Université de Montréal

Nouveau support de visualisation spatio-temporelle pour faciliter l'exploration et le partage de données environnementales

SFMN GeoSearch : un outil pour la recherche en foresterie au Canada

par
Rodolphe Gonzalès

Département de Géographie
Faculté des Arts et Sciences

Mémoire présenté à la Faculté des études supérieures
en vue de l'obtention du grade de M.Sc. en géographie

Avril 2009

Université de Montréal
Faculté des Arts et Sciences

Ce mémoire intitulé :

Titre : Nouveau support de visualisation spatio-temporelle pour faciliter l'exploration et le partage de données environnementales

Sous-titre : SFMN GeoSearch : un outil pour la recherche en foresterie au Canada

présenté par
Rodolphe Gonzalès

Département de Géographie
Faculté des Arts et Sciences

a été évalué par un jury composé des personnes suivantes :

Thora Hermann, présidente-rapporteuse
Jeffrey A. Cardille, directeur de recherche
Lael Parrott, codirectrice de recherche
Pierre J. H. Richard, membre du jury

Accord des coauteurs de l'article

L'article inclus dans le présent mémoire a été rédigé conjointement par Rodolphe Gonzalès, Jeffrey Cardille, Lael Parrott, Caroline Gaudreau et Gaël Deest. Cet article, intitulé « An interactive cartography and ecoinformatics tool to facilitate the exchange and visualization of Canada-wide forestry database », a été soumis à la revue *Ecological Informatics*. Sa date de publication n'est pas connue à ce jour.

À titre de coauteur, je suis d'accord pour que Rodolphe Gonzalès inclue dans son mémoire de maîtrise intitulé « nouveau support de visualisation spatio-temporelle pour faciliter l'exploration et le partage de données environnementales » l'article identifié ci-dessus.

Coauteur(e)	Signature	Date
Jeffrey Cardille	[information retirée / information withdrawn]	06-08-2009
Lael Parrott	[information retirée / information withdrawn]	28-04-2009
Caroline Gaudreau	[information retirée / information withdrawn]	26 avr. 2009
Gaël Deest	[information retirée / information withdrawn]	26 avril 2009

Résumé

Dans divers domaines des sciences, et particulièrement en écologie, les chercheurs accumulent dans le cadre de leurs travaux individuels des quantités de plus en plus grandes de données. Celles-ci sont potentiellement sous-exploitées : alors qu'elles pourraient être réutilisées par d'autres chercheurs dans d'autres projets, elles demeurent le plus souvent cantonnées aux laboratoires des scientifiques qui les ont produites. Cependant, nous voyons depuis les années 1990 des initiatives visant à mettre sur pied des politiques de partage de ces données. C'est ainsi que, mettant en œuvre les opportunités offertes par les technologies de l'information, certains grands réseaux de recherche se sont lancés dans des projets de partage de grande envergure. Ils ont constitué à travers les années — et mettent à disposition du public par le biais de serveurs web ou ftp — de grandes bases de données composées des contributions individuelles de leurs membres. Dans le même temps, de nouveaux outils cartographiques, les « globes virtuels » (tels que Google Earth ou NASA World Wind), offrent — de par leur caractère interactif, convivial et expressif — des opportunités inédites quant à la représentation de données scientifiques multidimensionnelles sur un support géographique. Aujourd'hui, ces outils pourraient être avantageusement intégrés aux services de distribution déjà existants ; ils pourraient servir, dans le cadre du partage de données possédant des informations de localisation (ce qui est souvent le cas pour les données environnementales), de support efficace à leur visualisation.

Le travail présenté ici concerne l'amélioration, par le biais de la cartographie dynamique et intuitive des globes virtuels, d'un des aspects des systèmes de

partage : l'interface entre le réutilisateur potentiel et la base de données qu'il souhaite consulter. Ce travail a mené à la mise au point de SFMN GeoSearch, une application intégrée créée pour un réseau pancanadien de recherche en foresterie. SFMN GeoSearch est un outil générique capable de gérer, d'afficher et de rendre disponible au téléchargement tout échantillon quantitatif possédant des informations de lieu et de temps. Le système se divise en deux grandes parties : une base de données contenant les jeux soumis par les membres du réseau d'un côté, et une application client de l'autre. Cette dernière, conçue pour assister efficacement les scientifiques dans leurs découvertes de jeux de données à réutiliser, est centrée sur un globe virtuel mettant en œuvre des techniques novatrices d'exploration et de visualisation, dans l'espace et dans le temps, de données scientifiques.

SFMN GeoSearch est disponible à cette adresse :

http://meta.geog.umontreal.ca/currentprojects/forests/

Mots-clés : Visualisation spatio-temporelle, Exploration géographique interactive, Globes virtuels, Geobrowser, Partage de données, Système d'Information Géographique.

Abstract

In many fields of science, particularly environmental sciences, researchers are accumulating ever-increasing quantities of field data related to their research projects. These data are potentially underused: they are for the most part kept in the laboratories where they were produced while they could be utilized by other researchers on their own projects. However, since the 1990's, we have seen many initiatives aimed at developing policies for sharing these large amounts of individually collected data. Some major research networks have been taking advantage of the opportunities offered by information technologies to develop large-scale data sharing projects. Through the years, these networks have been gathering large databases built from their members' individual contributions and are now making them available through web or ftp servers. Meanwhile, new cartographic tools, the so-called virtual globes (such as Google Earth or NASA World Wind) are now offering interactive, user-friendly and expressive new opportunities to display multidimensional scientific data on geographical media. Today, these modern cartographic tools could be integrated with distribution systems, to be used as an effective support for the visualization and exploration of georeferenced shared scientific data.

The work presented here focuses on the improvement of a particular aspect of data sharing systems through the dynamic and intuitive cartography of virtual globes: the interface between the potential second-hand user and the database he/she wants to consult. This work led to the development of SFMN GeoSearch, an integrated application created for a Canada-wide network of forestry researchers. SFMN GeoSearch is a generic tool capable of managing, displaying and making

available any sample bearing spatial and temporal information. The system is divided into two main parts: a database of datasets submitted by members of the network on one side, and a client application on the other. The latter, designed to effectively assist scientists in their discoveries of previously collected data to reuse for their own projects, is centered on a virtual globe implementing innovative techniques to explore and visualize large quantities of ecological data in space and time.

Keywords: Spatiotemporal visualization, Interactive geographic exploration, Virtual globes, Geobrowser, Data sharing, Geographic information System.

Table des matières

Liste des figures

Remerciements

Ce travail de deux ans dont le résultat final est présenté ici aurait été impossible sans les contributions, l'implication ou le soutien de nombreuses personnes autour de moi. Je souhaite adresser mes remerciements chaleureux à mes parents Roger et Annie Gonzalès pour leur compréhension et leur soutien sans faille depuis le vieux continent. À ma compagne Melissa White pour ses encouragements au quotidien et, accessoirement, pour la relecture des parties rédigées en anglais. À ma directrice, Lael Parrott, pour ses nombreuses contributions, ses encouragements et son enthousiasme indéfectible. À mon directeur, Jeffrey Cardille (qui est à l'origine de l'idée d'un outil intégré liant cartographie 3D et données partagées) pour ses contributions et commentaires enrichissants, ainsi que pour m'avoir fait confiance en me proposant ce projet. À Caroline Gaudreau et Gaël Deest pour leur enthousiasme, leurs conseils, et leur travail acharné à la concrétisation, ligne de code par ligne de code, de l'application SFMN GeoSearch. Je les remercie particulièrement d'avoir supporté avec une admirable patience les changements survenus au fur et à mesure de la progression de ma recherche et des problèmes techniques rencontrés. À Cristiane Albuquerque Martins, Clément Chion, Guillaume Latombe, Samuel Turgeon, Élise Filotas, Christine Grenier et Raphaël Proulx, membres ou ex-membres du laboratoire de Systèmes Complexes dirigé par Lael Parrott, pour participer quotidiennement à créer une ambiance de travail autant décontractée qu'intellectuellement stimulante. À Flavien Gillié pour sa relecture critique, ses talents linguistiques et sa gentillesse. Au personnel du Département de Géographie de l'Université de Montréal pour, selon les cas, leurs enseignements

ou leur aide précieuse dans les méandres de l'Administration. Enfin, à mon grand-père Jacques Gonzalès à qui je dédie ce travail.

Introduction

La collecte de données scientifiques, qu'elle soit initiée par des organisations gouvernementales, académiques ou privées, est mue par un effort constant depuis des siècles. Avec les avancées technologiques que nous connaissons depuis la seconde moitié du XXe siècle, ces efforts produisent des quantités de plus en plus grandes de données dans la plupart des domaines des sciences, et particulièrement dans les sciences environnementales. Exemple de ces progrès technologiques, l'informatique a à elle seule permis de faire progresser de manière exponentielle les capacités de traitement et d'entreposage des données produites par les scientifiques.

Les usages de la recherche, toutefois, n'ont pas évolué aussi vite que les outils. Typiquement en sciences environnementales, une équipe se rend sur un site d'étude pour collecter des données destinées à répondre à une question de recherche précise. De retour au laboratoire, ces données seront exploitées et, le cas échéant, des résultats seront publiés. Finalement, ces données liées à une étude particulière demeureront le plus souvent dans les ordinateurs de leurs auteurs. Ainsi, nous sommes dans une situation où des jeux de données collectés individuellement pour des recherches ponctuelles sont accumulés à travers les années, et sont potentiellement sous exploitées. Cette situation est dommageable pour la communauté dans son ensemble, car la réutilisation de ces données par d'autres chercheurs pourrait ouvrir des horizons vers lesquels l'auteur des analyses originales ne s'était pas tourné, et permettre des avancées qui profiteraient à la société dans son ensemble.

La communauté scientifique est consciente que la diffusion et le partage de l'information favorisent la création de nouvelles connaissances. Néanmoins, ce n'est que récemment que des initiatives concrètes ont émergé. C'est ainsi que nous voyons des milieux scientifiques s'organiser en réseaux dont le mandat est notamment de promouvoir l'interdisciplinarité des travaux de recherche et le transfert des connaissances parmi les chercheurs. En ce qui concerne les sciences environnementales, ces structures semblent particulièrement adaptées pour répondre à un besoin pressant : les questions auxquelles les chercheurs sont confrontés sont appréhendées selon une approche de plus en plus complexes. De nombreux phénomènes étudiés sont le résultat de différents processus, parfois non linéaires, qui se révèlent à travers les échelles spatiales et temporelles. Leur étude requière ainsi de grandes quantités de données multidisciplinaires étalées dans l'espace et dans le temps, et il est difficile pour des équipes seules de recueillir les données nécessaires à ces recherches.

Pour répondre à ce besoin en données, la mise en place de systèmes de partage efficaces, visant à rendre disponibles au plus grand nombre des jeux de données issus d'études individuelles, se révèle être une solution prometteuse. Si les récents progrès dans la qualité des réseaux informatiques rendent possibles ces échanges de données à grande échelle, le partage de données scientifiques s'avère être un domaine soulevant des obstacles à la fois comportementaux et techniques qui sont traités par des chercheurs issus d'horizons différents. Ces chercheurs apportent des contributions que l'on peut classer grossièrement dans deux disciplines : les sciences sociales d'un côté, et les technologies de l'information de l'autre. La première traite des bénéfices que peut apporter le partage entre scientifiques et identifie les obstacles qui lui sont liés. Elle tente de

dessiner les contours de situations de compromis qui satisferaient à la fois les intérêts particuliers des chercheurs, qui se soucient notamment de leurs droits intellectuels, et l'intérêt général, qui bénéficierait largement d'un partage généralisé des données de chacun. Dans la seconde, les praticiens s'attellent à apporter des réponses aux problèmes révélés notamment par les sciences sociales. Les travaux de ces derniers s'inscrivent dans les progrès de l'informatique, des technologies de l'information et du génie logiciel, convergeant vers la mise au point de systèmes de gestion des données scientifiques et de leurs métadonnées.

Dans le même temps, un autre domaine a profité des progrès de l'informatique grand public. La visualisation de données scientifiques sur support géographique a connu des avancées spectaculaires depuis quelques années. Elle a abouti au début de la décennie à de nouveaux outils de visualisation spatiale connus sous le terme de « globes virtuels ». Le grand public a maintenant accès à des modèles en trois dimensions du globe terrestre ou les cartes ne sont plus représentées de manière projetée, mais sont plaquées sur une maquette virtuelle de la planète. L'utilisateur peut, à l'aide de la souris de l'ordinateur, modifier son point de vue de manière dynamique et fluide et découvrir la Terre dans ses détails. Ce support, sur lequel des couches de données peuvent être superposées, peut ainsi — dans certains cas — constituer, pour des scientifiques souhaitant visualiser rapidement des données géoréférencées, une alternative avantageuse aux coûteux et compliqués Systèmes d'Information Géographiques (SIG).

Au carrefour des avancées en matière de représentation géographique et de partage de données, un espace prometteur se crée. Le travail présenté dans ce mémoire se trouve à cette intersection. Il vise à créer, dans le cadre d'un partenariat avec un réseau pancanadien de recherche en foresterie, un logiciel de

partage de données scientifiques mettant la puissance des nouvelles représentations géographiques au service de l'échange de données quantitatives géoréférencées. Un globe virtuel servira de support à la visualisation, dans l'espace et dans le temps, de données échantillonnées par différents chercheurs à des dates diverses. Sur ce support géographique à trois dimensions, sept dimensions différentes seront visualisées simultanément grâce à une iconographie idoine et à une utilisation originale de la troisième dimension géométrique. Le regroupement sur un support spatio-temporel simple d'emploi de données jusqu'alors dispersées parmi les membres du réseau offrira aux chercheurs une vue générale des travaux effectués. Il permettra d'appréhender les travaux des membres du réseau non seulement comme des efforts ponctuels, mais aussi comme les éléments d'un travail collectif aux frontières spatiales et temporelles beaucoup plus larges.

1. Contexte de la recherche

1.1. Le partage de données scientifiques

> « The demands to share data have also increased in response to a push for interdisciplinary research. It is difficult to pinpoint where the focus on interdisciplinary research first arose, but it has been spurred by the belief that the solution to today's complex, global problems are outside the realm of any one discipline to solve. » (Zimmerman, 2003, p. 2)

Malgré la tradition de collégialité de la communauté scientifique, le partage de l'information est jusqu'à récemment resté ponctuel, car laissé à l'initiative de l'individu (Stanley et Stanley, 1988). Par ailleurs, si les scientifiques sont invités à collaborer, à soumettre des résultats de recherche pour les faire évaluer par leurs pairs ou à rendre disponibles des données ayant mené à ces résultats, les moyens techniques de transmettre rapidement et efficacement de grandes quantités d'informations ont jusqu'à présent manqué. Toutefois, les progrès de l'informatique mettent aujourd'hui les scientifiques dans une situation inédite. Les capacités de stockage des ordinateurs personnels connaissent depuis l'arrivée des premiers disques durs au début des années 1980 une progression exponentielle ; ils offrent aujourd'hui les moyens techniques de stocker, copier, transmettre et manipuler facilement des bases de données de plus en plus volumineuses. De plus, les réseaux informatiques sont, au moins dans les pays riches, de plus en plus efficaces et permettent des transferts rapides de données.

Mais si la technologie rend en théorie possible une meilleure diffusion des données de recherche, des difficultés techniques subsistent, auxquels viennent

s'ajouter d'autres obstacles, culturels ceux-là, sur lesquels les sciences sociales se sont penchées au milieu des années 1980.

1.1.1. Le partage à l'échelle de la communauté scientifique

Parmi les travaux produits à cette époque, Sharing Scientific Data (Fienberg et coll. 1985) cernait, il y a presque 25 ans, bon nombre de points restant valides aujourd'hui. Les auteurs y font une liste synthétique des aspects positifs et des obstacles au partage des données scientifiques. Il est pertinent d'en énumérer certains, car ils constituent la fondation des efforts accomplis, et restants à accomplir, pour parvenir à une diffusion satisfaisante des données de recherche. Parmi les obstacles à surmonter, Fienberg et coll. (1985) identifient :

— les éventuels problèmes techniques liés, par exemple, à l'incompatibilité du matériel, des logiciels, des formats de fichiers contenant les données, etc. ;

— la documentation des jeux de données. Ce problème, qui est celui des métadonnées et de l'ontologie, est aujourd'hui un domaine de recherche très actif ;

— les questions de confidentialité de l'information ;

— le problème du coût. Les pratiques à mettre en œuvre lors de la collecte et du conditionnement des données afin que le partage et la réutilisation de celles-ci soient clairs et aisés impliquent des coûts en temps et en argent non négligeables. Ils peuvent, précisent Fienberg et coll. (1985), d'une part dépasser les bénéfices éventuels et, d'autre part entraîner des gains dont le chercheur/ partageur ne sera que rarement le bénéficiaire direct.

