienne des milliers de résultats pour une recherche, dont des résultats inutiles ou présentés de façon inutile pour l'utilisateur.

Afin de pouvoir bénéficier des technologies d'Internet et du Web, l'utilisateur doit obtenir un accès auprès d'un fournisseur de services Internet et disposer d'une connexion Internet (modem téléphonique, modem câble, etc.) et d'une adresse électronique. De plus, pour avoir accès à certains services, l'utilisateur pourrait devoir prendre un abonnement (gratuit ou non). Par contre, pour offrir de tels services, il faut se munir d'un serveur Web et d'un logiciel de protection spécialisé, appelé coupe-feu («firewall»).

SGBD et serveurs universels :

Parmi les principales technologies de pointe de l'information, l'une des premières a été le système de gestion de base de données, ou SGBD, dont des exemples sont Oracle DBMS, SQL-Server, Informix, Sybase, DB2, Access, etc. Mise au point il y a 30 ans, cette famille de solutions a maintenant atteint son apogée commerciale, spécialement avec l'approche relationnelle qui est en tête du marché depuis les 20 dernières années. Les **SGBD relationnels** permettent de définir la structure de la base de données, d'y introduire de l'information simple (chaînes de caractères, nombres, dates ou valeurs booléennes) et d'en vérifier l'intégrité, de manipuler les données et de les soumettre à des recherches, puis de générer des rapports automatiques (Date 2000). Ces systèmes peuvent être utilisés simultanément par plusieurs

utilisateurs, même par milliers, sans risque de panne ni de corruption des registres. Les données peuvent être stockées dans un seul endroit, ou bien être réparties sur plusieurs sites de façon complètement transparente pour l'utilisateur. L'accès aux données peut être direct, soit à l'aide d'interfaces-utilisateurs graphiques construites avec l'application, ou par le Web.

Influencés par le Web riche en média et les SGBD orientés objet à capacité multimédia qui sont apparus dans les années 1990, les SGBD relationnels ont évolué vers des SGBD hybrides relationnels et orientés objets, que l'on nomme **serveurs universels** (p. ex. Oracle 8i et Cartridges, Informix avec Datablades). Comme leur nom l'indique, ils sont universels car ils ne sont pas restreints aux types de données traditionnelles que l'on trouve dans les SGBD. On peut également y stocker, manipuler et rechercher l'information multimédia (p. ex. photos, cartes, sons, rapports complets, vidéos), qui peut servir entre autre au développement de sites Web. Par conséquent, on peut dire que les SGBD d'aujourd'hui sont bien adaptés à la révolution du Web.

Entrepôts de données et récents outils d'aide à la décision :

Bien que les SGBD aient été créés dans le but d'harmoniser les fichiers de données disparates, indépendants, redondants et spécifiques aux applications (figure 1), la plupart des organisations ont favorisé la création de bases de données isolées. Ainsi, en évoluant, les bases de données indépendantes avec redondances sont devenues des bases de données indépendantes avec recoupements. Généralement, on considère cette évolution comme une amélioration car les problèmes de recoupement et d'incohérence sont maintenant beaucoup plus faciles à gérer, particulièrement dans les décisions quotidiennes de niveau opérationnel. Néanmoins, le résultat n'a toujours pas l'aspect unifié d'un système intégré, où les données provenant de diverses bases de données sont dans l'ensemble cohérentes et prêtes à fournir rapidement de l'information

Bédard, Y. & W. Henriques, 2003, *Technologies de pointe de l'information pour la surveillance de la santé environnementale, Environnement et santé publique, Edisem, Vol. Chapitre 36, 27p.*

stratégique résumée et groupée pour une prise de décision à de hauts niveaux (Inmon et al 1996). C'est d'ailleurs ce que visent les **entrepôts de données** : «fournir une vision unifiée des diverses bases de données hétérogènes afin d'alimenter efficacement les outils d'aide à la décision utilisés pour la prise de décisions stratégiques» (Bédard, Merrett et Han 2000). Pour ce faire, l'entrepôt de données doit importer par lots en mode de lecture seule des sous-ensembles de bases de données sources (appelés systèmes sources) et les traiter/intégrer pour que l'information résultante à stocker soit conforme et correctement groupée (figure 2) (Poe 1995). Comme alternative, et dans la mesure où l'interopérabilité totale est instaurée, l'entrepôt peut faire le traitement ou la fusion des données à la volée (certaines solutions commerciales offrent des entrepôts virtuels qui se créent à la volée, mais se limitent à des petits entrepôts et à des données sources faciles à intégrer).