Les auteurs établissent également une liste de points positifs dont la communauté scientifique pourra tirer profit, notamment :

— la mise à disposition des données de recherche facilitera, par le biais de nouvelles analyses, la vérification et l'amélioration des résultats (on peut citer, pour illustrer ce propos, Campbell et coll. (2002), qui établissent dans une étude visant à évaluer l'impact de la rétention de données en génétique que plus d'un quart des chercheurs interrogés n'ont pu confirmer les résultats d'une recherche publiée à cause d'un refus de partage de la part d'un collègue) ;

— elle pourra également promouvoir de nouvelles recherches ayant pour base des données existantes...

— ... ou ouvrir l'interprétation des résultats à des points de vue différents.

Plus récemment, Arzberger et coll. (2004) reviennent dans An international Framework to Promote Access to Data sur des expériences de mise en place de cadres institutionnels liés à l'accès aux données et à leur diffusion. Ils établissent, presque vingt ans après Fienberg et coll. (1985), une liste de sujets sur lesquels les institutions doivent travailler pour assurer le succès des projets de partage de grande envergure. Si certains problèmes cités par Fienberg et coll. (1985) restent, au début des années 2000, non seulement valides, mais surtout en partie irrésolus (notamment les problèmes liés à la documentation des jeux de données et au manque de motivation objective à partager), Arzberger et coll. (2004) ajoutent qu'il faut réaliser plus de recherches sur les bienfaits du partage aux niveaux

économique et social afin de faire prendre conscience de l'importance du problème aux décideurs[1].

Un argument apparaît souvent dans la littérature citée ici : d'un point de vue éthique, si une recherche a été financée par le contribuable, il serait normal que les fruits de celle-ci soient redistribués vers le public (Arzberger et coll., 2004 ; Parr et Cummings, 2005). Arzberger et coll. (2004) précisent :

> « Open access to publicly funded data provides greater returns from the public investment in research, generates wealth through downstream commercialization of outputs, and provides decision-makers with facts needed to address complex, often transnational, problems. »

À l'échelle de la communauté scientifique, un consensus semble se dégager : de nombreux auteurs insistent, dans un contexte où la spécialisation scientifique toujours grandissante fait face au caractère multidisciplinaire des solutions offertes par la science à la société, sur les opportunités que peuvent offrir le partage et la diffusion des données scientifiques. À l'échelle de l'individu, néanmoins, des frictions se font sentir et l'idée d'un partage quasi général fait souvent face à la réalité des intérêts individuels.

1.1.2. Le partage du point de vue de l'auteur des données

Il est difficile pour une institution de prendre la décision de mettre en place un partage général et systématique de la recherche qu'elle encadre. Comme le rappellent notamment Fienberg et coll. (1985), Stanley et Stanley (1988) et Arzberger et coll. (2004), il n'est pas rare que des chercheurs se trouvent dans une

---

[1] Ce document est le résumé d'un rapport commandé par l'Organisation pour la Coopération et le Développement Économique (OCDE), qui se soucie principalement de questions politico-économiques.

situation où la diffusion de leurs travaux doit être proscrite, car pouvant par exemple toucher à des problèmes de sécurité nationale, ou enfreindre des règles de confidentialité ou de propriété intellectuelle. Allant plus loin dans cette direction, Data Sharing (Stanley et Stanley, 1988) analyse le partage du point de vue de l'auteur original des données. Il y est fait mention des contraintes et des aspects négatifs que peut induire sur l'auteur original une diffusion institutionnalisée et obligatoire. Les auteurs s'intéressent ainsi aux droits du chercheur détenteur des données originales, ainsi qu'aux légitimes réticences qu'il peut éprouver à mettre systématiquement ses données à la disposition de ses collègues. Parmi celles-ci, Stanley et Stanley (1988) avancent plusieurs arguments qui se révèlent importants dans le développement de l'idée de partage parmi les scientifiques. Ils évoquent :

— le coût, en temps et en argent, nécessaire à la documentation formelle et à la mise à disposition des données ;

— le manque de gratification, voire le risque de conséquences négatives à accepter de partager leurs résultats dans le cas où ils seraient soumis à une vérification ;

— un possible effet pervers sur la science en général : la mise à disposition de travaux déjà accomplis pourrait inciter des chercheurs à utiliser en priorité des données déjà disponibles plutôt que de les collecter eux-mêmes. Ceci réduirait dans l'ensemble le nombre de jeux de données originaux concernant l'étude d'un phénomène particulier et par conséquent, appauvrirait la connaissance scientifique ;

— l'inquiétude de l'auteur de la recherche concernant la compétence du demandeur à analyser ses données ;

— la perte du contrôle sur son travail.

Campbell et coll. (2002) confirment quatorze ans plus tard dans Data Withholding in Academic Genetics: Evidence From a National Survey que parmi les chercheurs refusant de divulguer leurs données de recherches :

— 80 % justifiaient leur choix par la quantité de travail nécessaire à la mise à disposition de celles-ci ;

— 64 % par leur souhait de réserver leurs données aux futures publications de leurs étudiants ou collègues proches ;

— 53 % préféraient se réserver leurs données pour leurs propres publications futures.

La même étude nous apprend que 35 % des chercheurs estiment que le partage a diminué dans la décennie précédente alors que 14 % perçoivent le contraire.

La question du pouvoir et de son lien à la détention de l'information (donc à la détention d'une certaine connaissance) est à la base du dernier argument énoncé plus tôt par Stanley et Stanley (1988). Cette relation a été théorisée par Michel Foucault dans les années 1970 avec son concept de « savoir-pouvoir » (Foucault, 1980). Cette théorie, bien qu'établie à un niveau supérieur à ce qui nous intéresse ici (il traite principalement de l'utilisation de la connaissance par les institutions pour soumettre des groupes d'individus), apporte une fondation théorique aux observations de Stanley et Stanley (1988) : la détention de la connaissance, c'est-à-dire, ici, de données scientifiques, est liée à un certain pouvoir ou au moins au maintien du détenteur de ce savoir à un certain rang

hiérarchique qui lui confère (toutes proportions gardées toutefois) une position dominante sur ses collègues.

Enfin, dans leur analyse, Stanley et Stanley (1988) énumèrent les raisons les plus courantes que peut avoir un chercheur à réclamer un jeu de données à un collègue. Ils distinguent six raisons qui, de par le fait, recoupent et complètent la liste des bénéfices établie par Fienberg et coll. (1985) citée plus haut (l'intérêt du partage pour la communauté rejoint donc ici l'intérêt pour l'individu), parmi lesquels :

— faire des analyses que le premier chercheur n'avait pas faites ;

— établir un lien entre le jeu de données en question et un autre qui a été collecté indépendamment ;

— faire une méta-analyse à partir de données collectées indépendamment par plusieurs chercheurs.

Stanley et Stanley (1988) sont donc eux aussi conscients de l'importance de la diffusion de la connaissance dans une communauté : encore une fois, leurs objections se dirigent surtout vers le partage obligatoire et institutionnalisé.

1.1.3. Contraindre le système à s'équilibrer

> « There is a fundamental dilemma embedded in database creation and management. At least on a perceptual level, the benefits derived from a database are greater for the user of the data than for the contributor of the data. » (Porter et Callahan, 1994)

Pour Porter et Callahan (1994), un système de partage idéal rassemblerait des individus qui mettraient leurs travaux à la disposition de leurs pairs autant

qu'ils utiliseraient ceux de ces derniers. Mais quel est l'intérêt pour un chercheur de passer du temps à préparer des données, à les classer, les ordonner, les documenter formellement dans le seul but de les rendre utilisables par un tiers ? Nous touchons là à un point important du partage au sein d'un groupe : le don provient de l'individu alors que le gain provient du groupe en tant qu'entité. Cela peut constituer un manque de motivation et induire, pour les raisons évoquées par Stanley et Stanley (1988), un effet de rétention de données. C'est également un phénomène étudié dans le domaine des mathématiques (appliquées aux sciences sociales et à l'économie, en particulier) de la théorie des jeux.

Un cadre théorique

La situation dans laquelle se trouve une communauté de scientifiques où un agent (le chercheur) doit coopérer (ici, partager sa connaissance) pour permettre au groupe d'en tirer un gain, peut être modélisée grâce au cadre conceptuel de la théorie dite du « dilemme du prisonnier à agents multiples », ou MPD (Multi-Person Prisoner's Dilemma) (Schelling, 1978), qui est une extension du dilemme du prisonnier classique, un jeu à somme non nulle[2] que l'on peut résumer ainsi : deux agents, dont le but est de maximiser leur gain, doivent décider simultanément de coopérer ou non. Le dilemme réside dans le fait que d'un point de vue individuel, la stratégie de ne pas coopérer est la plus payante, alors que du point de vue du groupe, le gain est bien plus élevé si tous les agents coopèrent. Ranganathan et coll. (2004) décrivent dans Incentive Mechanisms for Large Collaborative Resource Sharing l'utilisation qu'ils font de la théorie du MPD pour

[2] C'est-à-dire que la somme des gains et des pertes ne s'équilibre pas forcément, en d'autres termes, deux joueurs peuvent se trouver, par exemple, dans des situations « gagnant/gagnant » ou « perdant/perdant ».

modéliser le comportement d'agents tels que des chercheurs mis dans une situation de partage. Dans un système composé de *n* agents, où chaque agent possède un jeu de données qu'il peut partager ou non, le bénéfice retiré de la participation au système de partage, qui est fonction du nombre de contributeurs, est identique selon que l'on décide de partager ses données (ce qui implique un coût) ou que l'on décide de ne pas le faire. Ainsi et comme l'illustre le dilemme du prisonnier dans sa forme classique de deux joueurs, sans motivation extérieure, le caractère rationnel des agents fait que l'on ne peut pas atteindre une situation de coopération/partage qui serait bénéficiaire à tous (l'équilibre du modèle se situant à zéro contribution).

Des travaux ont été conduits afin d'étudier et de développer des mécanismes visant à lutter contre la tendance à la rétention de données. Golle et coll. (2001) étudient ce problème dans le contexte des réseaux de partage de données P2P[3] et proposent des solutions visant à inciter les agents à participer activement au système, qui se rapprochent de celles mises en œuvre par Ranganathan et coll. (2004). Ces derniers, qui prennent pour exemple le cas de chercheurs amenés à partager (ou non) leurs travaux, étudient — par le biais de simulations basées sur un modèle MPD — deux mécanismes qui permettraient d'amener le système vers un état plus équilibré :

— échange de jetons, où l'utilisateur doit « payer » le contributeur d'un jeu de données. C'est, en économie, le système classique d'échange à tarif fixe ;

---

[3] « Peer to peer », méthode de partage de fichiers par Internet, où un agent demandeur d'un fichier particulier est mis automatiquement et directement en relation avec un autre agent possédant ce fichier. La limite à ce système résidant dans le manque objectif de motivation à mettre ses propres fichiers en partage.

— système de « réputation », où un agent peut seulement télécharger les données d'agents ayant une note « réputation » inférieure ou égale à la sienne. Cette note peut être, par exemple, fonction du nombre de jeux de données mis à disposition de la communauté.

Ranganathan et coll. (2004) confirment de manière expérimentale l'efficacité de ces méthodes simples pour lutter contre la tendance naturelle à une forme d'égoïsme des agents rationnels dans un système de ce type. Ces mécanismes sont toutefois difficilement acceptés dans un réseau d'échange. C'est particulièrement vrai dans le milieu de la recherche scientifique qui est lié à une tradition de « récompense » sous forme de citations (Olson et coll., 1996 ; Ranganathan et coll., 2004).

Les mécanismes pratiques

L'expérience des réseaux de chercheurs qui partagent déjà leurs travaux apporte un éclairage important sur les choix pratiques en matière de politique visant à motiver les auteurs originaux. Dans Packaging and Distributing Ecological Data from Multisite Studies, (Olson et coll., 1996) évoquent ce problème et proposent d'utiliser le système traditionnel de la citation. Ces citations, précisent-ils, devraient être établies dans un format similaire à ceux des publications de journaux et ajoutées aux descriptions des données. Enfin, en ce qui concerne les travaux issus de plusieurs jeux de données combinés, des règles doivent être établies afin de créditer les différents auteurs impliqués. Porter et Gallahan (1994) font état des pratiques imposées aux groupes de recherche du réseau Long Term Ecological Research (LTER). Dans ce réseau, vieux de vingt-cinq ans et au sein duquel travaillent environ 1800 personnes sur vingt-quatre stations dispersées

principalement sur le territoire étasunien, le partage des données de recherche est régi par une série de normes. Celles-ci, développées par le *LTER ad hoc Committee on Data Access*, obligent les fournisseurs de données autant que les utilisateurs. Porter et Gallahan (1994) résument ces obligations : si les gestionnaires de données pour chaque site de recherche sont responsables de la mise à disposition continue et sur le long terme de ces données (ainsi, encore une fois, que de leur documentation), les utilisateurs doivent citer les auteurs des données originales. Ils sont également responsables des coûts liés au transfert des données, et doivent s'interdire la revente de celles-ci à un tiers.

À un niveau supérieur, une autre motivation objective — qui semble la plus évidente au premier abord — est liée à la mise en place d'un cadre juridique favorisant la mise à disposition des données de recherche (Zimmerman, 2003 ; Arzberger et coll., 2004). Dans Data Sharing and Secondary Use of Scientific Data: Experiences of ecologists, Ann Zimmerman décrit les rapports qu'entretiennent les institutions fédérales étasuniennes avec la question du partage des données ; les États-Unis ont légiféré depuis quelques années dans le sens du partage quasi total des données produites avec des financements fédéraux, à l'exception des données confidentielles ou liées à la sécurité nationale. C'est le FOIA (Freedom of Information Act), complété de plusieurs révisions et amendements datant du début des années 2000, qui établit ce cadre juridique. Mais on ne retrouve pas cette ouverture dans la plupart des autres pays. Par exemple, bien que des méta bases de données scientifiques soient établies et financées par des instances européennes (comme le European database of long-term experiments on soil organic matter (Smith et coll., 2002)), c'est dans le sens de plus de protection des droits d'auteur que l'Union Européenne légifère en 1996 avec une directive sur la

protection juridique des bases de données. Concernant le Canada, qui ne dispose pas de législation fédérale à ce propos, Arzberger et coll. (2004) indiquent que trois des principaux organismes de financement suivent trois politiques différentes. La question du cadre juridique est donc encore largement en chantier et les débats se poursuivent.

1.1.4. Les aspects pratiques de la réutilisation des données

Dans la quête pour un partage généralisé des recherches scientifiques, convaincre des chercheurs de mettre leurs données à la disposition d'un groupe est une première étape. Lorsque les données issues de plusieurs individus, ou groupes d'individus, doivent être rassemblées, d'autres obstacles apparaissent. En citant, plus tôt dans ce mémoire, les freins au partage de l'information, nous avons évoqué les problèmes de compatibilités matérielle et logicielle. Nous n'allons pas y revenir ici, mais plutôt nous attarder sur la gestion des problèmes liés aux incertitudes qui entourent un jeu de données (lorsque celui-ci est réutilisé par un chercheur qui n'en est pas l'auteur) et à l'hétérogénéité qui résulte de l'assemblage de données issues de jeux différents. Nous allons ensuite voir deux solutions possibles : la standardisation des méthodes et des matériels, et la documentation des données par le biais de l'ontologie et des métadonnées.

L'hétérogénéité des données

Comme le rappellent Jones et coll. (2001), la collecte de données en écologie est fortement marquée par le caractère individuel du travail qui est effectué : la prise de données répond aux besoins immédiats d'une équipe sur un site, ce qui implique que les données, une fois rassemblées, peuvent présenter des dissimilitudes dans plusieurs de leurs caractéristiques. Concrètement, cette

hétérogénéité peut être induite, pour ne citer que quelques exemples, par la différence des méthodes d'échantillonnage employées, par l'expérience du ou des scientifiques qui ont effectué le travail, par le matériel qui était à la disposition du chercheur au moment de la collecte, mais aussi par la question de recherche pour laquelle le travail est effectué. Tout ceci rend le processus de partage ardu, car les données transférées de l'auteur au réutilisateur sont le plus souvent accompagnées d'incertitudes sur les buts de la recherche originale, sur les moyens et les procédures mises en œuvre pour les atteindre et — par conséquent — sur la capacité d'un jeu à être réutilisé pour une autre recherche.

La gestion de l'incertitude qui entoure un jeu de données

> « The use of data outside their original context implies distance. For data to be reused, they must be able to travel beyond the location in which they were produced » (Zimmerman, 2008)

Dans une étude empirique consistant à recueillir et analyser l'expérience d'écologistes utilisant des données de seconde main, Zimmerman (2003) étudie les obstacles objectifs à la pratique du partage de données du point de vue du réutilisateur. Pour bon nombre de chercheurs interrogés, les problèmes peuvent être grossièrement classés dans deux catégories. La première est celle du manque de confiance dans ces données. Nous savons qu'un certain degré d'incertitude accompagne tout jeu de données, même s'il n'en est pas fait mention dans le travail final (Star, 1985), et il est aisément compréhensible qu'un chercheur ne peut utiliser, ou incorporer à ses travaux, des informations dont il soupçonne que la qualité est insuffisante. La seconde catégorie concerne la compréhension de la recherche originale et des données qui en découlent : le scientifique doit absolument comprendre les données qu'il souhaite réutiliser. C'est-à-dire qu'il doit

savoir par qui elles ont été prises, quand, comment, avec quel matériel, dans quel but, dans quelles conditions, etc.

> « (...) If I honestly could not figure out what they have done, then I would not use that data point. » (Zimmerman, 2008)

Ces problèmes découlent de la séparation des données et de leur contexte et constituent une « distance » (Porter, 1995) qu'il convient de réduire. La standardisation de la collecte, du traitement, de l'entreposage et de la gestion des données fait partie des méthodes pouvant aider à la réutilisation de celles-ci en dehors de leur contexte. On peut citer l'exemple d'une réutilisation de données dans le cadre d'une méta-analyse, où l'usage de pratiques communes s'avère alors très irnportant : il est difficile de comparer des données collectées par plusieurs équipes qui utilisent des méthodes de collectes ou du matériel dont on sait qu'ils biaisent les résultats. Car si un biais constant peut ne pas avoir de répercussions importantes sur une seule étude, l'accumulation de biais différents peut s'avérer catastrophique lorsque l'on considère plusieurs jeux de données à la fois.

La standardisation telle que définie ici implique de gros efforts à plusieurs niveaux. Ces efforts pourront parfois être coûteux en temps et en argent, ce qui rend la standardisation difficile à mettre en place concrètement (Jones et coll., 2001 ; Zimmerman, 2008). Elle devra par exemple être le résultat de concertations au niveau de la communauté scientifique et elle touchera potentiellement toutes les étapes du travail sur le terrain. Ainsi, il faut que les écologistes s'entendent sur des méthodes qui sont à la fois suffisamment générales pour convenir à la majorité des situations et suffisamment précises pour être vraiment utiles à la compréhension et à l'estimation de la qualité des données.

Renommée et expérience

Ressortent des entretiens d'Ann Zimmerman plusieurs points importants qui avaient déjà été décrits par des spécialistes des métadonnées tels que Michener (1998) et Jones et coll. (2001). En premier lieu, les subtilités et la volatilité[4] de la collecte sont telles que l'expérience du chercheur qui souhaite réutiliser ces données intervient comme un élément déterminant dans le succès de cette opération. Un individu ayant déjà collecté une donnée particulière dans le passé aura une expertise qu'une personne moins expérimentée dans ce domaine n'aura pas. Il saura ainsi juger si l'utilisation d'un jeu est risquée ou non. En deuxième lieu, un point qui revient souvent dans les témoignages des chercheurs interrogés par Ann Zimmermann concerne la renommée de celui qui a publié les données. Quand des données sont réputées délicates à collecter et qu'un soupçon pèse sur un jeu, la renommée de l'auteur de la recherche originale peut devenir un paramètre décisif dans la réutilisation de celui-ci.

Dans un domaine tel que la recherche en écologie, dont le cadre d'étude est constitué de très nombreux phénomènes avec des interactions plus ou moins bien connues, où des artefacts de ces interactions peuvent parfois masquer les phénomènes étudiés et où l'expérience et parfois l'appréciation du technicien ou du chercheur sur le terrain entrent pour une grande part dans la qualité du travail final, la notion de « confiance » (en soi même ou dans la personne à l'origine du travail) fondée sur des appréciations subjectives demeure donc valide aujourd'hui. Toutefois, insiste Ann Zimmermann, si la confiance est importante, elle ne suffit pas à assurer le succès du transfert des données : la capacité à comprendre le jeu

[4] Certaines données pouvant grandement varier selon le contexte.

en question est la clef du processus. Les gros efforts de documentation objective demeurent donc de la plus haute importance.

La documentation des données

> « In short, metadata describes the 'who, what, when, where, and how' about every aspect of the data » (Michener, 2006)

L'échange de données implique un échange social qui s'effectue traditionnellement par la communication directe entre l'auteur des données et l'utilisateur qui les lui demande. L'utilisateur des données de seconde main doit savoir, par exemple, quand la collecte a été effectuée, dans quel but, avec quel matériel, etc. Dans un système où le partage est centralisé, l'auteur des données est souvent séparé du chercheur susceptible de les réutiliser par une base de données informatique. Cette communication entre l'auteur original et celui qui réutilisera le travail, qui est rompue par la nature même du système, peut constituer un frein à la diffusion de l'information dans le réseau.

Nous savons que ces problèmes, liés à la séparation des données de leur contexte, peuvent dans certains cas être atténués par des efforts de standardisation. Une autre façon de réduire la « distance » citée plus haut réside dans la documentation des données de recherche par le biais de schémas de métadonnées riches et précis sur lesquels des groupes de chercheurs s'entendent.

Comme nous l'avons vu dans la section précédente et ainsi que le rappellent Michener et coll. (1997) et Zimmerman (2008), un jeu de données brutes n'est jamais parfaitement compréhensible par lui-même. La documentation qui lui est jointe se révèle être le meilleur outil objectif permettant d'appréhender des données destinées à être classées ou réutilisées.

La recherche liée aux métadonnées a pris de l'ampleur à la fin des années 1980. À cette époque, de grandes bases de données environnementales faisaient leur apparition et il devint nécessaire d'explorer leur contenu selon la description des jeux qu'elles contenaient. C'est, pour William Michener, la première fonction des métadonnées : servir de support à la découverte de données. Une deuxième fonction est de permettre la compréhension et l'utilisation des données. Cette dernière est également à placer à un niveau supérieur de complexité, car le nombre et la précision des métadonnées nécessaires pour accomplir cette tâche sont bien plus élevés que pour la découverte de données. En plus des données descriptives générales indispensables à celle-ci, la compréhension requiert des informations précises sur, par exemple, le contexte de la recherche (incluant les hypothèses, les caractéristiques du site d'étude, les méthodes mises en œuvre, etc.), ou sur les unités ou la précision des mesures (Michener et coll. 1997 ; Michener, 2006 ; Madin et coll., 2007). La troisième et dernière grande fonction des métadonnées est de permettre l'automatisation informatique de la découverte de données, de leur assimilation et de leur analyse. Ainsi, la standardisation de ces schémas de métadonnées est essentielle à la bonne communication au sein d'une communauté.

William Michener est à l'origine d'une des premières tentatives de description de données écologiques : en 1987, il établit avec ses collègues une première série de trente paramètres destinée à décrire le contexte de la recherche. En 1997, soixante-quatre métadonnées sont établies et supportées par l'*Ecological Society of America's Committee on the Future of Long-term Ecological Data*. Ces métadonnées devaient notamment permettre de faire en sorte que les données soient toujours exploitables après la cessation d'activité de leurs auteurs (Michener

et coll.,1997). Depuis, les communautés liées à différents domaines des sciences se sont attelées à l'établissement de standards de schémas de métadonnées de plus en plus complets, avec par exemple, pour l'écologie, l'*Ecological Metadata Language*. Ils permettent un début d'automatisation de la découverte, de la gestion et du traitement de données brutes.

Vient s'ajouter à cela l'ontologie, qui permet de capturer les observations de manière plus souple et plus subtile que ce qu'offre un système de métadonnées en établissant des relations sémantiques entre les données. L'ontologie permet de mettre en place une hiérarchie de classes et d'instances de classes, dont les liens sont fonction du sens qui les lie (Madin et coll., 2007) :

> « As a simple example, if instances of biomass are defined as instances of weight in a particular domain ontology, then data about biomass will be discovered when searching for data about weight (...). Moreover, these data are compatible through being semantically classified as weights, and can potentially be merged. » (Madin et coll., 2007)

L'ontologie propose ainsi une solution aux limites de la documentation par les métadonnées seules. Elle apporte, par le biais d'une classification sémantique, une souplesse dans la recherche des données et une meilleure description du contexte de la recherche.

Limites

Comme la standardisation des méthodes de collecte des données, la mise en place de pratiques de description de jeux de données est accompagnée de coûts : lorsqu'une équipe a terminé les travaux de terrain et les analyses, quand les résultats de ces recherches sont parus et que le ou les chercheurs sont prêts à passer à un autre projet, il ne reste souvent plus de fonds ni de temps pour la

documentation des données (Michener et coll., 1997). Dans un réseau qui souhaite rassembler des jeux de données collectés par plusieurs chercheurs, l'intervention humaine dans la documentation et la gestion des données est souvent indispensable : une personne doit alors jouer le rôle de bibliothécaire de données numériques, sa tâche est notamment de convertir, de classer et de documenter les jeux de données des chercheurs (Westbrooks, 2004). Cette méthode de gestion est plus efficace que lorsque les chercheurs documentent entièrement par eux-mêmes leurs jeux de données, mais elle est aussi plus coûteuse.

Il faut également garder à l'esprit que même avec une base de données bien administrée, seul l'auteur d'un jeu de données connaît les conditions exactes de la collecte, ses subtilités ou les anomalies éventuelles dans les résultats ; une bonne communication est donc, une fois de plus, indispensable à la transrnission efficace de l'information. Ainsi, une fois la grande importance de la formalisation de métadonnées établie, il convient de rappeler, comme le montre l'étude de Zimmerman (2008), qu'au delà de la documentation standardisée des données telle que décrite dans la section précédente, la communication verbale ou écrite entre chercheurs reste, même lorsqu'elle est ponctuelle, de la plus grande importance :

> « There is no unique, minimal, and sufficient set of metadata for any given data set, since sufficiency depends on the use(s) to which the data are put. » (Michener et coll., 1997)

1.2. Les progrès dans la représentation géographique de données scientifiques

> « Oui, les savants maîtrisent le monde, mais seulement si le monde vient à eux sous forme d'inscriptions en deux dimensions, superposables et combinables. C'est toujours la même histoire depuis Thales au pied des pyramides. » (Latour, 1999)

Nous avons vu que les avancées de l'informatique personnelle ont posé, dès les années 1980, les fondations permettant l'émergence d'un nouveau paradigme de gestion et de partage des données produites par la science. De son côté, le domaine de la visualisation spatiale des données scientifiques a également profité des progrès rapides et constants dans les capacités de calcul des ordinateurs personnels. L'informatique a permis, avec les SIG, des améliorations notables en terme d'accès, d'exploitation et de visualisation des données scientifiques. Ils permettent notamment la création et la superposition rapide de couches de données géoréférencées et offrent aux scientifiques ainsi qu'aux décideurs des outils d'analyse et d'aide à la décision très efficaces. Toutefois, nous connaissons depuis le début des années 2000 une conjonction d'évolutions techniques qui place la communauté scientifique dans une situation inédite. D'une part, le niveau atteint dans la qualité de l'affichage sur moniteur et d'autre part l'explosion de la quantité d'informations (notamment spatiales) disponible ont permis de mettre en place de nouveaux outils cartographiques qui autorisent l'exploration et la visualisation spatiale de données scientifiques.

Nous allons ici laisser de côté la visualisation de données sur support papier et sur support informatisé « traditionnel » (que l'on connaît avec les SIG), pour nous concentrer sur les nouvelles méthodes de représentation cartographique que

nous permettent les progrès récents de l'informatique. Nous aborderons également dans cette section la recherche effectuée dans les domaines de la visualisation spatio-temporelle de données scientifiques, et de la cartographie plus abstraite des réseaux sociaux.

1.2.1. Le globe virtuel : un nouvel outil de visualisation spatiale

> « It's like the effect of the personal computer in the 1970s, where previously there was quite an élite population of computer users. Just as the PC democratized computing, so systems like Google Earth will democratize GIS » (Goodchild, 2008)

> « (...) virtual globes are set to go beyond representing the world, and start changing it. » (Butler, 2006)

Du touriste cherchant son chemin, au scientifique tentant de discerner des les patrons spatiaux du phénomène qu'il étudie, les cartes constituent pour tous un outil permettant la compréhension de l'environnement géographique dans lequel nous évoluons. Alors que nous utilisons des cartes projetées selon diverses méthodes depuis l'Antiquité, le support de celles-ci et, par extension, la manière que l'on a d'appréhender l'objet cartographique, n'a pas fondamentalement changé depuis des siècles. Nous voyons néanmoins depuis le début des années 2000 une évolution intéressante dans le domaine de la cartographie. Depuis 2001, année durant laquelle des techniciens de la société Keyhole produisent un globe virtuel qui deviendra en 2005 le très populaire Google Earth (GE), plusieurs globes virtuels sont apparus. Parmi ces applications, outre GE, NASA World Wind (WW)[5]

---

[5] WW est une application distribuée gratuitement depuis 2004 et dont le code source est « libre », c'est-à-dire qu'il est lisible et modifiable par tous.

et Virtual Earth[6] de Microsoft sont les plus populaires. Comme le terme le laisse penser, un globe virtuel est semblable à bien des égards aux représentations tangibles du globe terrestre tels qu'ils existent depuis la fin du XVe siècle. C'est donc, grossièrement énoncé, un modèle en trois dimensions de l'ellipsoïde terrestre sur lequel est plaquée une collection d'images géoréférencées issues de satellites de télédétection.

En comparaison de la cartographie informatique, les globes virtuels apportent de nombreuses de nouveautés intéressantes. Parmi celles-ci, l'interactivité : l'utilisateur peut interagir avec le modèle par le biais de la souris de l'ordinateur en contrôlant la position de son point de vue par rapport à la représentation du globe. À la manière d'une caméra virtuelle qui filmerait le globe (l'écran de l'ordinateur serait alors le cadre filmé par cette caméra). Il est ainsi possible de s'approcher ou de se reculer du globe, de se déplacer dans toutes les directions, d'incliner son point de vue, etc. Une autre particularité importante des globes virtuels est qu'ils possèdent plusieurs jeux d'images de définition plus ou moins fine : alors que la caméra s'approche du globe, des images plus précises se substituent aux précédentes plus grossières (Tooth, 2006). Aussi, depuis quelques années, les irrégularités verticales de la surface de la planète ont été ajoutées au modèle : en inclinant la vue, l'utilisateur voit apparaître le relief, modélisé en polygones, de la zone observée.

L'objet cartographique et la compréhension du contexte spatial

L'arrivée de l'informatique est relativement très récente dans l'histoire de la cartographie et les usages et les habitudes de travail liés aux cartes sur support

---

[6] Comme GE et WW, Virtual Earth est distribué gratuitement. Toutefois, son code source est confidentiel.

papier ont été presque directement transférés au nouveau support informatique, sans changer radicalement notre perception de l'objet cartographique. Pour la première fois depuis l'informatisation des cartes, toutefois, l'arrivée des globes virtuels pourrait induire un changement dans les rapports qu'un individu entretient avec la représentation et la compréhension de son espace géographique. Il y a plusieurs raisons à cela, parmi lesquelles trois points se démarquent :

— le mode d'utilisation permet pour la première fois une exploration fluide, simple et intuitive dans toutes les dimensions de l'espace géographique (Goodchild, 2008) ;

— il est possible d'adapter les visualisations à ses besoins. Cette personnalisation se fait par l'ajout de contenus aux fonds de cartes proposés par défaut. Nous reviendrons sur ce point dans la section suivante ;

— le succès de ces logiciels est à présent clair : à titre d'exemple, 350 millions de personnes ont téléchargé GE depuis le mois de juin 2005 (Ohazama, 2008). Nous sommes donc en présence d'un outil dont la base d'utilisateurs est très importante (Schöning et coll., 2007).

Dans cette masse d'utilisateurs, au moins deux grands types semblent se dégager : les scientifiques d'un côté et le grand public de l'autre.

Les globes virtuels et le grand public

Pour le grand public, il semblerait que ce genre d'outil soit pour le moment utilisé afin de répondre à des questions simples, telles qu'« où se trouve l'hôpital le plus proche », ou pour identifier un lieu déjà connu (par exemple sa propre résidence) sur l'imagerie satellitaire (Butler, 2006 ; Schöning et coll., 2007). Ainsi,

pour l'utilisateur moyen, une fois la période d'exploration qui suit la découverte de l'outil passée, les tâches accomplies sont le plus souvent ponctuelles et ne sont que rarement issues d'une véritable réflexion spatiale :

> « Too often the gift of spatial context provided by these technologies is under-used by applications that only provide functionality and ignore the users' inclination to explore and learn » (Schöning et coll., 2007)

Schöning et coll. (2007) avancent dans Improving interaction with Virtual Globes through Spatial Thinking: Helping Users Ask 'Why?' que ce problème est lié à deux caractéristiques du système. D'une part, l'interface physique qui nous permet d'interagir avec le logiciel : nous utilisons pour le moment la souris de l'ordinateur, mais celle-ci est en passe de se voir préférer des méthodes plus immersives telles que des écrans tactiles. D'autre part, l'information qui est délivrée en surimpression de l'imagerie satellitaire : la communication de l'ordinateur vers l'utilisateur d'informations contextuelles sur sa recherche est pour le moment trop faible. Les auteurs montrent dans leur étude qu'au prix d'efforts dans ces deux domaines, on peut espérer que les globes virtuels pourront libérer leur potentiel pédagogique et induire chez les utilisateurs une compréhension contextuelle de leurs recherches. Cela leur permettrait d'aborder des questions de la vie quotidienne de manière spatiale ou, pour reprendre l'expression de Schöning et coll. (2007), d'ajouter le « *pourquoi ?* » à la question « *où* se trouve *quoi* ? ». Nous n'en sommes certes pas encore là pour le moment, mais les progrès effectués dans ce sens montrent que les globes virtuels ont le potentiel de changer le rapport du non-cartographe à l'objet cartographique.