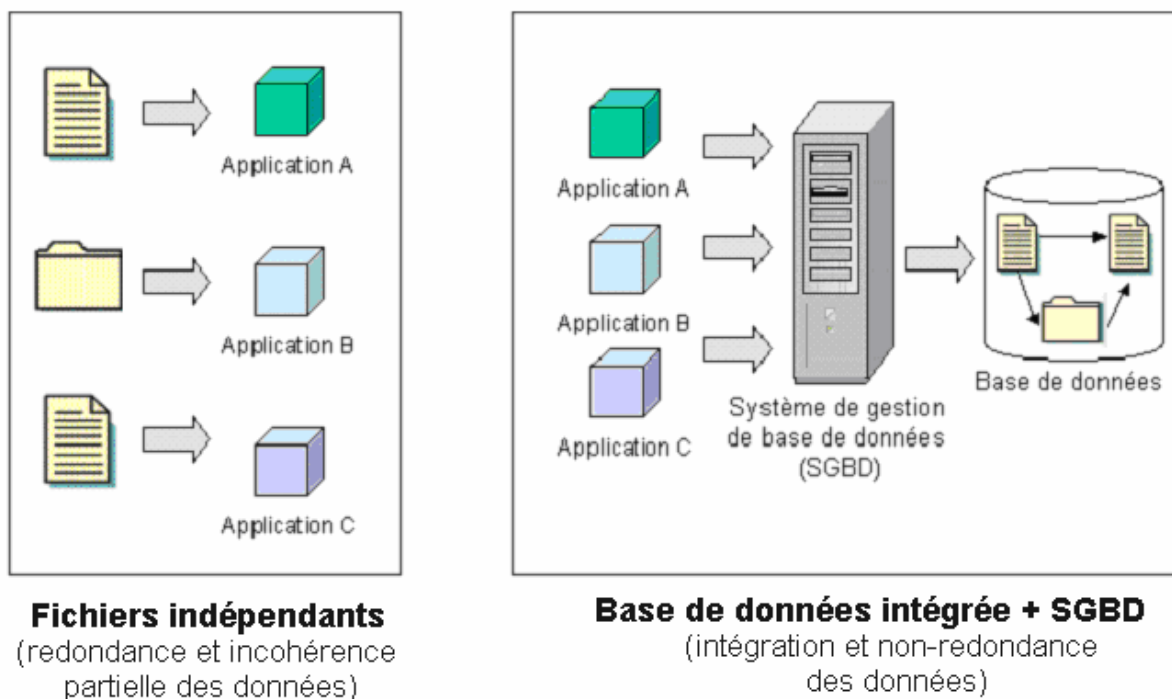


Figure 1 : Fichiers indépendants par rapport à base de données intégrée + SGDB

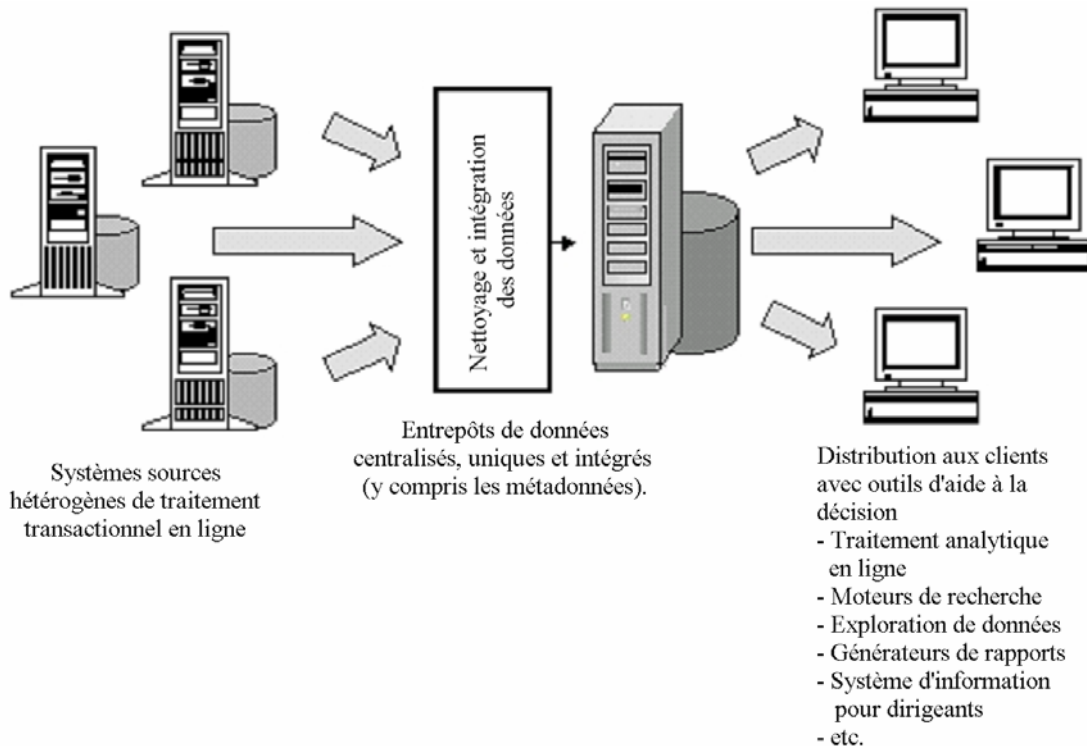


Figure 2 : Systèmes sources + entrepôt de données + outils d'aide à la décision

Lorsque les données sont mises à jour dans les systèmes sources (et que les données périmées sont archivées), la nouvelle information est ajoutée dans l'entrepôt sans remplacer l'information précédente. C'est cette caractéristique qui donne aux entrepôts de données la possibilité d'analyser les données au fil du temps, un élément clé dans la prise de décision (Brackett 1996). Par conséquent, les entrepôts de données sont habituellement considérés comme la principale source d'information en matière de découverte de connaissances et de veilles économiques.

Une importante différence entre les entrepôts de données et les bases de données traditionnelles est le fait que la base de données est conçue pour traiter un grand volume de courtes transactions détaillées de niveau opérationnel sur de petits groupes de données, alors que l'entrepôt de données est conçu pour traiter de petits volumes de longues transactions agrégées de niveau stratégique sur de grandes quantités de données. Lorsque le volume de données est trop élevé

pour atteindre ces objectifs contraires avec une base de données unique, on utilise deux différents modèles (ou technologies) de bases de données : le modèle relationnel-objet du SGBD traditionnel pour les opérations de type transaction, ou le modèle multidimensionnel des entrepôts de données pour les opérations de type analyse et la découverte de connaissances (c.-à-d. l'aide à la décision). Dans les deux cas, l'interrogation et la navigation de la base de données doivent être faciles pour l'utilisateur, et les temps de réponse doivent être très courts. Des exemples de systèmes d'entrepôts de données sont Red Brick, Essbase et Oracle Express.

Dans le cas où on voudrait simplement créer un entrepôt de données pour une partie de l'organisation, il est possible, avec la même technologie, de bâtir un **mini-entrepôt** spécialisé et spécifique à un seul domaine, en puisant les données d'un sous-ensemble de systèmes sources et en obtenant de l'information plus sommaire. Il est fréquent de trouver plusieurs mini-entrepôts de données dans une organisation et afin d'éviter un autre niveau d'informations isolées qui se recoupent (c.-à-d. l'évolution de fichiers indépendants vers des bases de données indépendantes, vers des mini-entrepôts indépendants!), on recommande habituellement de construire de ces mini-entrepôts en plus d'un entrepôt de données unique à l'échelle de l'organisation. Dans de tels systèmes, les mini-entrepôts offrent de l'information sectorielle plus agrégée et présentent l'information de façon spécialisée, ce qui permet d'accéder plus rapidement à l'information que dans le scénario de l'entrepôt.

Afin de permettre l'extraction de connaissances utiles de l'entrepôt/du mini-entrepôt de données, il faut disposer d'un outil d'aide à la décision comme un générateur de requêtes et de rapports, d'un logiciel de traitement analytique en ligne (OLAP) et d'outils de forage de données :

- a) **Générateurs de requêtes et de rapports** : Ces outils (p. ex. Impromptu, Crystal Report) facilitent la création de requêtes et de rapports en remplaçant l'interface SQL standard basée sur une approche technique (langage SQL) par une interface-utilisateur plus intuitive, habituellement basée sur le langage naturel (p. ex. langage clair) ou par de meilleures interfaces graphiques basées sur les requêtes/rapports. Ils offrent également certaines fonctionnalités complexes permettant d'écrire avec l'interface SQL, ainsi que des capacités de présentation plus avancées. Finalement, ils disposent d'une couche sémantique qui remplace les noms de fichiers et les champs non évocateurs utilisés et intégrés dans la structure même de la base de données (c.-à-d. ce qu'il faut savoir lorsqu'on utilise le langage SQL) par des noms réels et évocateurs faisant partie des activités quotidiennes de l'utilisateur.
- b) **Traitement analytique en ligne (OLAP)** : Il s'agit là de la catégorie la plus populaire d'outils d'aide à la décision. Ces outils offrent des fonctionnalités uniques permettant d'explorer d'importantes quantités de données de façon rapide, intuitive et interactive. Ce forage ponctuel des données basé sur la découverte est possible grâce à la structure multidimensionnelle des entrepôts de données (appelé hypercube). En effet, dans cette structure, l'utilisateur peut passer directement d'un niveau détaillé d'information à un niveau plus agrégé/synthétisé (faire un zoom avant ou arrière) (*voir figure 3A*), naviguer d'une catégorie d'information à une autre (*voir figure 3B*) et faire des corrélations avec les données, les filtrer, en faire la coupe, etc. Les outils OLAP possèdent des fonctions automatiques permettant de se concentrer sur les tendances et les exceptions prévues, mais leur principale force est de permettre la navigation entre les différents formats d'affichage de la même information (tableaux, graphiques standard et 3D, cartes). Les outils OLAP les plus connus sont Powerplay et Business Objects.