Les globes virtuels et les scientifiques

Les scientifiques sont souvent, de par leurs travaux (et l'utilisation qu'ils font parfois des SIG), plus au fait de l'outil cartographique que le grand public. Selon les disciplines dans lesquelles ils exercent, ils sont également habitués à considérer les phénomènes qu'ils étudient dans de multiples dimensions, dont les dimensions spatiales. Les globes virtuels offrent de ce fait des voies nouvelles à l'utilisation de la cartographie en science. Tooth (2006) cite plusieurs arguments montrant l'importance de ces outils comme support à la réflexion géographique : en tant que géomorphologue, l'auteur conçoit la combinaison de la visualisation des caractéristiques topographiques d'une région et des possibilités d'examiner la scène à différentes échelles comme une opportunité nouvelle d'établir, depuis son bureau, des hypothèses quant aux processus liés à la formation du paysage. De la même manière, un travail consistant à rassembler des images aériennes pour observer la variation de patrons à grande échelle demandait de longues journées de recherches, l'utilisation de l'imagerie d'un globe virtuel permet d'effectuer ces recherches beaucoup plus rapidement. Pour Butler (2006), ces nouveaux outils cartographiques peuvent être un point d'entrée vers les SIG ; beaucoup de scientifiques dont les recherches pourraient tirer des bénéfices des SIG se tiennent éloignés de ceux-ci, car leur emploi est encore largement réservé à des spécialistes.

Goodchild (2008) explique les raisons de cette simplicité d'accès par le fait que les globes virtuels évitent le plus possible l'emploi de références techniques habituellement utilisées dans le métier. L'utilisateur n'a pas besoin de se soucier du système de projection, de la déformation qui en résulte, du géoréférencement des images, des fausses couleurs habituellement utilisées sur ces images en

télédétection, etc. De plus, le concept d'échelle est partiellement éludé puisque celle-ci change dynamiquement avec le point de vue de l'utilisateur ; un mesurage entre deux points se fait ainsi automatiquement le long de la géodésique, dans l'unité de la mesure sur le terrain, sans poser plus de questions à l'utilisateur. Mais si les fonds de cartes interactifs que constituent les globes virtuels ont un attrait en eux-mêmes, l'intérêt pour les scientifiques repose toutefois largement sur la capacité à représenter leurs propres données et à leur donner rapidement et simplement un sens dans leur contexte spatial.

1.2.2. La création et la visualisation de contenu rapporté

Butler (2006) et Goodchild (2008) rappellent très justement que les globes virtuels que nous avons aujourd'hui ne permettent pas d'effectuer les analyses auxquelles les SIG nous ont habitués[7]. Selon eux, le véritable intérêt des globes virtuels, outre la capacité à manipuler le globe à travers les échelles et selon des axes que nous interdit (sauf longs efforts) la cartographie informatique traditionnelle, réside dans leur capacité à visualiser, dans les trois dimensions spatiales, des données de toutes sortes. Pour bon nombre de scientifiques, l'intérêt d'utiliser ces outils devient frappant : ils n'ont désormais plus besoin de mettre en place de gros et coûteux SIG pour visualiser des échantillons géoréférencés. L'emploi de méthodes rapides d'importations de données autorise une visualisation simple et expressive. Il permet également le partage de ces données avec des collègues.

La plupart des globes virtuels utilisent, comme dans les SIG traditionnels, un système de visualisation par couches. La couche de base demeure généralement

---

[7] L'ampleur de l'intérêt chez les professionnels du SIG — tels que ESRI — pour ce nouvel objet cartographique laisse toutefois espérer une intégration progressive de certains outils d'analyses.

l'imagerie aérienne fournie par défaut. Sur celle-ci, un grand nombre d'autres couches — dont le niveau de transparence est ajustable — peut être ajouté. Ces couches de données géoréférencées peuvent être de types vectoriel (polygones, polylignes, points) ou matriciel (Goodchild, 2008). L'intérêt de ces plates-formes comme support à la visualisation de données réside donc principalement dans deux points : l'ajout et le partage de contenu.

L'intégration de nouveau contenu

Il est rapide et relativement aisé (pour certaines tâches, le degré de facilité varie néanmoins selon les logiciels) d'intégrer de nouvelles données directement sur le globe. Cette intégration peut se faire par différentes méthodes :

— par numérisation manuelle (point par point) ;

— par l'intégration directe de données prises en temps réel grâce à un système de positionnement par satellites ;

— par la conversion de couches vectorielles créées dans d'autres SIG au format *Shape*[8]. Il existe d'ailleurs déjà des outils développés par des programmeurs indépendants permettant d'exporter, directement depuis l'application ArcGIS, le travail en cours au format KML, afin d'être inclus dans GE notamment (Goodchild, 2008).

Pour un scientifique, ces données intégrables aux globes virtuels peuvent être de types et d'origines diverses : dans The Web-wide world, Butler (2006) cite l'exemple d'un biologiste qui suit les déplacements de morses au Groenland en plaçant des récepteurs GPS (*Global Positioning System*) sur des animaux et en

[8] Format de fichier de la société ESRI, éditrice de la suite logicielle ArcMAP dont ArcGIS fait partie.

affichant leurs positions directement dans GE. Il cite également les travaux d'un climatologue qui utilise GE pour visualiser, de manière dynamique et selon des points de vue divers, des données qu'il collecte par avions et satellites. Aussi, tout comme nous pouvons afficher des données collectées sur le terrain, il est possible de représenter des données modélisées issues de recherches diverses.

Pour ce genre d'utilisation, il est important de considérer les avantages et inconvénients des deux principales applications : GE et WW. Cette dernière étant, d'une manière générale et pour le moment, plus adaptée à l'affichage de données dynamiques que le produit de Google, qui — pour sa part — propose à l'heure actuelle plus de souplesse dans l'affichage de données statiques.

1.2.3. Les globes virtuels comme plates-formes de partage du contenu

La gratuité, le grand nombre d'utilisateurs et la souplesse d'utilisation de ces outils les rendent attractifs comme plate-formes de partage de données géoréférencées. Le partage, toutefois, demeure un domaine où les méthodes mises en place par les différents globes virtuels divergent. À titre d'exemple, GE exploite le format de fichier KML (fondé sur une structure XML, donc non compilé et modifiable par l'utilisateur) qui permet l'exportation et l'irnportation de couches de données. Ce format a d'ailleurs été nouvellement accepté comme standard par l'*Open Geospatial Consortium*. La société Google catalogue également dans une gigantesque base de données les fichiers KML (parfois appelés *mashups*) créés par d'autres utilisateurs. Ils laissent, pour le meilleur et pour le pire, le soin à la communauté d'utilisateurs d'évaluer la qualité du contenu par un système de notation. WW, de son côté, utilisait jusqu'à récemment un format fondé lui aussi sur le schéma XML, mais il converge actuellement vers une plus grande intégration du

format particulier KML, tout en acceptant déjà l'importation du format *Shape* d'ESRI.

De plus en plus, cette facilité de partage d'informations séduit les scientifiques qui voient dans ces outils, non seulement une solution simple pour visualiser leurs propres données, mais également une excellente plate-forme de partage de données avec leurs collègues (afin, par exemple, de mettre en commun des jeux de données) ou avec les décideurs et le grand public (afin de vulgariser et de communiquer des recherches pouvant influencer la prise de décision (Butler, 2006 ; Goodchild, 2008)).

Cette capacité à produire des visualisations spatiales précises, et à les mettre facilement à disposition d'un grand nombre d'utilisateurs[9] a été récemment mise à profit dans le domaine de la gestion de catastrophes naturelles. Ce type de situation nécessite un partage rapide et efficace d'informations géoréférencées de grande qualité (typiquement de l'imagerie aérienne de définition submétrique) et en quasi temps réel afin d'organiser les secours. Comme le rapportent Nourbakhsh et coll. (2006), GE a joué un rôle important dans les jours qui ont suivi la catastrophe de l'ouragan Katrina à la Nouvelle-Orléans aux États-Unis : une collaboration entre les organismes de secours et le *Global Connection Project* s'est mise en place afin de fournir des informations géographiques à jour et de grande qualité aux équipes de sauvetage. Celles-ci ont pu ainsi coordonner leurs efforts en visualisant ces images dans GE. Elles ont pu déterminer, par exemple, les zones dont l'accès par la route était rendu impossible.

Cette expérience a été renouvelée lors du séisme au Pakistan en 2005 : Google ayant rapidement rendu publique une série d'images satellitaire de haute

---

[9] Potentiellement au moins plusieurs millions dans le cas de GE et de son format KML.

définition prise le lendemain de la catastrophe. Ce genre d'imagerie, notent Nourbakhsh et coll. (2006), n'est habituellement pas accessible au public, mais la gravité des événements a imposé des efforts de la part des détenteurs de l'information. La souplesse et la facilité d'usage des globes virtuels ont permis de la rendre disponible et directement exploitable par n'importe qui possédant un ordinateur et une connexion à Internet.

1.2.4. Limites

Si les globes virtuels apportent une interactivité et une continuité dans l'exploration cartographique, ils enlèvent également une partie de la structure narrative à laquelle les atlas et les cartes sur papier nous ont habitués. La manipulation interactive du globe et la transition d'une région à une autre étant fluide, dynamique et continue, la compréhension de la géographie est plus intuitive et plus directe qu'avec une série de cartes statiques découpant le monde en rectangles représentés à des échelles différentes, telles que l'on peut en trouver dans un atlas. Ces derniers, toutefois, conduisent grâce à l'intégration de données factuelles relatives à chaque carte, ainsi qu'à un cadrage réfléchi des différentes zones géographiques représentées, à une compréhension et à une analyse que les globes virtuels n'offrent pas encore.

Dans le même ordre d'idée, alors que la carte papier simplifie et symbolise, par la sémiologie graphique, ce qui est représenté, l'approche choisie pour les globes virtuels est celle d'un pseudo-réalisme dont l'objectif est de reproduire au plus près l'aspect visuel de la Terre[10]. Par exemple, si le relief est représenté sur les cartes papier par des symboles (habituellement des courbes de niveau), les

[10] Cet objectif se heurte toutefois à des limites discutées plus loin dans cette section.

globes virtuels invitent l'utilisateur à observer la Terre de profil pour voir se dessiner le relief de la zone observée. Aussi, les cartes papier représentent les lieux d'intérêt selon une symbologie précise et décrite dans une légende. Les données affichées sont ainsi sélectionnées, triées et hiérarchisées par le cartographe dans le but de transmettre les informations le plus clairement possible. Les globes virtuels, de leur côté, se contentent d'afficher une photographie des lieux. À charge de l'utilisateur de l'interpréter. Sur ce point, les globes virtuels tendent toutefois à se rapprocher de la cartographie traditionnelle en permettant le téléchargement et l'affichage, en superposition de l'imagerie nue, de couches de données supplémentaires. Ainsi, alors que les globes virtuels offrent une compréhension intuitive de la géographie, les cartes papier synthétisent -à l'heure actuelle- mieux l'information.

L'utilisation du globe virtuel comme d'un outil cartographique soulève également des problèmes liés à la précision et à l'homogénéité des données affichées. Concernant la précision, le géoréférencement des éléments matriciels est parfois, notamment dans Google Earth, en désaccord avec d'autres informations superposées. Par exemple, les tracés des réseaux routiers montrent parfois des décalages de plusieurs dizaines de mètres avec le tracé figurant sur les images, ce qui témoigne de manques de précisions dont les origines sont difficiles à retracer. Si cette faiblesse peut ne pas être dérangeante pour certaines utilisations (par exemple pour des visualisations à petite échelle), cela pourrait vite limiter l'intérêt de cet outil pour des applications requérant une meilleure précision. Concernant l'homogénéité, la mosaïque des images aériennes est construite à partir de clichés produits à différentes dates, ce qui peut poser des problèmes de compréhension. Par exemple :

— il y a une rupture dans la continuité visuelle du globe une fois les images placées côte à côte. Ceci induit des situations où une zone potentiellement intéressante peut être divisée en plusieurs images la représentant sous différents états ;

— la date des images est très importante dans la compréhension de ce qui est représenté. Les saisons peuvent induire sous nos latitudes de grandes différences visuelles. Malheureusement, ces informations de dates sont absentes des visualisations (pour Google Earth, seules les dates liées aux droits d'auteurs sont stipulées, les informations sur les images elles-mêmes sont difficiles, voire impossibles, à obtenir).

D'une manière générale, les choix des images affichées, qui ont une grande importance dans la compréhension des cartes, ne sont pas clairement exposés aux utilisateurs. Quels sont les critères principaux au choix d'une image par rapport à une autre ? La définition de l'image ? La densité du couvert nuageux ? Le contrat avec un fournisseur particulier ? etc.

Pour un emploi plus professionnel, la simplicité d'utilisation des globes virtuels s'accompagne aujourd'hui d'une limitation dans ce qu'il est possible d'accomplir avec le produit. Si cette limite tend à s'amoindrir avec l'ajout constant de nouvelles fonctions, le scientifique doit toutefois, au moment de l'évaluation de ses besoins, prendre cette limitation en compte avant de préférer un globe virtuel à un autre mode de représentation géographique.

Concernant l'utilisation de globes virtuels comme support au partage et à la visualisation de données, il n'existe pas encore de système simple et directement utilisable pour échanger et afficher celle-ci directement dans l'application. Les échanges se font donc encore largement par l'envoi de fichiers (de type KML) d'un chercheur à un autre.

Au delà des considérations techniques, certaines de ces applications sont produites par des entreprises dont les politiques commerciales peuvent avoir une importance. Google offre une API (*Application Programming Interface*) performante et en constante évolution pour Google Maps et, depuis quelques mois, pour GE (quoique dans une version *bêta* encore très jeune). Celle-ci permet d'intégrer les systèmes de cartographie Google au sein même du navigateur Internet. La puissance commerciale de Google, ainsi que son exposition dans les médias, en fait une marque vers laquelle de nombreuses personnes choisissent de se tourner spontanément, créant une masse critique d'utilisateurs qui, *de facto*, finit par imposer le choix des technologies de la compagnie. Encore une fois, au moment du choix qu'un chercheur doit faire entre une technologie et une autre, les avantages et les inconvénients doivent être soigneusement considérés. Par exemple, le code source des applications de Google n'est pas libre, toute l'activité est gérée dans les locaux de Google et un utilisateur est susceptible de passer par les serveurs de la société pour afficher la moindre donnée. Ceci implique des problèmes à plusieurs niveaux qui peuvent dans certains cas limiter la notion de gratuité de ce type d'outil :

— confidentialité : les données affichées grâce aux services de cartographie Google peuvent être directement lisibles et exploitables par la société. Cela devient un problème important quand il s'agit, par exemple, de données de

recherches. Comme nous l'avons vu précédemment, un des arguments pouvant freiner le partage des données est la réticence que certains chercheurs éprouvent à perdre le contrôle de celles-ci ;

— dépendance à la politique commerciale de Google : le système mis en place à l'aide des outils de Google devient dépendant du bon vouloir de la compagnie à continuer de fournir le service gratuitement tout au long de la durée de vie du projet que le scientifique met en place ;

— exploitation des données à des fins commerciales : la stratégie commerciale de Google (et la raison d'exister de services innovateurs et coûteux en termes de recherche et de développement) repose sur l'exploitation des informations que les utilisateurs consentent à lui laisser afin, par exemple, de soumettre à ces derniers des publicités commerciales précisément ciblées. Des questions éthiques peuvent légitimement émerger de ce genre de pratiques dans le cadre de la recherche scientifique.

Pour finir, si la base d'utilisateurs des globes virtuels est importante (en nombre d'utilisateurs), celle-ci reste géographiquement et socialement hétérogène : l'utilisation de ces logiciels est directement liée aux infrastructures permettant l'accès à Internet, à l'équipement en ordinateurs et à l'éducation de l'utilisateur potentiel. Ces trois points sont très liés au niveau socioculturel dans lequel l'utilisateur évolue (Bucy, 2000). Le mythe d'une société où l'information (géographique ou non) voyagerait librement de son auteur au reste du monde et où chacun aurait le pouvoir de produire cette information demeure encore aujourd'hui de l'ordre de l'utopie. Ainsi, la « démocratisation de l'information

géographique » projetée par Goodchild (2008) ne semble pas encore à portée de main.

1.3. Visualisation multidimensionnelle dans un espace à trois dimensions

> « Why is our space three-dimensional? The physicist Paul Ehrenfest found that planetary orbits in N-dimensional Euclidean space R^N are stable if and only if N=3, precluding other dimensional universes from having a long career » (Inselberg, 1990)

Dans de nombreux domaines de la recherche scientifique et principalement dans les recherches environnementales, la collecte de données s'effectue par sites. Pour chacun de ces sites, de nombreuses données sont généralement échantillonnées à des dates diverses. Ainsi, la nature même du travail à effectuer place souvent la personne chargée de l'analyse devant une collection de données multidimensionnelles parfois très complexes dans leurs interconnections. La représentation de ce type de données sur un support cartographique traditionnel (à deux dimensions), de manière à ce que des tendances puissent être rapidement observables, constitue un défi sur lequel se sont penchés de nombreux chercheurs, principalement depuis le début des années 1990 (Kreuseler, 2000).

Parallèlement, nous avons vu qu'une des grandes contributions des globes virtuels dans le domaine de la cartographie concerne l'apport de la troisième dimension spatiale, qui demandait auparavant des efforts particuliers et des logiciels onéreux. Pour l'affichage de données multidimensionnelles, le passage de deux à trois dimensions représente un progrès important, mais souvent insuffisant compte tenu du nombre d'informations à afficher. Dans le cas le plus classique de représentation de données géoréférencées dans un espace géographique à trois

dimensions, deux de celles-ci sont employées pour représenter la position géographique du site d'étude, et la troisième est disponible pour l'affichage d'une troisième valeur.

Étant contraintes aux trois dimensions de notre espace, de nombreuses méthodes ont été développées afin d'y représenter un nombre supérieur de dimensions. Certaines des plus marquantes sont présentées dans l'article de Ferreira de Oliveira et Levkowitz (2003), From visual data exploration to visual data mining: a survey. Dans Visualization of Geographically Related Multidimensional Data in Virtual 3D scenes, Kreuseler (2000) s'emploie à transférer certains de ces concepts, créés pour des représentations sur des supports à deux dimensions spatiales, telles que *parallel coordinate* (Inselberg, 1990), *dimensional staking* (LeBlanc et coll., 1990), ou *shape coding* (Beddow et coll., 1990), au contexte d'un espace géométrique à trois dimensions. Il applique également certaines techniques de visualisation par le biais d'icônes ou de codes de couleurs.

1.3.1. Plusieurs dimensions dans une icône

Chernoff (1973) présente dans un article intitulé The Use of Faces to Represent Points in k-Dimensional Space Graphically une méthode de visualisation multidimensionnelle originale permettant à l'observateur de comprendre rapidement des relations potentiellement complexes entre diverses variables. La méthode proposée consiste à dessiner un visage dont les caractéristiques (forme du visage, position et taille de la bouche, du nez, des oreilles, etc.) correspondent à des données échantillonnées et à autant de dimensions (dix-huit au total). Cette méthode est particulièrement intéressante, car elle permet d'augmenter facilement le nombre de dimensions affichées en gardant

un bon niveau de lisibilité (voir figure 1). C'est, dans l'esprit, ce que reprennent plus tard Bergeron et Grinstein (1989) : l'utilisation d'icônes, dont certains paramètres sont contrôlés par des variables, permet dans leur cas de visualiser sept dimensions simultanément : la position utilisant deux dimensions (les auteurs ne travaillant pas spécifiquement sur support géographique), les cinq autres — correspondant à divers attributs de l'icône — étant réservées aux autres variables échantillonnées. En théorie, précisent Bergeron et Grinstein (1989), l'attribut couleur à lui seul contient trois dimensions (la teinte de la couleur, sa saturation et sa valeur), toutefois, la capacité du cerveau humain à identifier précisément ces trois dimensions à la fois n'est pas évidente. Il convient donc de garder constamment à l'esprit qu'une des limites au nombre de dimensions affichables simultanément réside dans la capacité de l'utilisateur moyen de les discerner. Par exemple, la méthode de Chernoff (1973) a pour limite que les différents attributs d'un visage humain sont peu comparables entre eux (il est difficile de comparer un attribut lié à la taille du menton avec celui qui est lié à celle des yeux, par exemple), ainsi que le fait qu'un observateur humain se concentre inconsciemment sur certains aspects d'un visage plus que sur d'autres (Ferreira de Oliveira et Levkowitz, 2003).

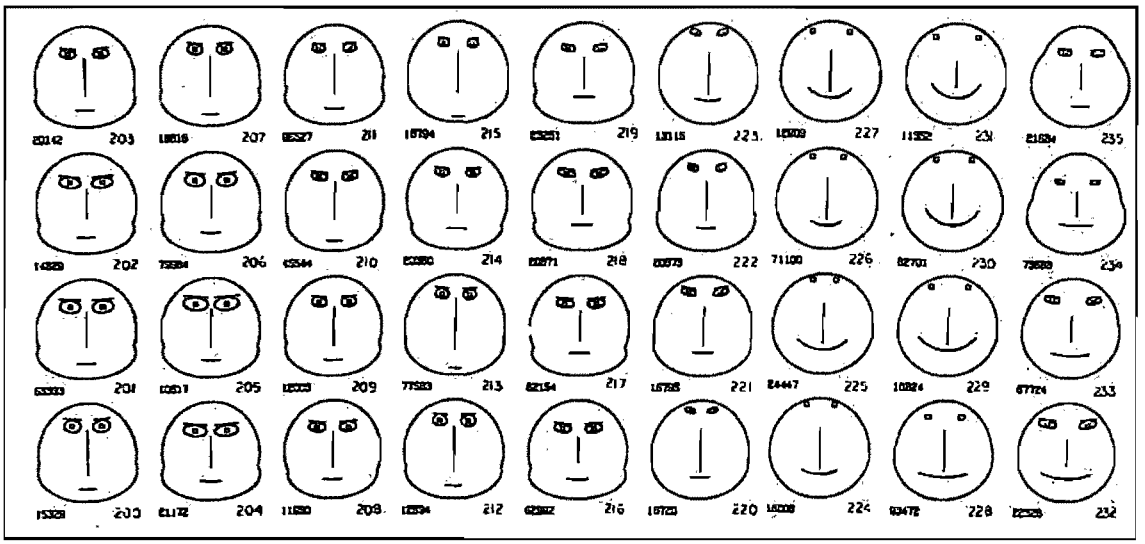

Figure 1 : Exemple de visualisation multidimensionnelle d'échantillons de données géologiques selon la méthode de H. Chernoff (1973).

Dans la même lignée que précédemment, bien que destinée à des jeux de données de nature un peu différente, Pickett et Grinstein (1988) proposent une méthode impliquant des icônes sous forme de bâtons (voir figure 2). Elle consiste à attribuer des propriétés d'inclinaison, d'orientation, de taille, d'épaisseur et de couleur à des segments qui seront placés sur un repère à deux dimensions (le nombre de dimensions affichables simultanément est alors de sept). Ce genre de visualisation s'applique bien à des situations où les points sont échantillonnés à intervalles proches et réguliers ; l'ensemble des segments formant, une fois les propriétés appliquées, une texture à partir de laquelle il est possible de déduire des tendances ou des patrons.

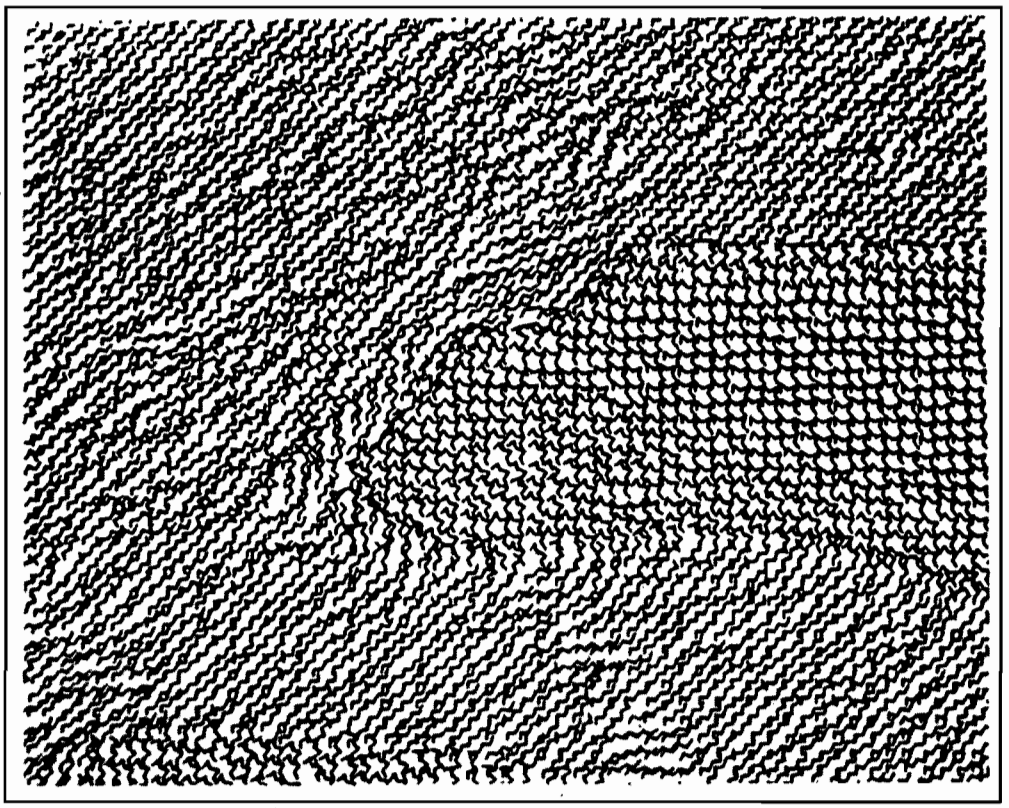

Figure 2 : Exemple de visualisation multidimensionnelle par bâtons (Pickett et Grinstein 1988).

1.3.2. Visualisation spatio-temporelle

> « Pictures as well as words can be used to tell stories, and must somehow incorporate time because stories are of changes and without time would not exist. Specifically, we consider how maps present geographic information narratively - how they help us to tell geographic stories. » (Vasiliev, 1997)

L'analyse et la compréhension de phénomènes écologiques sont souvent étroitement liées aux dimensions spatiales et temporelles. Ainsi, une base de données échantillonnée sur le terrain, ou produite à partir de modèles contient souvent des indications à caractères à la fois spatial et temporel. La dimension

spatiale est représentée par les coordonnées géographiques correspondant, par exemple, au lieu d'échantillonnage, et la dimension temporelle par une ou plusieurs dates, ou par une durée. Lorsqu'il s'agit de visualiser ces données, le support cartographique permet effectivement la représentation des dimensions spatiales, mais la question de la visualisation simultanée de la dimension temporelle reste entière.

Bien que l'implémentation du temps dans les représentations géographiques s'avère souvent problématique (il n'existe pas vraiment de méthode s'appliquant à tous les cas de figure), de nombreuses méthodes permettant de montrer l'évolution de phénomènes à la fois de manière spatiale et temporelle ont été mises au point depuis plusieurs décennies. Ces recherches, qui se sont grandement accélérées avec l'arrivée des SIG, peuvent se diviser en trois grandes catégories : les cartes représentant l'évolution d'une ou plusieurs variables par le biais d'animations, les cartes statiques qui incluent la dimension temporelle (grâce à la mise en œuvre d'une symbologie adaptée) et les cartes exploitant la troisième dimension spatiale (les deux premières étant le plus souvent réservées au support cartographique projeté).

Les cartes animées

Comme l'explique Vasiliev (1997), l'idée d'afficher successivement les différents états d'un phénomène sur un fond de carte a été théorisée dans les années 1950 par Norman Thrower dans Animated Cartography, mais c'est dans les années 1990, avec les progrès de l'affichage informatique, qu'il a été possible de mettre en place des systèmes de visualisation de cartes animées interactives dans lesquels l'utilisateur peut choisir, par exemple, les variables à observer (Monmonier, 1990). L'amélioration constante des capacités d'affichage des

ordinateurs personnels a permis des progrès spectaculaires qui sont notamment mis en œuvre aujourd'hui dans les globes virtuels. Celui-ci sert de support à l'animation de l'iconographie qui représente les valeurs d'une ou plusieurs variables. Une frise chronologique, affichée en surimpression de la cartographie, indique la date de début, la date de fin et la position sur cette échelle de l'image affichée à un temps *t*.

Les cartes statiques

Il s'agit ici de représenter l'évolution dans le temps d'un phénomène par le biais d'une iconographie particulière. Un mouvement sur la carte peut par exemple être représenté par une flèche, ou un événement ponctuel indiqué par une étiquette contenant une date. Un exemple connu de carte statique incluant des informations temporelles est celui de la Carte figurative des pertes successives en hommes de l'armée française dans la campagne de Russie 1812-1813 par M. Minard. La carte représente à la fois le déplacement de l'armée (par une ligne montrant le chemin que parcoururent les troupes) et les effectifs de celle-ci qui s'amenuisent au fur et à mesure de leur progression vers Moscou et de leur retraite vers la France (par l'épaisseur de la ligne). La carte est accompagnée d'une courbe, liée à la représentation du déplacement par des lignes verticales, montrant la température à différentes étapes de la retraite de l'armée française. De manière simple et claire, Minard réussit véritablement avec cette carte à raconter visuellement une histoire. Compte tenu des capacités des ordinateurs personnels, les cartes animées sont aujourd'hui plus utilisées que les cartes statiques. Ces dernières demeurent toutefois utiles, car elles ne sont pas dépendantes de l'informatique et peuvent être imprimées sans que le sens ou l'information contenu par la carte n'en soit altéré.

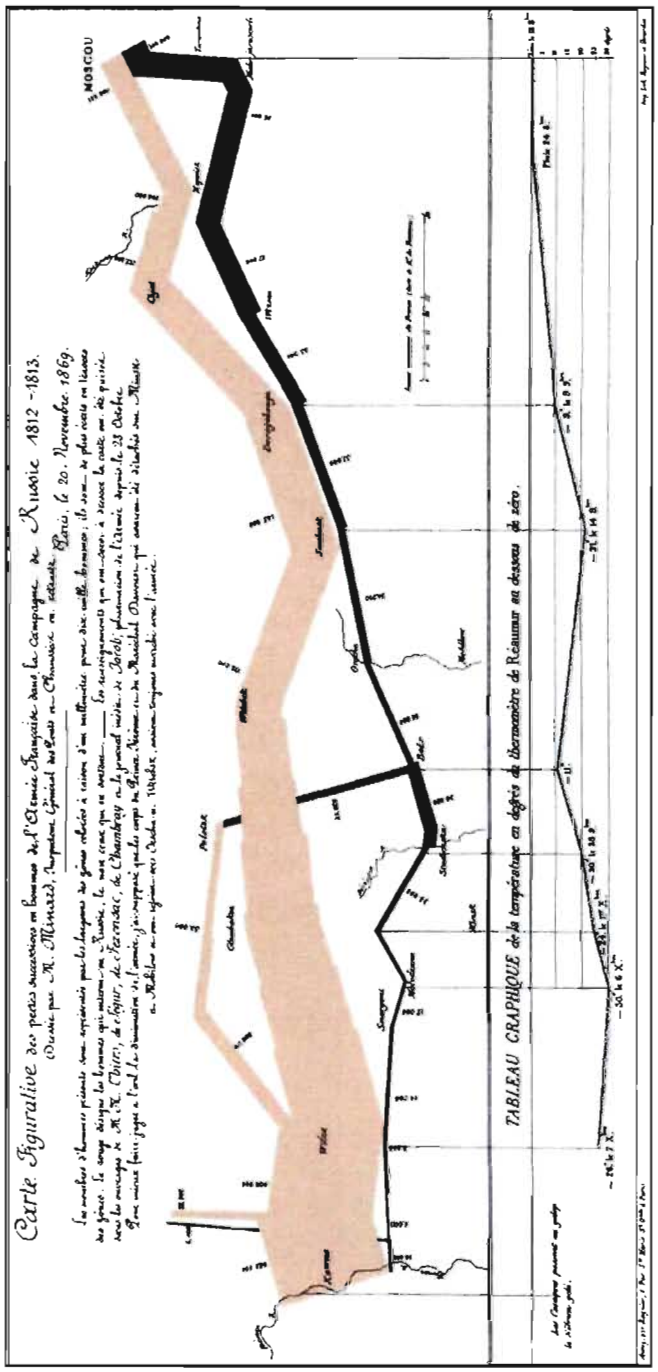

Figure 3 : Carte figurative des pertes successives en hommes de l'armée française dans la campagne de Russie 1812-1813 par M. Minard (1869).