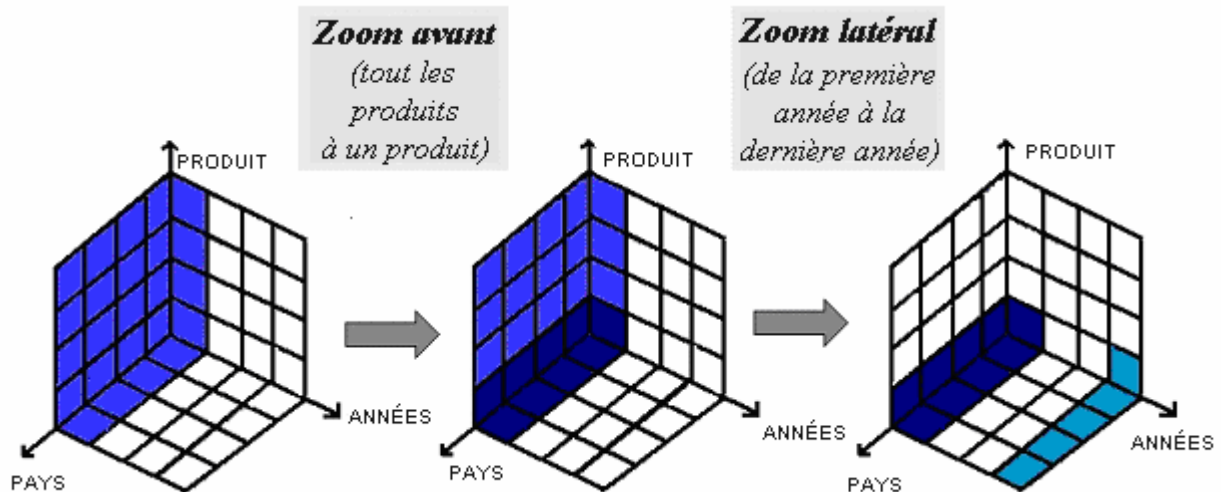


Figure 3 : Navigation OLAP

- c) **Forage de données** : Cette catégorie d'outils de découverte de connaissances vise à automatiser la recherche de structures, de corrélations ou de tendances cachées dans de grands hypercubes et à faire automatiquement des prédictions basées sur les données historiques. Le forage de données est complémentaire aux produits OLAP moins coûteux, dans le cas où il est demandé par l'utilisateur et constitue un travail particulièrement pénible. Parce qu'ils sont «automatiques», ces outils ont mené à la découverte de structures inattendues et complexes, et accéléré le forage d'entrepôts volumineux. Pour accomplir leur travail, les outils de forage de données utilisent certaines techniques complexes comme le réseau neuronal, les arbres de décision, les algorithmes génétiques, l'induction de règles et le groupement à liens simples.

Il est de plus en plus courant de trouver ces technologies déjà incorporées dans les logiciels statistiques et les SGBD, ce qui en améliore les qualités d'aide à la décision. Les produits d'aide à la décision sont souvent des modules complémentaires qui viennent se greffer aux moteurs de recherche du SGBD ou de l'entrepôt/du mini-entrepôt spécialisé et qui sont offerts par des tierces

parties. Toutefois, on remarque actuellement une tendance dans les principaux serveurs universels à intégrer à la fois les capacités multidimensionnelles et les applications frontales d'aide à la décision.

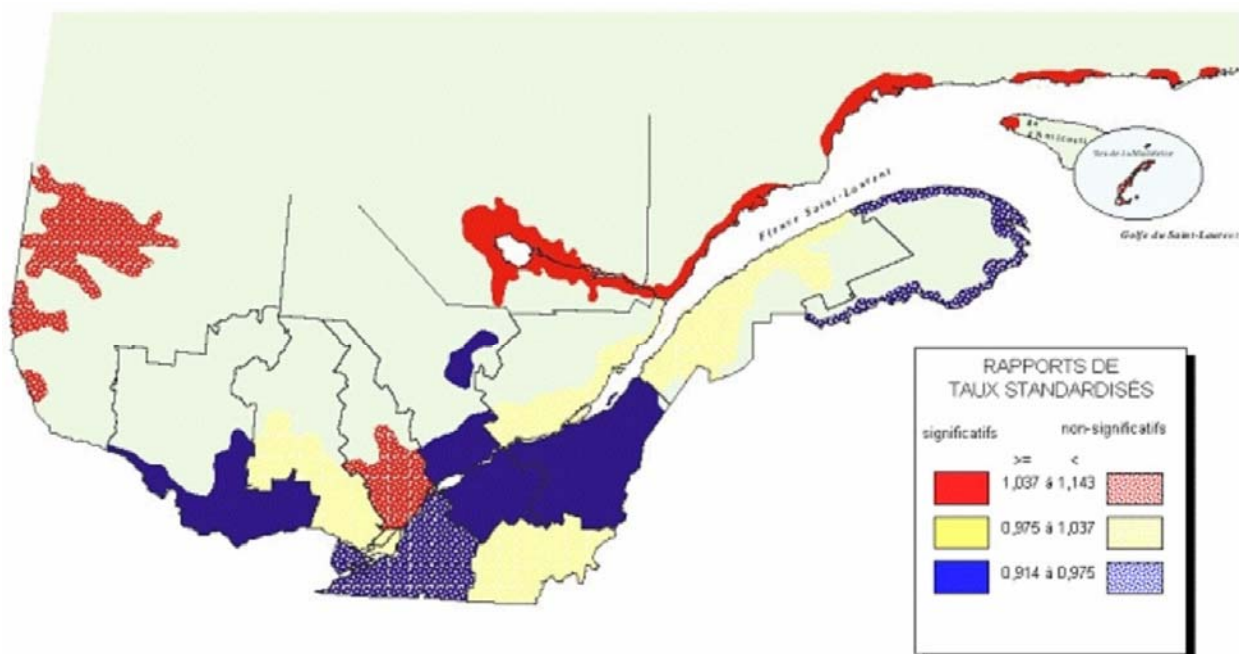
Finalement, il est possible de rendre les entrepôts et les mini-entrepôts de données accessibles par le Web en associant au navigateur un générateur de rapports et de requêtes, ainsi qu'un système d'information pour dirigeants (outil de rapport réactif en mode lecture sur le modèle du tableau de bord). On trouve également sur le marché la technologie nécessaire à l'implantation de serveurs OLAP et des moteurs de forage de données fonctionnant sur le Web.