Exploiter la troisième dimension spatiale

L'emploi de la troisième dimension spatiale comme espace de visualisation temporelle a été formalisé dans les années soixante par Torsten Hägerstrand, qui a développé le concept de cartographie temporelle. Il s'agit de lier les informations spatiales et temporelles par la représentation de « chemins spatio-temporels ». Ceux-ci utilisent la troisième dimension spatiale, c'est-à-dire l'altitude, pour représenter la date d'un événement géoréférencé. Le « cube spatio-temporel » (Hägerstrand et coll., 1967 ; Hägerstrand, 1970), repris par de nombreux chercheurs (encore une fois, lorsque l'informatique permit de concrétiser ces idées) dont Kwan (2000) et Kraak (2000), est une illustration de ce concept (figure 4). Le cube spatio-temporel consiste à représenter des données dans une portion de territoire inscrite dans la base d'un cube ou d'un rectangle (dans ce cas, on parle d'« aquarium spatio-temporel »). Les évolutions spatiale et temporelle des variables sont dessinées sur la même carte (qui est préférablement représentée en vue de 3/4) ; les dimensions spatiales sont toujours représentées horizontalement, alors que la dimension temporelle est représentée verticalement. L'utilisateur est alors capable de comprendre visuellement la date et le lieu d'un événement. Plus un point est placé à une altitude élevée, plus l'événement lui correspondant s'est déroulé récemment sur l'échelle temporelle. À l'inverse, un point proche du sol est lié à un événement plus ancien (Andrienko et coll., 2003)

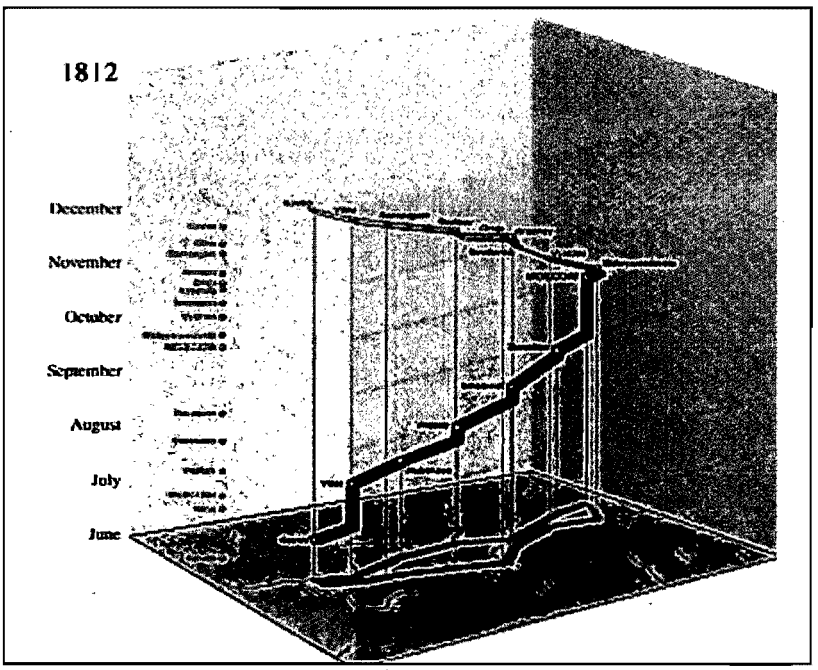

Figure 4 : Représentation de la carte de M. Minard de 1869 selon la méthode du cube spatio-temporel (ITC-Minard).

1.4. La cartographie d'un réseau

> « Jameson saw the correlation of abstract knowledge and imaginary figures as key to understanding contemporary symbolic structures, and regaining the capacity to act within them. » (Abrams et Hall, 2006)

Nous nous éloignons maintenant de la cartographie géographique pour nous intéresser à une autre cartographie, plus abstraite et faisant appel à des codes et des conventions nouvelles, celle des réseaux.

1.4.1. Le contexte théorique

L'organisation de la vie est telle que des interactions, des liens, des interférences, émergent lorsque des sujets évoluent dans le même environnement.

Cela se retrouve par exemple dans les écosystèmes où les interactions et les interdépendances entre plantes, mammifères, ovipares, etc., forment des réseaux (trophiques, par exemple) qui sont le sujet de recherches actives.

Pour les humains, les réseaux sont, aux niveaux politique, social, culturel, économique, militaire, académique, etc., des composantes dominantes de nos sociétés (Jameson, 1991). Ainsi, chacun évolue dans plusieurs réseaux d'interdépendances. Pour un individu quelconque, un de ces réseaux pourrait être, par exemple, celui de sa famille, un autre celui de son travail, un troisième encore celui de ses amis, etc. Dans tous ces réseaux, qui ont alors au moins un point en commun (notre individu), des hiérarchies émergent. Dans le réseau « travail », s'il s'agit d'une grosse entreprise, toute une organisation d'interdépendances s'est établie : les membres du réseau des actionnaires de l'entreprise, par exemple, sont probablement aussi connectés à d'autres entreprises, créant de ce fait un autre réseau chapeautant toute une série d'agrégats d'individus par le simple jeu des connexions. On peut ainsi entrevoir que, de proche en proche, notre individu est un élément d'une structure décentralisée gigantesque, abstraite et particulièrement difficile à appréhender.

Les théories liées aux réseaux recouvrent un large spectre d'intérêts et sont par conséquent étudiées dans différents domaines des sciences (notamment dans les sciences sociales et les mathématiques). Chacun dans son domaine, ces milieux académiques ont apporté des contributions fondamentales à la compréhension de ces réseaux.

En mathématiques/systèmes complexes, une des avancées les plus marquantes est celle de l'effet dit « petit monde » (Milgram, 1967 ; Watts, 1999), qui pose et explique le paradoxe que dans de nombreux réseaux, la moyenne du

nombre de liens par lesquels il faut passer pour rejoindre deux individus pris au hasard est étonnamment bas, indiquant que les interconnections sont complexes, denses et structurées de manière particulière. On parle souvent de six degrés de séparation, mais ce nombre, parfois utilisé comme un titre accrocheur, varie selon les réseaux et les types de liens étudiés (Watts, 2003 ; Barabàsi, 2003).

Dans les sciences sociales, les philosophes du post-modernisme ont développé des théories liées à la place des individus dans la société. La théorie des réseaux est un outil puissant pour aider à la compréhension des sphères d'influences imbriquées qui agissent sur une personne ou sur un groupe, ainsi que des transferts divers dont ils sont soit les bénéficiaires, soit les instigateurs, soit les intermédiaires.

À ce propos, Schwartz et Wood (1993) écrivaient dans Visualizing Network Data :

> « We are currently in the midst of a networking revolution. Data communications networks such as the Internet now connect millions of computers, cellular phones have become commonplace, and personnal communications networks are in the developmental stages. In parallel with the ever increasing network sizes has been a concomitant increase in the collection of network measurement data. Understanding this data is of crucial importance as we move to a modern, information-rich society. »

C'est en effet dans le contexte de la mondialisation, de la déréglementation des échanges commerciaux et de l'explosion de la complexité des réseaux d'échanges (d'information, de devises, de biens, etc.) qui en résulte, que Frederic Jameson décrit dès le début des années 1980 le besoin d'une « esthétique de la cartographie cognitive » pour combler notre incapacité à représenter mentalement les grands réseaux décentralisés dans lesquels nous sommes intégrés en tant

qu'individus (Jameson, 1991). Il insiste sur l'importance politique et sociale de visualiser les sphères d'influences qui agissent sur chacun à différents degrés, afin de comprendre sa position en tant qu'individu, ou groupe d'individus, au sein de la société.

Représenter concrètement un réseau

Cette cartographie cognitive des réseaux est aujourd'hui communément employée pour représenter des interactions diverses au sein de groupes ; le moyen d'y parvenir est de dessiner un graphe, ou sociogramme (déjà décrit par Jacob L. Moreno dans les années 1930), où l'on note les nœuds ainsi que les lignes (directionnelles ou non) qui les lient. Le résultat est une constellation de nœuds interconnectés (voir figure 5). En plaçant correctement les différents nœuds selon leurs connexions, il est possible d'avoir une vue d'ensemble des interactions qui se produisent dans le réseau étudié. On peut alors facilement voir, même dans des réseaux extrêmement compliqués comme ceux que forment — par exemple — les liens trophiques entre les espèces d'un écosystème, quels nœuds sont fortement interconnectés, quels nœuds servent de ponts entre différents groupes, où se forment des cliques, des groupes, sur quels autres groupes ils agissent, etc.

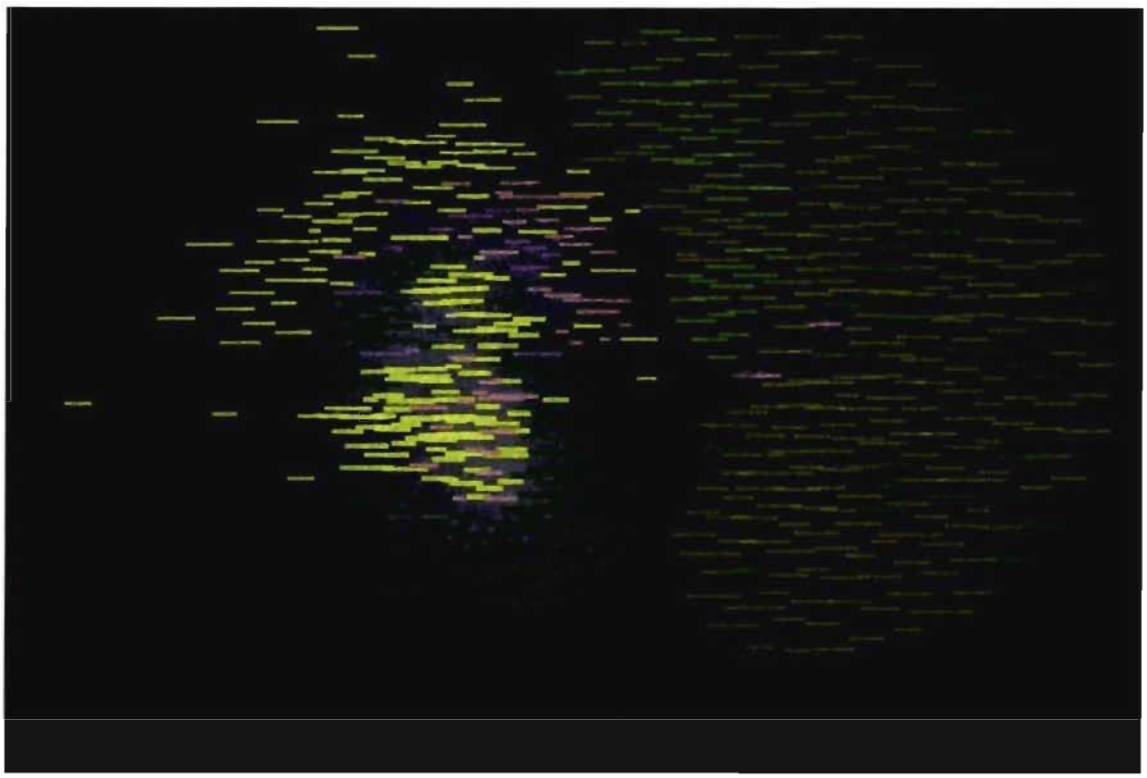

Figure 5 : Graphe représentant les relations trophiques entre différentes espèces présentes sur le mont Saint-Hilaire, Québec. Chaque espèce est représentée par un nœud de couleur différente et les liens de prédation sont représentés par des flèches grises ayant pour origine le prédateur et pour destination la proie.

Des informations supplémentaires peuvent également être visualisées sur le même graphe ; on peut représenter, par exemple, une variable relative à un nœud du réseau (ou à un lien entre deux nœuds) par une couleur, on peut également faire varier la taille de ces attributs visuels pour visualiser une autre variable. Toutes sortes de données qualitatives ou quantitatives peuvent alors être représentées facilement.

Schwartz et Wood (1993) rappellent également que si un réseau représente de manière abstraite un ensemble d'interconnexions, celui-ci peut contenir des

informations spatiales : les nœuds pouvant ainsi être ancrés à des positions géographiques et représenter, par exemple, des échanges à travers différentes régions du monde.

1.4.2. Le scientifique et son réseau social

Bien qu'à l'échelle d'un groupe de scientifiques, l'intérêt des membres soit plus local que ce qui a motivé les recherches en sociologie citées plus haut, les principes généraux restent valides. Qu'il s'agisse de collaborations directes entre des chercheurs, ou de l'utilisation d'un travail déjà effectué comme point de départ d'une nouvelle idée, la recherche repose en grande partie sur les interactions des chercheurs avec leur environnement professionnel. La connaissance et la compréhension qu'ont les chercheurs de cet environnement ont donc une grande importance, tant pour la carrière des individus que pour les avancées scientifiques qui peuvent découler de leurs collaborations.

Comme nous l'avons vu, la cartographie seule, sous forme de graphe (et sans analyse particulière), d'un réseau permet de visualiser simplement et rapidement des interconnexions compliquées et peut aider à comprendre la structure générale de celui-ci. Il est également possible de regarder plus près dans cet enchevêtrement de liens et de nœuds pour tirer des informations à l'échelle de l'individu. On peut ainsi se servir du graphe comme d'un outil d'aide à la communication dans un réseau au sein duquel les membres, potentiellement dispersés géographiquement, ne se connaissent pas forcément. Selon le type de liens affichés, on peut alors savoir quelles relations existent entre des chercheurs et répondre à des questions telles que, pour n'en citer que quelques-unes : qui travaille sur quoi ? Qui travaille avec qui ? À quel groupe/université appartient une

personne ? Quels chercheurs s'intéressent à un sujet particulier et quelles pourraient être les collaborations possibles ? À qui s'adresser pour obtenir un type spécifique de données ? Etc.

2. Problématique et objectifs

2.1. Rappel du contexte

La littérature présentée dans le premier chapitre traite de deux grands domaines : le partage des données parmi les chercheurs d'un côté, et la visualisation spatio-temporelle de l'information scientifique de l'autre. Ces domaines sont issus d'efforts de recherche séparés, le travail présenté ici vise à les lier afin de créer un outil de partage plus visuel, plus intuitif et plus efficace.

La première partie de la mise en contexte établit l'importance du partage des données scientifiques. L'Humanité est de plus en plus confrontée à des changements environnementaux se manifestant à différentes échelles et dont les causes, souvent complexes, sont mal connues. Pour appréhender ces questions, les scientifiques tendent à opter pour des approches holistiques où les problèmes sont considérés de manière multidisciplinaire à différentes échelles spatiales et temporelles. Il est ainsi souvent nécessaire de recueillir de grandes quantités de données différentes, échantillonnées sur de vastes étendues spatiales et sur de longues durées. Une équipe ou un laboratoire seul n'a souvent ni le temps ni les moyens financiers de recueillir ces données. L'alternative consistant à utiliser une partie du travail d'autres chercheurs s'avère donc particulièrement intéressante. Dans ce contexte, les communautés scientifiques ont besoin d'outils permettant de mettre simplement et efficacement en œuvre ce partage.

La deuxième grande partie montre les progrès accomplis dans le domaine de la visualisation, dans des espaces spatio-temporels, de données scientifiques. Les méthodes présentées, souvent antérieures à l'informatique, se sont révélées

et ont pris l'ampleur que l'on connaît aujourd'hui grâce à celle-ci. Il est ainsi aujourd'hui possible, notamment grâce à la cartographie 3D, de mettre en œuvre des méthodes efficaces de visualisations multidimensionnelles sur des supports géographiques intuitifs et interactifs.

Les données dans leur contexte spatio-temporel : la cartographie pour faciliter l'exploration de données partagées

En sciences environnementales, les dimensions spatiales et temporelles sont importantes pour appréhender nombre de phénomènes. La mise en contexte des données partagées peut ainsi être facilitée par l'intégration d'outils permettant leur visualisation sur un support spatio-temporel. La dimension spatiale est déjà (ou sera bientôt) représentée dans des systèmes de partage de données telles que OBIS-SEAMAP (Halpin et coll. 2006), Ecotrends (Servilla et coll. 2008) ou GBIF-MAPA (Flemons et coll. 2007). Toutefois, ces outils proposent des modules de visualisation à deux dimensions spatiales (sous forme de cartes projetées) qui s'avèrent vite limités quant au nombre de dimensions affichables simultanément (par exemple, la dimension temporelle n'est pas représentée dans les applications citées). D'autres systèmes commencent à proposer des visualisations 3D reposant sur Google Earth. Ils nécessitent néanmoins des transferts manuels de fichiers entre le module de recherche et le module de visualisation. Ces transferts, potentiellement répétitifs et fastidieux, tendent à freiner la découverte de données. Toutefois, des avancées logicielles récentes autorisent l'intégration de méthodes efficaces et performantes de visualisations multidimensionnelles aux interfaces de partage.

2.2. Le projet

Nous sommes liés dans ce projet au Réseau de Gestion Durable des Forêts (RGDF[11]), un réseau de recherche pan-canadien des Centres d'Excellence CRSNG qui a réuni depuis 1995 environ quatre cents membres issus de divers secteurs d'activité[12]. Dans ce réseau comme dans de nombreux autres, la recherche de chaque membre est communiquée par plusieurs moyens, notamment :

— par le biais des publications dans les revues spécialisées, ou par les rapports (rapports d'avancement et rapports finaux) que le RGDF publie ;

— par le biais des conférences organisées annuellement par le réseau.

Dans chacun de ces cas, si l'objet et les résultats des recherches des membres du réseau sont publiés et relativement bien connus, les données elles-mêmes ne sont pas distribuées systématiquement.

Création d'un nouvel outil de partage

L'objectif de ce travail est de créer pour le RGDF une application permettant aux membres de ce réseau géographiquement dispersé d'explorer, de visualiser et de partager des données collectivement produites depuis presque quinze ans. Cette recherche a donc mené à la mise au point d'un outil de partage propice au travail de groupes géographiquement dispersés, où les données de chacun sont mises à la disposition de la communauté dans son ensemble.

---

[11] Ou SFMN (Sustainable Forest Management Network), en anglais.

[12] Si la majorité des membres est issue du milieu académique, les secteurs gouvernementaux, industriels, non gouvernementaux et aborigènes sont représentés.

L'outil de partage se divise en deux parties. Le serveur d'un côté, qui contient la base de données construite à partir des soumissions des membres du réseau. L'application client de l'autre. Cette dernière est bâtie à partir de plusieurs modules différents, tous issus de logiciels libres.

Parmi ces modules, nous avons :

— la partie cartographique, qui est construite à partir de la bibliothèque de programmation NASA World Wind Java (worldwind.arc.nasa.gov). Contrairement à Google Earth qui contient une batterie de méthodes de visualisations standardisées, NASA World Wind Java constitue une base cartographique 3D presque nue (seuls le modèle 3D et les images satellitaires sont fournis par défaut) à partir de laquelle il est possible d'expérimenter et de développer des méthodes de visualisation nouvelles ;

— la partie permettant l'exploration de la base de données, qui exploite la bibliothèque Java *prefuse* (Heer 2004, Heer et coll. 2005) à travers deux modules : un menu en arborescence d'un côté (où les données sont mises à disposition de l'utilisateur à travers une hiérarchie de domaines et de sous-domaines) et un graphe de l'autre. Ce dernier représente les interconnexions des membres du réseau et permet de sélectionner des données à visualiser selon leur contexte social.

Ces modules sont rassemblés dans un programme Java (créé pour cette recherche) qu'il est possible de décliner facilement en application Web afin d'être exécutable dans un navigateur Internet. Une description plus complète de l'ensemble des modules de l'application est présentée dans le chapitre 3.

3. Article

An interactive cartography and ecoinformatics tool to facilitate the exchange and visualization of Canada-wide forestry database

Facilitating Data Exchange Through Interactive Cartography

Rodolphe Gonzales[1], Jeffrey A. Cardille[1*], Lael Parrott[1], Caroline Gaudreau[2], Gaël Deest[3].

[1] Département de Géographie, Université de Montréal, Montréal, QC, Canada.

[2] École de Technologie Supérieure, Montréal, QC, Canada.

[3] Département de Mathématiques, Université d'Angers, France.

* Corresponding author: Jeffrey A. Cardille.

Submitted to Ecological Informatics

Keywords : Spatiotemporal visualization, Interactive visual exploration, Virtual globes, Geobrowser, Data sharing, Geographic Information Systems, Data visualization, Web portal.

3.1. Abstract

Recent dramatic advances in computer networks and information technologies have created exciting new possibilities for sharing and analyzing scientific research data. Although individual datasets can be studied efficiently using remarkable hardware and software applications, many scientists are still largely limited to considering data collected by themselves, their students, or closely affiliated research groups. Increasingly widespread high-speed network connections and the existence of large, coordinated research programs suggest the potential for scientists to access and learn from data from outside their immediate research circle. If viewable over a common, simple framework, the information contained in such assembled databases could provide greater perspective on an individual's research results. Meanwhile, the arrival of new and user-friendly spatial visualization tools, such as World Wind or Google Earth, has quickly delivered revolutionary new ways to interact with geographic information. The scale and speed of their adoption for various uses is striking: more than 100 million users downloaded Google Earth in its first year, and many millions of user-created maps have been made using Google Maps technology. We are developing a web–based application that facilitates the sharing of scientific data within a research network using the now-common «virtual globe» in combination with advanced visualization methods designed for geographically distributed scientific data. Two major components of the system enable the rapid assessment of geographically distributed scientific data: a database built from the information submitted by the members, and a module featuring novel and sophisticated geographic data visualization techniques. Users can submit datasets to the system's database, explore and visualize its content through a spatiotemporal

interface taking advantage of the virtual globe, and download data for their own analysis. By enabling scientists to share results with each other, the system provides a new platform for important meta-analyses and the analysis of broad-scale patterns. Here we present the most recent developments in the creation of the SFMN GeoSearch platform for the Sustainable Forest Management Network, a pan-Canadian network of forest researchers who have accumulated data for more than a decade. Through the development and dissemination of this new tool, we hope to help scientists, students, and the general public to understand the depth and breadth of scientific data across potentially large areas.

3.2. Introduction

Independently accumulated datasets are for the most part bound to personal computers in laboratories, while an aggregation of these individual efforts could be used to address the complex ecological questions with which humanity is confronted. In ecology particularly, scientists are often working on problems that need to be tackled from a multidisciplinary perspective with data covering wide spatial and/or temporal spaces (Michener 1997, Andelman et al. 2004, Madin et al. 2007). Although collaborations between specialized researchers from different disciplines are frequent, once a study is completed by a research team, results are typically communicated through the usual peer-reviewed articles, publications and conferences, and the data that led to the published results are seldom released outside of the laboratories from where they emerged. The existence of these disconnected individual research data creates a situation where bits of information about a given subject are geographically and/or socially scattered within the community, and therefore not exploited to their full potential.

In this context, research over the last fifteen years related to data management, meta-databases, metadata and ontology, has been pushing towards freeing the data's potential by making them available beyond the scope of their original projects. Meta-databases, built from spatially and temporally scattered datasets, are confronted with many problems related to both the heterogeneity and the formal description of their content (Michener 2006). Ontology and metadata is, in ecology, an active field of research that is leading to more efficient ways to gather, document, organize, and redistribute data coming from heterogeneous studies for second-hand uses. Tools to set up data sharing frameworks like Metacat (Jones 2001), or ecology-oriented online systems based on aggregated databases, like OBIS-SEAMAP (Halpin et al. 2006), Ecotrends (Servilla et al. 2008) and GBIF-MAPA (Flemons et al. 2007), are successfully contributing to pushing scientific communities towards a new paradigm of broader data sharing (Guralnick et al. 2007).

Meanwhile, in the last five years, the emergence of new mapping methods, aimed at cartographers and non-cartographers alike, allow for novel ways to display large amounts of data on a geographical medium. These methods take advantage of the spread (at least in the richest parts of the world) of increasingly fast Internet connections, as well as of the progress made in consumer grade computer graphics. Commonly called «virtual globes», or «geobrowsers», they propose interactive three-dimensional models of the earth on which all sort of content can be downloaded from remote servers and shown in quasi real-time. Among others, applications like Google Earth (earth.google.com) or NASA World Wind (worldwind.arc.nasa.gov) are popularizing this new cartography. They are providing to a broad audience of academics, stakeholders and general public a

kind of Geographic Information Systems (GIS) that is, although –at least in term of analysis capabilities- not as sophisticated as their professional counterparts, easier to use, highly interactive, and capable of inducing a good understanding of ecological data through their geographic context (Tooth 2006, Foresman 2008, Goodchild 2008).

At the crossroad of data sharing and new cartographic media

With the growing interest in data sharing, we have seen in the last few years different initiatives taking advantage of new computer cartography as a means to visualize georeferenced shared scientific datasets. Within this new cartography, two kinds can be distinguished: two-dimensional, projected maps, and three-dimensional models of the earth. Projects like OBIS-SEAMAP (http:// seamap.env.duke.edu/) and GBIF-MAPA (http://gbifmapa.austmus.gov.au/) are good examples of the common integration of cartography with ecological databases: all within a website, users make queries and the resulting datasets are displayed on a two-dimensional map of the Earth. More recently, three-dimensional virtual globes are increasingly used as a geographic visualization medium. Most of the projects using them are built around Google's Keyhole Markup Language (KML) format, an increasingly popular XML-based file format originally designed around Google Earth. It describes the geographical positions of points, lines, shapes or images in order to build and exchange preset data visualizations on virtual globes. Users generally download a KML file pointing to one or several remote datasets to display them in Google Earth or in other virtual globes. In turn, users can create a KML out of their own data and submit it to a remote server. Systems using an external virtual globe to display queries made in another application facilitate knowledge sharing by presenting different datasets in their

spatial context. But unlike data sharing tools using two-dimensional cartography, they still require users to put a set of different tools to work in order to go through the process of searching and displaying data. The lack of continuity in the process of exploring data and visualizing them could be, for many scientists, an obstacle to the use of geographically explicit shared data exploration systems.

We present here the latest progress made on SFMN GeoSearch, a data sharing and visualization application built for the Sustainable Forest Management Network (SFMN). The SFMN is a Canada-wide research group connecting individuals from heterogeneous sectors of activity and promoting multidisciplinary collaborations in research related to the sustainable management of forests in Canada. It has been a member of the Network of Centres of Excellence (NCE) since 1995, and has gathered over time more than 450 people active in a broad range of fields.

The focus of our research is to provide ecology-oriented research networks with a generic, user-friendly, and integrated tool that takes advantage of the latest in three-dimensional mapping technologies. Instead of harnessing separate programs to search and to display information, we built an integrated system where the tools needed for database exploration, data visualization, and communication between datasets authors and potential second-hand users are tightly interconnected, easy to use and accessible from the same online application. In this article, we will present some aspects of SFMN GeoSearch, focusing on its modules related to the spatiotemporal visualization and the exploration of shared datasets.

3.3. Structure of the application

The SFMN GeoSearch system provides a set of heterogeneous modules and techniques, each playing a distinct role in accomplishing various tasks. These tools, once integrated in a single interface allow members to: (1) Submit their own datasets to the system's centralized database; (2) Explore in intuitive ways the sum of all data already submitted; (3) Visualize chosen data on a multidimensional geographic medium; (4) Discuss, annotate and comment on the submitted datasets (figure 6). When appropriate, users can download datasets for their own use (including meta-analysis, addition to their own datasets to increase spatial or temporal scales, etc.). In term of its architecture, our system can be divided in two main parts: a user interface on one side, and a user-populated database on the other. The interface was made from various open source Java libraries and encapsulated in a Java applet, making it compatible with most popular web browsers. The database was built on Postgre SQL; it parses, organizes and stores georeferenced variables from the datasets in order to support our spatiotemporal visualization system.

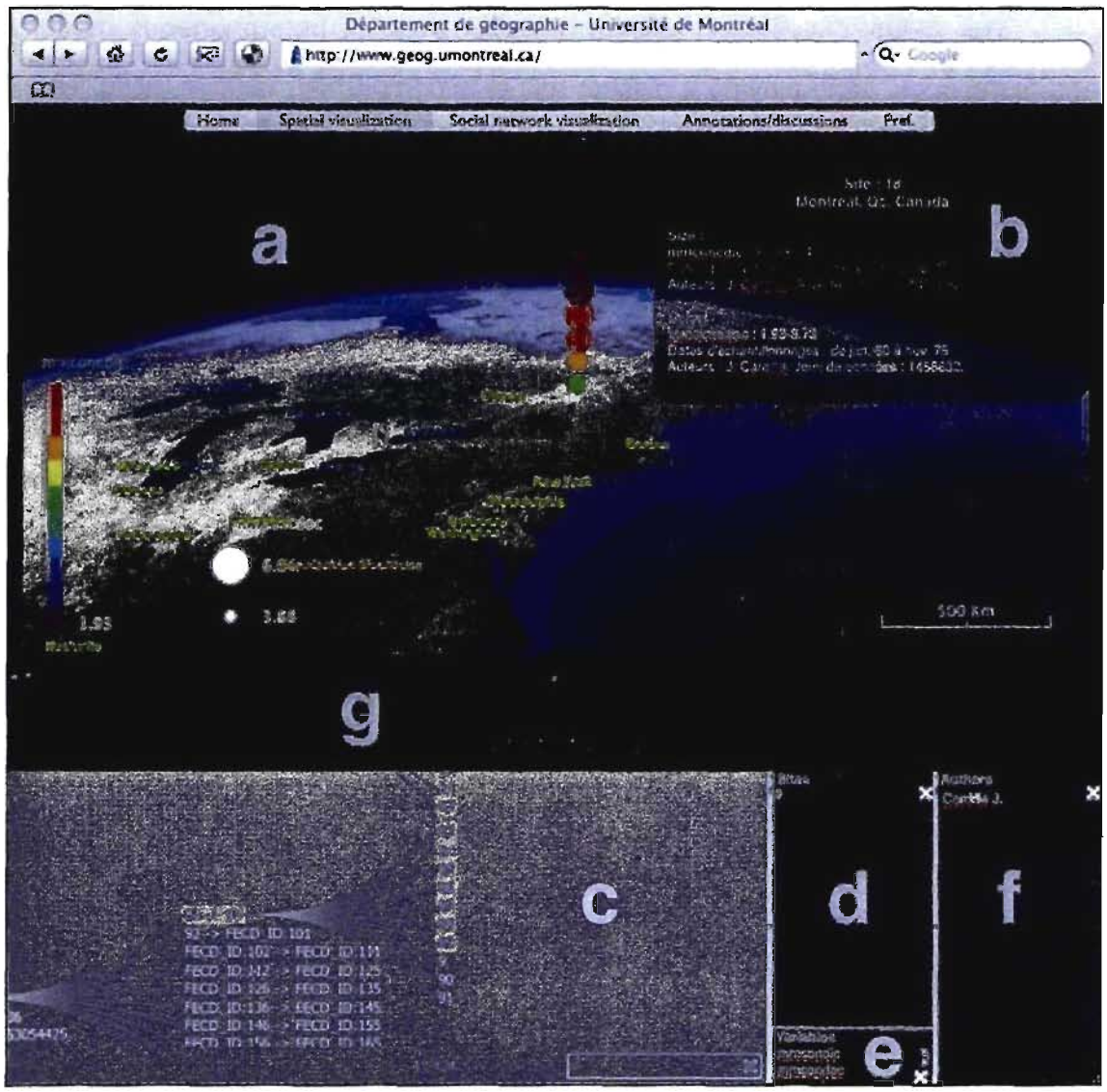

Figure 6: The main interface of the application, showing the spatiotemporal visualization (a) of data selected in "c"and "g" and filtered through lists in "d", "e" and "f". «B» provides a space to display extra information about the selected point. The displayed data are fictionous and used only to demonstrate the visualizations capabilities of the application.

3.3.1. Exploration of the database

SFMN GeoSearch's database is accessed and explored through queries built by users. While traditional GIS mainly use SQL (Structured Query Language) to retrieve, analyze data and build geographic visualizations from a database, we chose to keep the queries simple from a user's point of view. We designed two means to retrieve information from the database while keeping users connected to the context of their data exploration:

3.3.2. Hierarchical tree exploration

The entire database can be browsed through a textual exploration in a dynamic hierarchical tree view. The tree view represents the content of the database through a hierarchy of categories. Categories closer to the «root» of the tree are more general, while the ones closer the «leaf», the last items of the tree, are more precise. The exploration can be done based on variables, projects, sites or authors of studies. As a category, or node, is selected, more branches expand to display what belongs under it. For instance, if a user is looking for all research done in the province of Québec, he/she would first select the node «Sites», then «Countries», then «Provinces», and finally «Québec». All research located in Québec will then be made available. The process (four mouse clicks) is more intuitive than a SQL query. It also holds a better sense of context as users jump from more general categories to more particular ones while keeping a clear idea of his/her position in the ramifications of the tree (see figure 6-C).