Systemes d'information géographique et technologies associées :

Les organisations possèdent de l'information qui comprend habituellement certaines données géographiques (adresse, code postal, pays, province/état ou emplacement sur une carte de coordonnées géographiques). Par le passé, beaucoup d'organisations ont tenté de faire un meilleur usage de ces données géographiques dans le processus de prise de décision (p. ex. analyse de marché, mise en place de nouveaux services, études d'impact sur l'environnement, plan d'aménagement du territoire). C'est au début des années 1980 que la convergence de la cartographie numérique et des systèmes de gestion des bases de données (SGBD) a donné naissance aux premiers **systemes d'information géographique** (SIG) commerciaux comme ArcInfo et Intergraph MGE. Ces produits ont permis aux organisations de représenter leurs données tabulaires sur des cartes numériques et de produire ainsi des cartes thématiques (figure 4). Rapidement, des fonctions d'analyse spatiale ont été ajoutées. Entre le milieu des années 1980 et le milieu des années 1990, on a introduit certaines fonctionnalités comme la création de corridors de données spatiales en couches permettant de dresser des profils démographiques pour

Bédard, Y. & W. Henriques, 2003, *Technologies de pointe de l'information pour la surveillance de la santé environnementale, Environnement et santé publique, Edisem, Vol. Chapitre 36, 27p.*

une zone géographique spécifique (Heitgerd, 1994), l'intersection spatiale permettant d'identifier les régions favorables à la prolifération d'un vecteur de maladie (Byron Wood, JPL CA), et l'analyse de réseau (p. ex. trajet le plus court, acheminement, définition des zones couvertes pour un délai de transport donné). Combinées aux technologies d'information géographique comme l'association d'adresses et les systèmes de positionnement global (GPS), la cartographie numérique et l'analyse spatiale marqueront la prochaine révolution dans les technologies de l'information, en modifiant notre façon d'examiner les ressources de données pour la surveillance de l'environnement et la santé. La récente décision de fournir des coordonnées précises à l'aide du système GPS sans l'introduction d'une erreur de 100 mètres témoigne de la grande confiance accordée à l'analyse basée sur l'emplacement déjà accessible à tous. Sont également disponibles certains outils perfectionnés d'analyse spatiale des données, y compris les progiciels de géostatistiques, qui assistent les utilisateurs dans l'examen des données spatiales (Konstanine 1999).



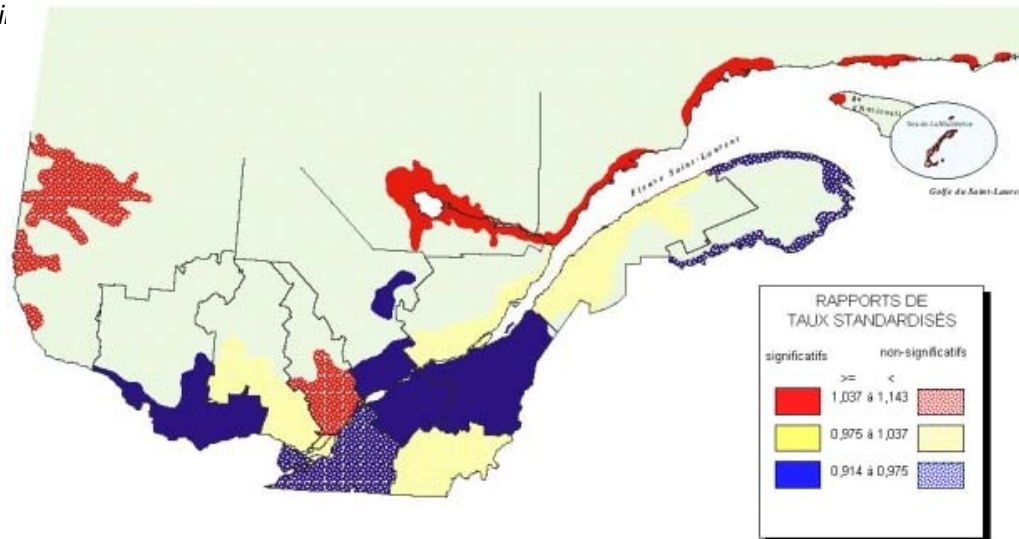


Figure 4 : Exemple d'une saisie d'écran montrant une carte thématique (cancer pour la province de Québec)

Depuis le milieu des années 1990, nous avons été témoin d'une pénétration du marché de masse par les produits dérivés des SIG, phénomène qui a été rendu possible par l'introduction d'ordinateurs personnels plus rapides et la baisse de prix des fichiers de données et des SIG de table. Les utilisateurs non traditionnels des SIG découvrent rapidement les avantages de représenter leurs données sur des cartes, en les superposant sur d'autres cartes pour en faire l'analyse spatiale (voir Longley et al 1999 pour plusieurs exemples).

On constate la même tendance de pénétration de marché pour les grandes organisations et les produits haut de gamme. En effet, la capacité des **serveurs universels à gérer des données géographiques** et à afficher des cartes avec des **visionneurs peu coûteux** dans une architecture client-serveur (figure 5) exerce une importante influence sur le marché. C'est notamment le cas de produits tels ArcView, Geomedia et MapInfo.

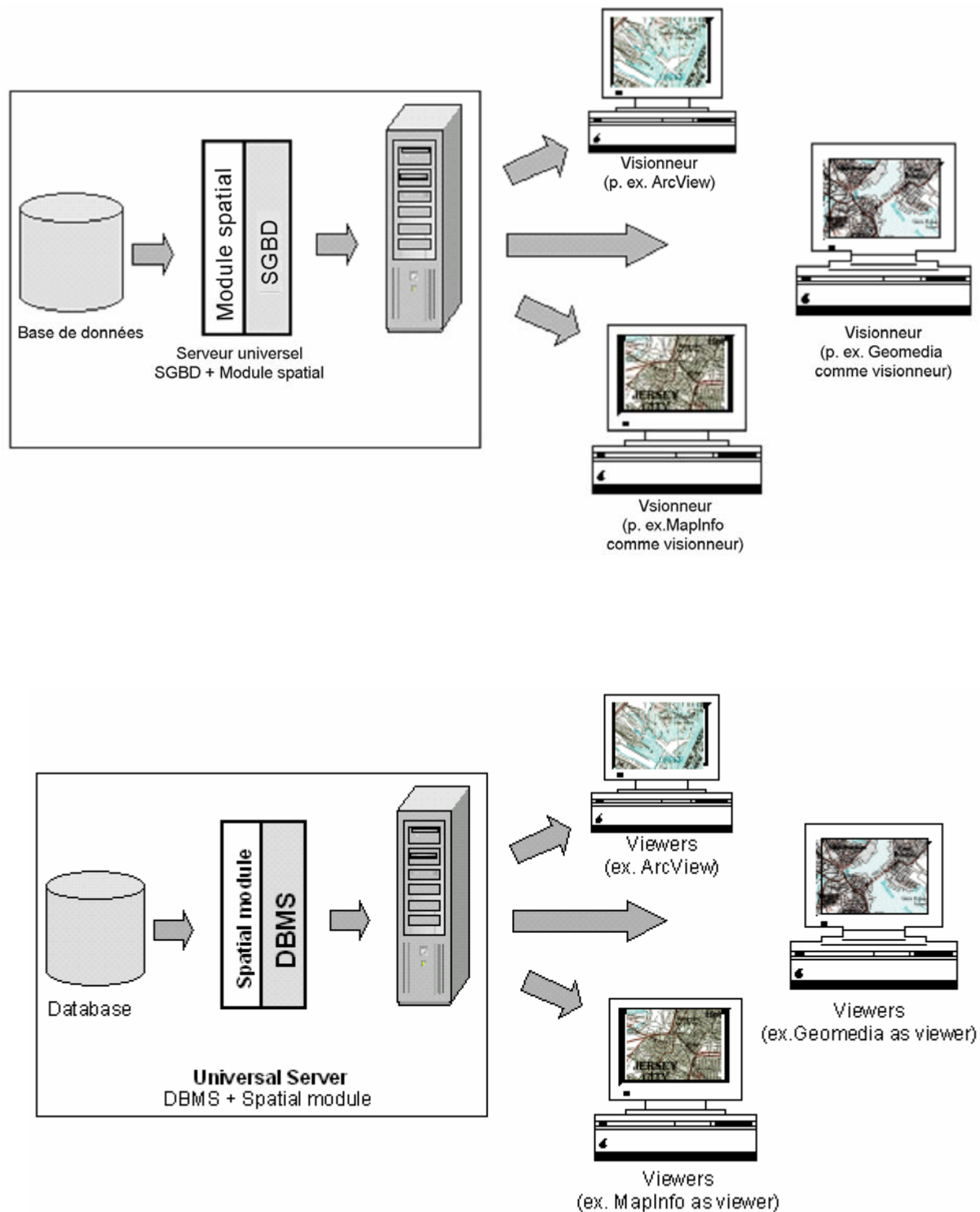


Figure 5 : Serveur universel et visionneurs

Un phénomène similaire se produit avec les **serveurs Web à vocation géographique** qui offrent la possibilité de créer des cartes sur le Web (figure 6). Ces serveurs, dont quelques exemples

Bédard, Y. & W. Henriques, 2003, *Technologies de pointe de l'information pour la surveillance de la santé environnementale, Environnement et santé publique, Edisem, Vol. Chapitre 36, 27p.*

sont Geomedia WebMap, MapXtreme, MapGuide, Model Server et Internet Map Server, sont utilisés pour différents types d'applications (localisation d'adresses, apprentissage en ligne, planification de voyages, etc.) et pour enrichir les technologies traditionnelles du Web. Deux exemples de ces technologies traditionnelles sont (1) les **bibliothèques numériques de données géographiques** qui permettent l'accès, la recherche et le téléchargement de cartes numériques, de photos aériennes et d'images satellites à partir d'un site du gouvernement ou de commerce électronique, et (2) le **L-commerce** (pour «location-commerce» en anglais, ou commerce électronique géolocalisé) qui fournit des cartes personnalisées indiquant un emplacement spécifiquement demandé par l'utilisateur et la route à emprunter pour s'y rendre.

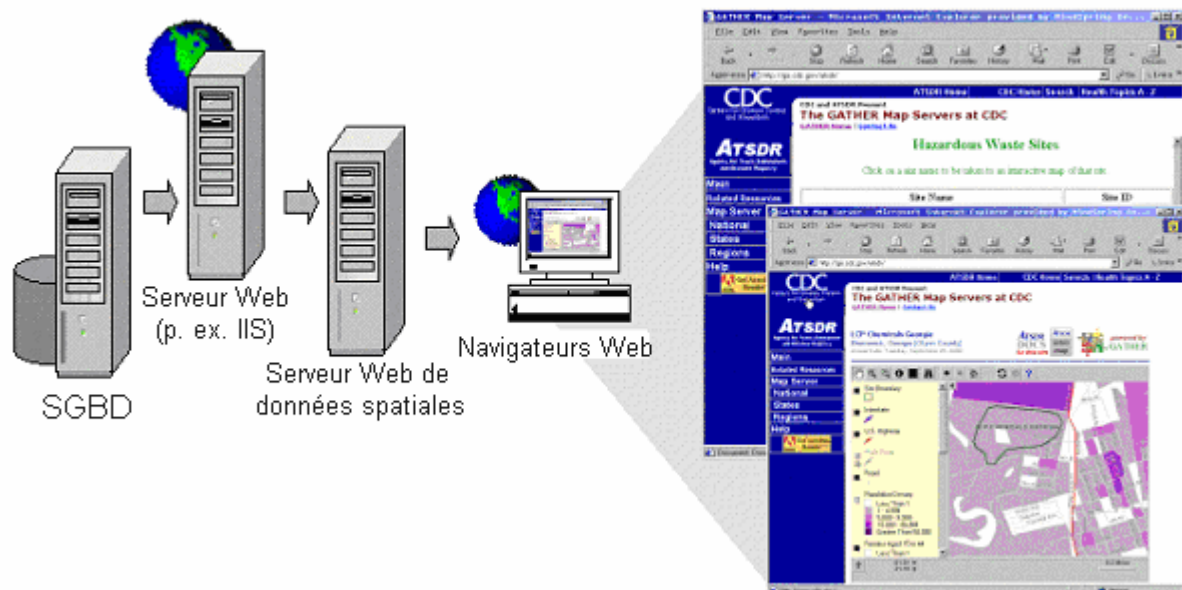


Figure 6 : Exemple d'un serveur Web de données spatiales, le serveur de cartes GATHER du CDC

Le marché des SIG se dirige petit à petit vers des normes ouvertes qui seront acceptées à un niveau international (p. ex. ISO et OpenGIS Consortium), des solutions d'interopérabilité (p. ex. OGDI, ou Open Geospatial Data Store, qui est une suite logicielle à utiliser directement dans les

Bédard, Y. & W. Henriques, 2003, Technologies de pointe de l'information pour la surveillance de la santé environnementale, Environnement et santé publique, Edisem, Vol. Chapitre 36, 27p.

applications afin de lire et d'intégrer les données spatiales à la volée) et des outils très efficaces pour fusionner les données géographiques (p. ex. FME, ou Feature Manipulation Engine de Safe Software, qui est un logiciel de nettoyage/d'intégration par lots très rapide pour les données spatiales) permettant d'intégrer ou de traiter des données géographiques provenant de diverses sources. C'est d'ailleurs pourquoi les premiers projets d'entrepôts, de traitement analytique et de forage des données spatiales quittent les laboratoires de recherche pour faire leur entrée sur le marché de l'application. Les SIG et les technologies qui y sont associées ont maintenant intégré le marché des technologies de l'information grand public.

Aperçu des besoin en technologies de l'information pour la surveillance de la santé environnementale

Au cours des 20 dernières années, on a constaté une importante amélioration dans les technologies de l'information. De plus en plus d'information est disponible pour les millions d'internautes soucieux de connaître l'impact des rejets dans l'environnement sur la santé des habitants de la planète. Les bases de données permettent d'accéder à des rapports sommaires annuels sur les rejets chimiques, préparés par exemple avec l'inventaire des rejets toxiques de la US Environmental Protection Agency. Les utilisateurs qui disposent d'un navigateur peuvent visualiser des cartes et des données concernant les rejets industriels pour une région ou un code postal en particulier, à l'aide de sites Web comme celui du système de cartes de pointage du Environmental Defense Fund (Figure 7).

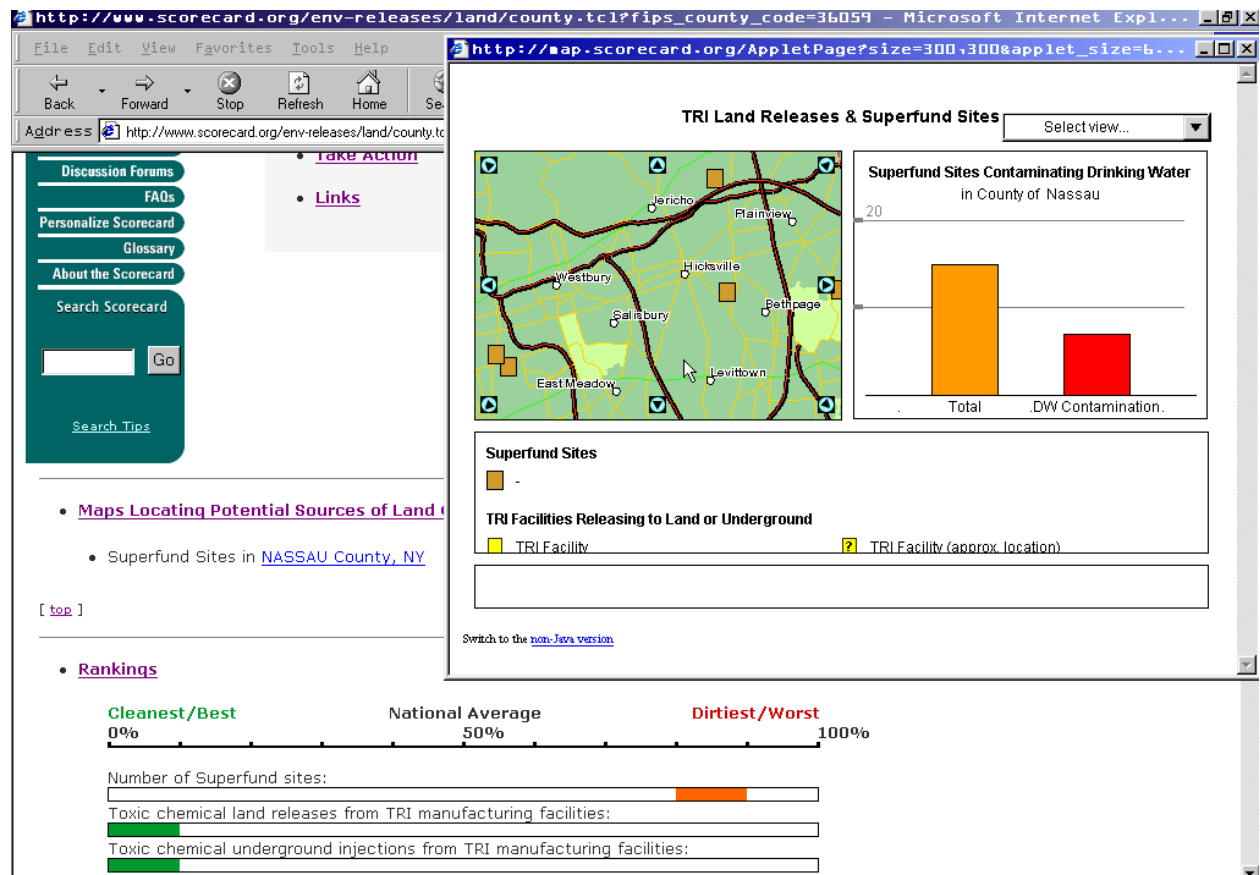


Figure 7 : Système de cartes de pointage du Environmental Defense Fund, qui montre à la fois l'information en formats cartographique et tabulaire.

Même si cette information est utile, et en admettant qu'une telle interface encourage l'accès libre à de l'information critique pour le maintien de la santé de l'environnement, il n'est pas suffisant de signaler les volumes de produits chimiques. On doit avoir une façon plus appropriée de mesurer l'impact des volumes de produits chimiques, ou plutôt la toxicité des composés, données largement dérivées de recherches toxicologiques en laboratoire et d'études sur l'exposition aux produits chimiques en milieu de travail. Sur le plan technique, il serait possible de concevoir une interface qui permette de relier les données de temps et de volume des rejets de produits chimiques à un algorithme analysant le risque pondéré d'effets secondaires sur la santé de chaque substance dans une certaine zone géographique. Cela exigerait toutefois l'expertise de

plusieurs disciplines convergentes. On a souvent dit que les SIG et les technologies d'analyse des données spatiales servaient davantage à développer des questions plus ciblées qu'à répondre à des questions complexes. Néanmoins, les SIG montrent à l'utilisateur toute l'information de façon unique, dans un format facilement compris par les citoyens intéressés à en savoir plus sur leur environnement. Le défi technologique consiste à mettre au point des outils qui convertissent des bases de données volumineuses contenant de l'information sur les rejets environnementaux dans des systèmes de bases de données relationnelles (SGBDR) en cartes qui illustrent le risque pour la santé humaine dans une certaine zone géographique. Un exemple de cette approche est donné à la Figure 8 : on y voit des sites de déchets dangereux (Superfund) dans la région des Grands Lacs, données pondérées par la présence de produits chimiques représentant un plus grand risque pour la santé. Ces données sont basées sur la dose relative requise pour provoquer un effet nuisible sur la santé (Roney et al 1998). En associant l'information sur la toxicité aux données de concentration, on peut développer des cartes qui représentent plus fidèlement les zones géographiques problématiques sur le plan de la santé environnementale. Cet exemple simple utilise l'approche décrite pour interroger les sources d'information (base de données HazDat du ATSDR) afin de fournir une seule carte des sites répondant à certains critères spécifiques, mais omet de considérer les sources additionnelles de contamination environnementales : rejets industriels actuels, absorption des produits chimiques par consommation de poissons, de viandes et de grains, ainsi que l'impact sur la santé d'habiter en région urbaine ou à proximité de routes et autoroutes achalandées. Plus on ajoute d'information pour obtenir un portrait plus complet de l'impact des produits chimiques sur l'environnement, plus le portrait se complique. Certes, il est utile de connaître les volumes de produits chimiques dispersés dans l'environnement, mais des outils sont encore nécessaires pour répartir ces données sur une zone géographique par classe chimique (p. ex. métaux lourds) et par système organique

Bédard, Y. & W. Henriques, 2003, Technologies de pointe de l'information pour la surveillance de la santé environnementale, Environnement et santé publique, Edisem, Vol. Chapitre 36, 27p.

atteint (p. ex. neurotoxines). L'émergence d'outils informatiques qui prennent les renseignements connus sur les produits chimiques spécifiques et les convertissent en carte interactive décrivant les zones de risque pour la santé en résumant et en pondérant les renseignements sur les rejets offrira une nouvelle façon d'aider le commun des mortels à déterminer l'impact du monde industriel sur la santé dans sa communauté.

Des bases de données conviviales sont déjà en construction dans de nombreux pays : Catelan et al (2000) en citent d'ailleurs de nombreux exemples dans une étude détaillée. Ces bases, principalement utilisées en conjonction avec les SIG et les outils OLAP pour le forage et la visualisation des données, facilitent les communications en utilisant des cartes pour superposer géographiquement les données, les analyser et y repérer les tendances.

De plus, les applications de pointe et les interfaces-utilisateurs basées sur l'entreposage des données spatiales et le traitement analytique des données spatiales, comme les projets SPHINX (Alberta) et ICEM-SE (Québec), fourniront ce genre d'outils qui donneront accès à de l'information agrégée, à des données détaillées, à de l'information externe ou interne (avec les permissions d'accès appropriées), ainsi qu'à des fichiers de données agrégées ou indépendantes. Des interfaces-utilisateurs simples et rapides sont des éléments clés de tels projets novateurs. En plus, l'ICEM-SE inclura une bibliothèque électronique de données géospatiales pouvant être interrogées graphiquement sur une carte : cela permettra de connaître l'existence d'information intéressante pour une région spécifique. Dans une récente étude canadienne sur les besoins (Bédard et al 2000), on a établi comme priorité la capacité d'identifier un type de problème de santé dans une zone géographique spécifique.

Bédard, Y. & W. Henriques, 2003, *Technologies de pointe de l'information pour la surveillance de la santé environnementale, Environnement et santé publique, Edisem, Vol. Chapitre 36, 27p.*

Cette étude a également mis en relief le fait que les spécialistes canadiens de la santé utilisaient la technologie SIG pour rehausser la visualisation et la communication de leurs données, en améliorer l'analyse et faciliter la prise de décision. Par contre, ces spécialistes reconnaissent que la technologie n'est pas utilisée à son plein potentiel en raison de la difficulté d'obtention des résultats, de la complexité du logiciel, du manque de formation et aussi du manque d'engagement organisationnel.

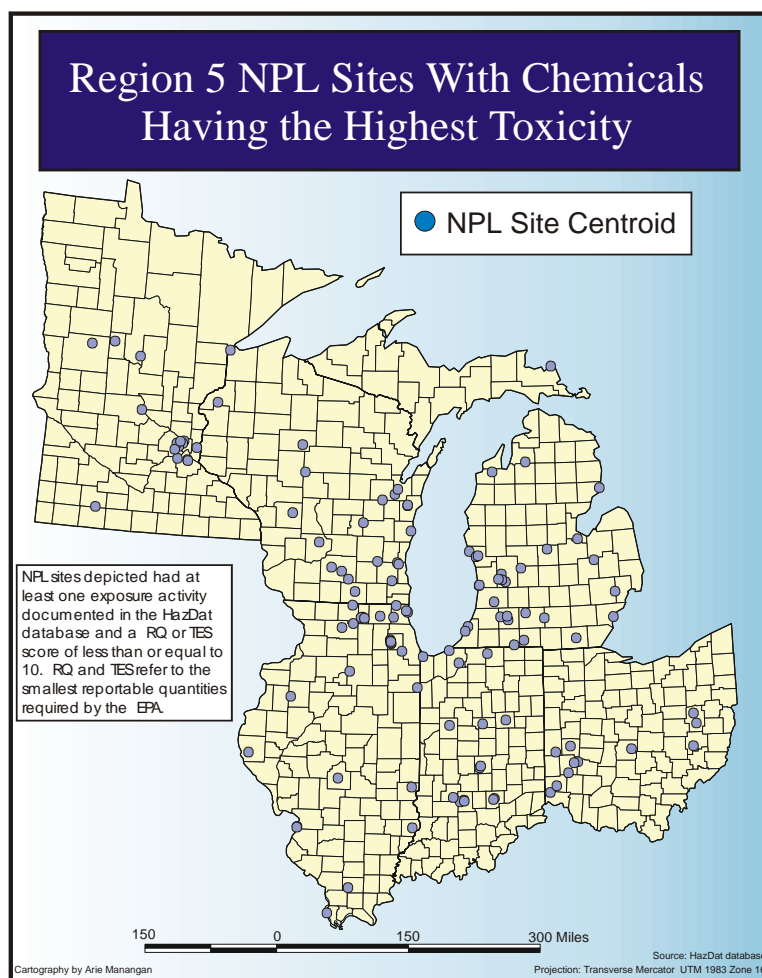


Figure 8 : La base de données HazDat du ATSDR a été interrogée afin d'identifier seulement les sites présentant des expositions hors site documentés et contenant des substances considérées comme les plus toxiques.

Applications, avantages et ressources nécessaires

À mesure que notre planète devient plus petite et plus industrialisée, les populations plus sensibles aux effets néfastes de l'exposition aux produits chimiques devront élire domicile là où les substances chimiques anthropiques sont plus rares. Les personnes âgées ou en âge de procréer, ainsi que nos jeunes, ont le droit de savoir où les taux de pollution dépassent le seuil de nocivité pour la santé. Une fois ces endroits identifiés, des mesures provisoires de protection de la santé devront être prises car ces produits chimiques demeureront vraisemblablement dans l'environnement de nombreuses années avant que la décontamination se produise. L'information spatiale pourra alors servir à identifier et à caractériser la population de ces régions et à encourager les individus et les communautés à contre-carrer la charge chimique additionnelle en adoptant des modes de vie plus sains, comme consommer des repas plus nutritifs, arrêter de fumer, faire régulièrement de l'exercice et quitter périodiquement l'environnement urbain. Il sera possible d'opter pour une approche plus pratique à la protection de la santé une fois que nous aurons fait le nécessaire pour examiner toutes les contributions régionales à l'exposition aux produits chimiques, et non seulement les apports industriels. Ainsi, l'aménagement du territoire pourra avoir d'importantes répercussions sur l'état de santé, car les charges chimiques dans l'environnement seront considérées comme un composant de l'aménagement global, limitant de ce fait l'impact combiné des diverses activités industrielles sur une communauté en considérant l'interaction des multiples installations sur de grandes zones géographiques. De façon similaire, l'effet des conditions topographiques (proximité de grandes étendues d'eau, présence de vallées profondes pouvant restreindre la dispersion des produits chimiques) doit être examiné dans un contexte géographique pour considérer les effets bénéfiques ou nuisibles de certains biomes sur le processus de régénération de l'environnement. Cette vision holistique représente certes un défi important, du fait qu'il faut trouver, obtenir, combiner et analyser l'information en

provenance de différentes sources. Les plus récentes technologies de l'information comme celles décrites précédemment doivent être implantées à grand échelle, facilement accessibles et utilisables par l'utilisateur peu versé en matière de technologies.

Défis d'une mise en oeuvre réussie et suggestions pour orientations futures

L'évaluation de la dose à laquelle un composé devient nuisible pour la santé humaine n'est pas une tâche facile. Les toxicologues doivent considérer une myriade de facteurs quand vient le temps d'extrapoler les données obtenues lors d'études sur les animaux aux humains (Williams-Johnson et al 1996). Les réseaux informatiques et les ordinateurs centraux sources contiennent des milliards de registres d'échantillons analytiques recueillis dans des sites de déchets dangereux, qui documentent les niveaux de polluants prioritaires dans l'air, l'eau et le sol. Au lieu de considérer la toxicité individuelle des composés en parties par milliard, il pourrait être plus raisonnable de considérer la toxicité *relative* des composés (possiblement l'effet sur les organes cibles) comme approche innovante dans l'évaluation de la toxicité des produits chimiques sur une vaste zone géographique. On peut rehausser les études épidémiologiques traditionnelles en considérant les possibilités qu'offrent les SIG de pointe, comme les statistiques spatiales et la superposition des données géographiques sur d'autres ensembles de données (p. ex. résultats du modèle d'exposition, données socio-démographiques, utilisation du terrain, topographie). De telles études basées sur les SIG caractérisent mieux l'exposition et les niveaux de risque et indiquent les zones où cela vaudrait le coup d'approfondir la recherche. Ces zones peuvent faire l'objet d'un examen exhaustif à l'aide de données plus détaillées sur la santé et l'environnement existant dans les systèmes sources, une fois placées dans un contexte géographique. En d'autres mots, ces systèmes offrent la capacité d'aller plus loin que l'examen des indicateurs de santé individuels (p. ex. regarder le taux de cancer dans un comté spécifique et

comparer les résultats avec ceux obtenus dans la province ou le pays pour déterminer où se situe la région par rapport au taux d'atteinte national) et aident à réduire l'information géographique biaisée dans les études écologiques ou épidémiologiques cas-témoin.

Cependant, pour comprendre les défis que représentent l'intégration des récentes technologies de l'information dans le travail quotidien des hygiénistes du milieu, il faut regarder les activités associées à l'accès, à l'évaluation de la pertinence et à la normalisation des données (Gosselin et al 2000). Autrement dit, le seul fait de connaître l'existence de ces données est un défi, alors on peut facilement imaginer qu'il en va de même pour l'obtention (coût, confidentialité) et la transformation de ces données en format utilisable (restructuration, enregistrement, validation, agrégation, géocodage, etc.).

D'autres défis importants identifiés par Gosselin et al (2000) concernent l'accès à la formation (à la fois formelle et informelle, y compris l'accès à un soutien technique) et l'obtention de fonds destinés au maintien et au développement des aptitudes d'utilisation des outils fondés sur les preuves. Ainsi, le vrais défis qui restent à surmonter sont plus administratifs et politiques que techniques, davantage associés à l'identification et à l'accès des données qu'à la technologie, concernant plus l'utilisation des systèmes, la formation et le support que la technologie... car la technologie requise est déjà là.

Conclusions

Plus les technologies informatiques deviennent conviviales, plus le fossé qui sépare les utilisateurs doués en informatique et ceux qui le sont moins devrait diminuer. Nous sommes maintenant à une époque où les aptitudes en informatique occupent une grande place dans notre quotidien. La technologie vient rapidement changer le type et la quantité d'information

accessible sur Internet et le World Wide Web, les systèmes modernes de gestion de bases de données et les serveurs universels. De plus, les outils cartographiques comme les SIG permettent de dresser un portrait simple des données complexes qui concernent la santé à l'échelle de la planète. D'autres technologies comme l'entreposage de données offrent une information à différents niveaux d'agrégation, fournissant la base nécessaire pour les outils et l'analyse de prise de décision. La pertinence de ces technologies et la volonté de les utiliser à l'avenir ont été démontrées dans une récente étude pan-canadienne des besoins, à laquelle ont participé des spécialistes de la santé (Gosselin et al 2000), ainsi que dans les nombreux exemples fournis par Catelan et al (2000). Dans un avenir plutôt rapproché, nous serons capable de parler à notre ordinateur et de lui demander «montre-moi le meilleur endroit pour vivre avec ma famille, donne-moi des statistiques locales et régionales sur la santé dans cette région par rapport aux régions avoisinantes, et indique-moi le meilleur chemin pour accéder aux services de santé dans les environs», tout cela selon certains critères pré-déterminés, et l'écran affichera les solutions possibles à la question complexe. Et tout cela sera basé sur de l'information dont nous disposons déjà, mais que l'on commence seulement à faire interagir pour obtenir des réponses à notre portée. Il est important de se rappeler ici que l'accès aux données, à la formation et au soutien technique n'est principalement pas un défi technologique, mais plutôt un défi organisationnel, administratif et politique. La technologie et l'expertise sont déjà disponibles, les besoins existent déjà et d'importants volumes de données sont déjà en cours d'utilisation, dans quelques pays. Il faudra disposer d'une vision à long terme, de nombreuses collaborations et de solides partenariats pour arriver à mettre en valeur le plein potentiel des technologies de pointe de l'information dans les activités en santé environnementale de ces pays. Toutefois, pour donner une portée mondiale à cette vision, on devra déployer des efforts considérables dans ces trois sphères que sont l'accès aux données, la formation et le soutien technique.

Bédard, Y. & W. Henriques, 2003, *Technologies de pointe de l'information pour la surveillance de la santé environnementale, Environnement et santé publique, Edisem, Vol. Chapitre 36, 27p.*

Références

Bédard Y, Gosselin P, Jerrett M, Elliott S, Mowat D, Moore J, Goddard M, Catelan R, Gingras A et Poitras P. *Recent Technological Trends vs Users' Needs in Health Surveillance, a Canadian Study*. Conférence internationale sur l'information en santé «Infocus 2000», organisée conjointement par l'Institut canadien d'information sur la santé et la Canada's Health Informatics Association, Vancouver, Canada, June 24-27 2000, 9 pp.

Bédard Y, Merrett T et Han J. «Fundamentals of Spatial Data Warehousing for Geographic Knowledge Discovery», dans *Geographic Data Mining and Knowledge Discovery*, coll. Research Monographs in GIS (éditeurs de la collection : Peter Fisher, Jonathan Raper, Taylor & Francis). H. Miller et J. Han (éditeurs), 2000 (en cours de publication).

Brackett MH. *The Data Warehouse Challenge: Taming Data Chaos*. New York, John Wiley & Sons, 1996, 579 pp.

Catelan R, Gosselin P et Bédard Y. «Évolution récente de la pénétration de la géomatique dans le domaine de la surveillance en santé : revue bibliographique», dans *Revue Internationale de Géomatique*. Paris, Éditions Hermès, 2000.

Date C.J. *An Introduction to Database Systems*, 7^e édition. Addison-Wesley, 2000, 938 pp.

Gisler W. «The Uses of Spatial Analysis in Medical Geography: A Review», dans *Social Science and Medicine* 1986;23:963–73.

Bédard, Y. & W. Henriques, 2003, *Technologies de pointe de l'information pour la surveillance de la santé environnementale, Environnement et santé publique, Edisem, Vol. Chapitre 36, 27p.*

Gosselin P., Bédard Y, Jerrett M, Elliott SJ, Catelan R, Poitras P et Gingras A. *GIS and OLAP in Health Surveillance: Needs Analysis for Successful Integration*. Rapport final préparé pour Santé Canada, Ottawa, 10 février 2000, 72 pp.

Inmon WH, Richard D et Hackathorn D. *Using the Data Warehouse*. New York, John Wiley & Sons, 1996, 285 pp.

Kulldorff M. «Statistical Methods for Spatial Epidemiology: Tests for Randomness», dans *GIS and Health in Europe*. Ed. A Gatrell, M Löytönen, London, Taylor & Francis, 1998.

Longley PA, Goodchild MF, Maguire DJ et Rhind DW. *Geographical Information Systems: Principles, Techniques, Applications, and Management*, 2^e édition. New York, John Wiley & Sons, 1999, 1101 pp.

Petridou E, Revinthi K, Alexander FE, Haidas S, Kolioukas D, Kosmidis H, Piperopoulou F, Tzortzatos F, Trichopoulos D. «Space-Time Clustering of Childhood Leukemia in Greece: Evidence Supporting a Viral Etiology», dans *British Journal of Cancer* 1996;73:1278–83.

Poe V. *Building a Data Warehouse for Decision Support*. Prentice Hall, 1995, 210 pp.

Roney N, Henriques WD, Fay M, Holler J, Susten S. «Determining Priority Hazardous Substances Related to Hazardous Waste Sites», dans *Toxicology and Industrial Health* 1998;14(4) :521-31.

Williams-Johnson MM, Henriques WD et Fay RM. «Investigating Ratios of Health Effect Levels Using ATSDR's HazDat Database: Extrapolation Methodologies in Quantitative Risk Assessment», dans *J. Clean Technol., Environ. Toxicol. & Occup. Med.* 1996;5(4):347-60.