3.3.3. Selecting data according to their authors' interconnections

Although the hierarchical exploration of the database is satisfactory in most cases, it shows some limitations when it comes to queries based on social interconnections. In a research network such as the SFMN, members are related to each other through different characteristics. Some may have worked together in the past, some may share the same research interests, some may belong to the same sector or institution. Tightly connected social networks emerge from these interconnections. As we saw, one way to display data collected by one particular person is to look in the tree view for the person's name. The researcher's work is then available under it. From this simple query, one logical next action to extend the search could be, for instance, to see who works on the same subjects. This cannot easily be done through selecting the next name on a list. In the tree, members are sorted in alphabetical order, and no list exists with specific criterions such as «people working on the same project». This raises an important limitation: hierarchical explorations of data cannot capture the relationships from which social networks are built. For this reason, we have chosen to include a graph-based view of the SFM research network in SFMN GeoSearch, which more readily facilitates exploration of the database from a social context.

Graph-based methods of data exploration help anchor people-oriented data searches in their social contexts, and provide a better general understanding of the social interactions within the community. Throughout the XXth century, social sciences, in their efforts to understand human interactions, have developed methods to analyze and map social networks (Scott 1991). These analysis methods, based on mathematical graph theories, led to powerful network visualizations (also called sociograms) where nodes, representing each member of the network, are connected to other nodes when sharing a particularity. The result

is a drawing of interconnected nodes that effectively maps a network of relationships and offers a clear presentation of the global structure of the social interactions at work between members.

3.3.4. Exploring interconnections and selecting data via a graph of the network

New methods of interactive network visualization such as *prefuse* (Heer 2004, Heer et al. 2005) have recently emerged and can be used towards designing better mediums for understanding social networks, as well as to explore, discover and select people-related resources in a database. Working from the concept of the traditional sociograms described above, *prefuse* innovates by taking advantage of recent advances in computer graphics and computational speed to present dynamic and, most importantly, interactive views of social networks. Once a graph has been shaped by selecting the links that should represent the social connections in the network, users can interact with the drawing by zooming in and out, or by panning in all directions of the plan. This helps focus on specific people and follow connections from one scientist to another (figure 7 and 8). In a research network such as the SFMN, where nearly each node is directly related to potentially valuable resources, this intuitive way to explore graphs proves itself to be: (1) Highly suitable to locate and access people-related data of potential interest in order to share them among members (Schwartz 1993). While alphabetically sorted lists of members falling in similar categories could technically provide access to data according to people's social connectivity, they would lack the constant sense of context that makes exploring a social network, and discovering its underlying resources, as easy as reading a map; (2) Useful for members of a

dense, geographically scattered network to comprehend their personal connectivity to the rest of the community. Beyond one or two degrees of separation, it is hard to mentally map one's own social connectivity, let alone understand other members' surroundings. Having an interactive and organized map of the network gives users an overall view of the general structure of their community and, by doing so, promotes awareness of their social environment. (3) Illustrate the role that a large centre of excellence plays in connecting researchers within a field, by allowing comparisons of the network connectivity before and after the centre's existence.

Figure 7: An example of a social network represented as a graph with *prefuse*. Members are connected with each other if they have worked on the same projects (green edges), or

if they belong to the same institution (blue edges). *Prefuse* takes advantage of a force-driven layout system to find an equilibrium point where sub-groups, hubs, isolates etc. are made visible. This visualization helps finding data through their authors' social interconnections.

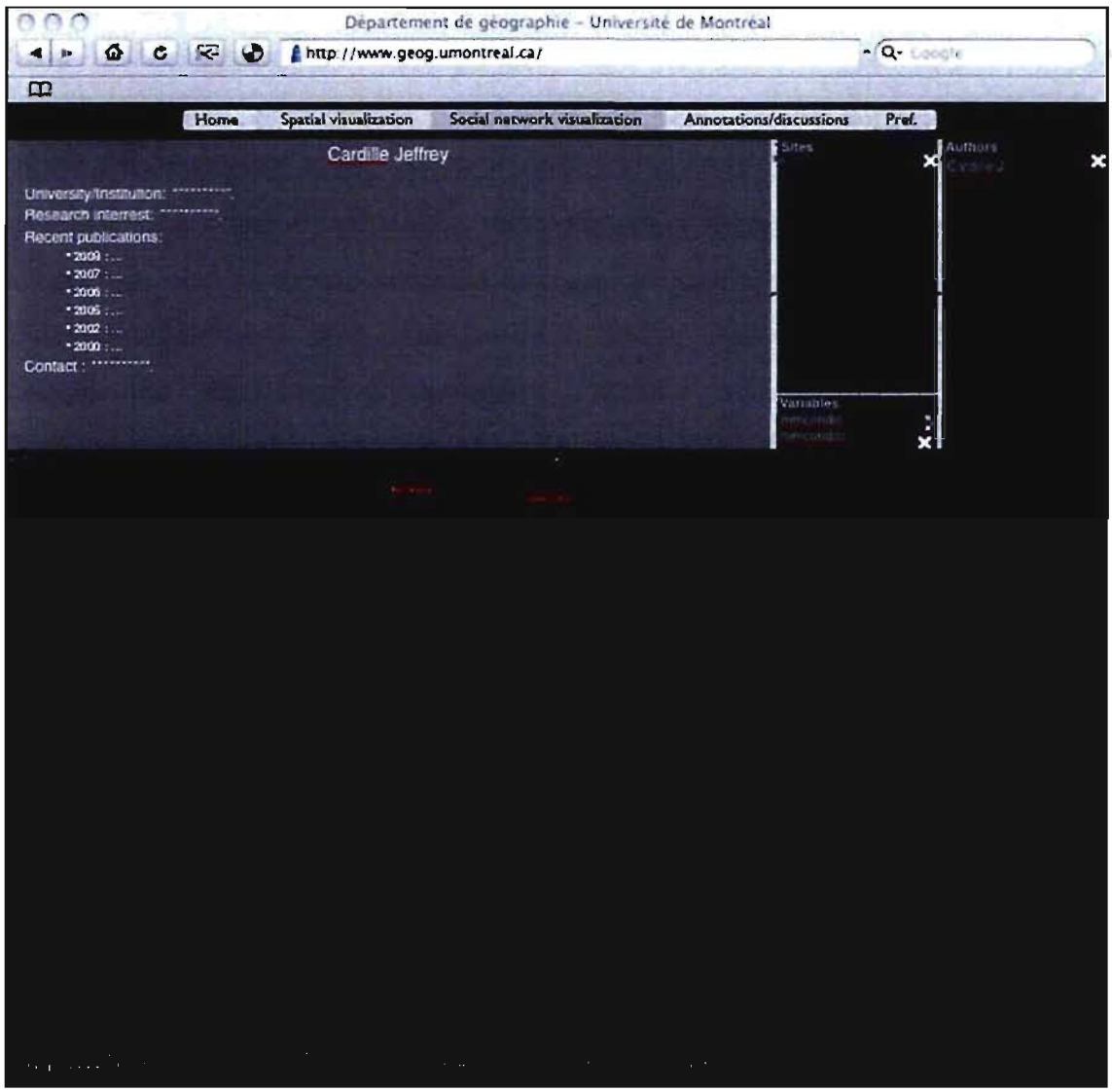

Figure 8: Implementation in our system of the *prefuse* network visualization library. It allows intuitive explorations of the database via an interactive graph of people's interconnections. Each node represents a member of the network, while the lines show different kinds of possible connections between the members (here, a blue line links two

people having worked on the same projects, and a green line link researchers belonging to the same institution).

3.3.5. Selecting data according to date

Considering the value of the temporal dimension in interpreting change and discovering patterns in ecological data, it is important that the exploration of the database include a way to segregate data along a time scale. The database is populated with variables sampled at different places and at different times. Some sites have had the same variable sampled many times on the same spot. Similarly, a variable taken from a variety of datasets sampled across a wide geographic range is unlikely to have been sampled on the same date. This makes exploring and displaying time sensitive variables problematic: without taking into account their date of sampling, visualizing different datasets simultaneously could potentially be misleading. Therefore, users do need to be able to explore the database and select variables to be displayed according to their temporal dimension. We address this issue by displaying at all times a dynamic time scale (figure 6-G) linked to the geographical visualization described in the next section. Every time a query is done and a visualization processed, the time scale adjusts itself to show the range of dates related to the displayed data, from the oldest sample to the most recent one. Users then (1) have a precise idea of the range of time stamps related to the data they are dealing with, and (2) are able to narrow down the search to certain dates in order to concentrate on the evolution of a variable within a time window of interest.

3.4. Spatiotemporal visualization of field data

New digital cartography applications are available today, featuring immersive methods to visualize and interact with georeferenced layers of information. There are two groups within the large family of geographic visualization clients available today. Some of these applications, like Google Maps or Yahoo! Maps, use projected maps or satellite imagery on a flat, two-dimensional medium. Others, like World Wind, Google Earth or Microsoft Virtual Earth, feature a full three-dimensional environment where a model of the earth, around which stitched-together satellite images of different resolutions are wrapped, can easily be interacted with. These systems effectively free users from preset geographic visualizations by providing intuitive ways to interact with the cartography. The user controls the position of his/her point of view with the computer mouse, as if manipulating a virtual camera from a remote location in outer space. Similarly, it is easy to move further back to visualize large features of the globe, zoom in to get closer to a specific region to observe details through satellite images of finer resolution. Users can also roll, or tilt the sphere, explicitly showing the topography of the area, and allowing the exploration of an accurate model of the earth's geography in both its horizontal and vertical dimensions. Indeed, these applications do provide, through the continuity of the maps and the liberty of movement, a more intuitive and higher level of spatial understanding. For these reasons, we believed such scale-free geographic exploration medium would be an important addition to our system, and we worked towards a tight integration of NASA World Wind's (WW) virtual globe technologies within our ecological database exploration tool. At the beginning of the project in 2007, several systems were offering a fully interactive three-dimensional visualization of the globe. However, only WW had made available their source code as a Java library, allowing us to compile the

program as a web browser-compatible Java applet. Furthermore, unlike products like the ones offered by Google, Yahoo! and Microsoft, WW's code is open source, and does not require any data to travel through third party servers. We believe this is an important point as our system could potentially be dealing with sensitive data.

While WW easily displays data spatially, visualization of queries in the temporal dimension has to be addressed in order to provide important context information to users. In the context of a meta-database, where data have been sampled by different people at different times and at different geographical positions, a medium explicitly displaying their spatiotemporal contexts provides strong visualizations in term of the understanding it can bring to the user. Along with spatial information indicated by two geographic coordinates, sampled variables stored in the database carry a time stamp. This extra dimension relative to each and every sample allows us to show their relative temporal values along with their geographic position. Thus, while spatial information is represented by plotting a point on the surface of the map, time is represented through the vertical, third spatial dimension of the virtual globe. These three values (latitude, longitude, and time) use the three geometric dimensions available on the 3D model of WW. Displaying the values of the variables then requires other means of visualization.

3.4.1. Visualizing variables

Aiming for a simple and intuitive method to display variable values — while continuing to show their positions in space and time —, we designed a multidimensional visualization system based on a set of coded icons. While multivariable geographic visualization systems do exist (Kreuseler 2000, Andrienko 2003), we decided to restrict the display to two variables chosen in the database by

the user for each space-time location for reasons of clarity. In addition to showing variable values, we designed icons that give additional information –when applicable- about the dispersion of the means of the two variables displayed. With this added information, the system is effectively showing more than 5 dimensions of data.

There are different ways to show information within an icon. Color, relative size, and shapes coding are effective and commonly used to display information (Bergeron and Grinstein 1989). We decided to use this system and designed a set of icons to provide users with clear indications of the value of the variables he/she chooses to display. These variables can display one or two variables in two different situations: the «top» view showing variables in two dimensions, and the alternative «tilted» view, taking advantage of the third dimension to display the time dimension.

Displaying one variable

When only one variable is to be displayed, the application shows a set of icons positioned on the sites where the variables have been sampled. The variables' values are indicated by their relative sizes. As a rule, the lowest value will always be assigned to the same icon size (about 10 pixels in diameter). Similarly, the highest value will always be allocated to the same icon size (30 pixels in diameter). All other values will be distributed along a linear scale between these two extremes. No matter what changes of view the user chooses to do (zooming, panning, etc.), these sizes stay the same and match a scale displayed on the left side of the view window, showing the two extreme sizes as well as their corresponding values (figure 9).

Figure 9: Visualizing one variable per site through the size of the icons.

Displaying two variables

When two variables are selected, the color of the icon comes into play. Values are separated in ten classes along a linear scale, and matched to a scale of colors (in the following order, from lowest to highest: purple, dark blue, light blue, cyan, turquoise, green, light yellow, dark yellow, orange, and red). The lowest values are then automatically set to purple, while the highest ones are assigned to red (figure 10).

Figure 10: Visualizing two variables per site: one through the color, the other one through the size of the icons.

If the value tied to the color is missing on some sites, a grey icon shows the size. If, on the other hand, the missing value is the one relative to the size, a set of fuzzy icons (lacking a clear border), of constant size (10 pixels) is used by the system (figure 11).

Figure 11: Fuzzy icons show variables linked to colors only, with no size information.

Mean values

As stated before, one of the advantages in keeping such a large database is to be able display variable evolution through time as well as space. This creates situations where the same variables have been sampled several times at different dates on the same site. All these data would then share, if not the same time stamp, the same spatial coordinates. Instead of having icons layered one upon another, with the latest date masking the other icons, we chose to automatically calculate the mean and the standard deviation of these stacked values. The only icons displayed are then the average of all the values belonging to the same point in space. However, this implies that icons should also bear a visual cue giving users an idea of the dispersion of data around the calculated mean.

Visualizing standard deviation

We need two standard deviation cues: one for the mean related to the size of the icons, and one for the mean related to the color of the icons. Once again, trying to find solutions that would not require users to spend a long time observing the screen, we chose clarity over detailed value representation. For the value relative to the size of the icon, a 50% transparent stroke is applied to the icon; the thickness of the stroke corresponds to the standard deviation. For the color-related values, a half-disc of which the transparency is relative to the value of the standard

deviation, is superimposed to the icon. Each of these visual representations falls into four classes (figure 12). For each icon, the choice of the class is relative to all the other standard deviations visualized at the same time on the globe. In other words, the lowest standard deviation values displayed at one time falls under the lowest cue class, the highest value fall into the highest one. The other standard deviation values are classified according to a linear scale.

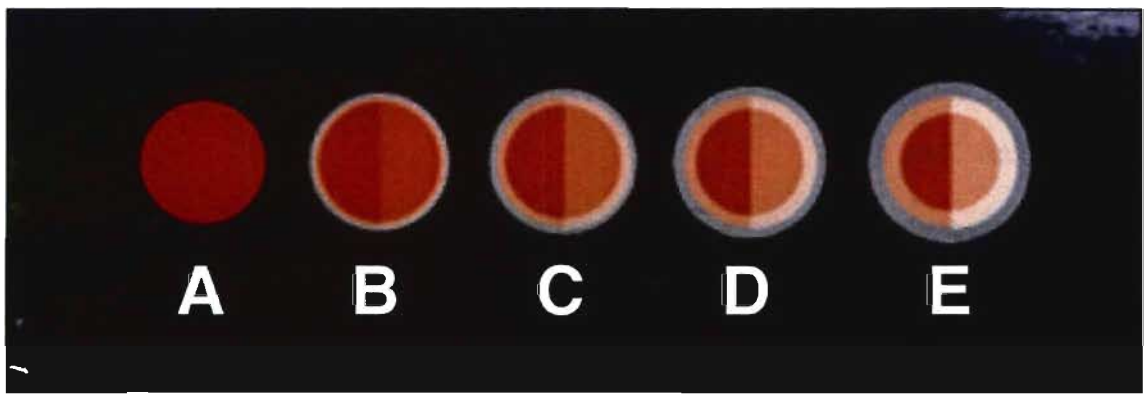

Figure 12: Icon A shows data that haven't been averaged or aggregated. Icons B to E show increasing values of standard deviation related to variable linked to color (the half-disc becomes whiter as standard deviation increases), and to variable linked to the size of the icon (the stroke at the border of the icon becomes wider as standard deviation increases).

Time evolution of variables in tilted view

The limitations of the default «top» view of the globe, which lead to the need for averaging stacked icons, can be effectively overcome by taking advantage of the third geometric dimension of WW. As users tilt their point of view, the vertical dimension of the virtual earth becomes visible, showing the topography of the region of interest at the same time as opening a new space to display data. This extra dimension permits the visualization of data according to their «position» in time. For each icon, an elevation index is attributed according to the date of

sampling. For instance, if a variable has been sampled five times on the same site, watching the scene from a tilted point of view will display each variable on a vertical time scale, where the oldest is represented closer to the ground, the most recent is shown higher towards outer space, and the three others are distributed in the space in between (figure 13a). A blank space is used if values are missing for certain dates, which leaves an empty slot. Alternatively, even if data have been sampled only once per site on a wide spatial range, tilting will quickly and effectively show an indication of the time value of the samples, potentially helping to interpret a variable's evolution in space with an explicit reference to its sampling date (figure 13b).

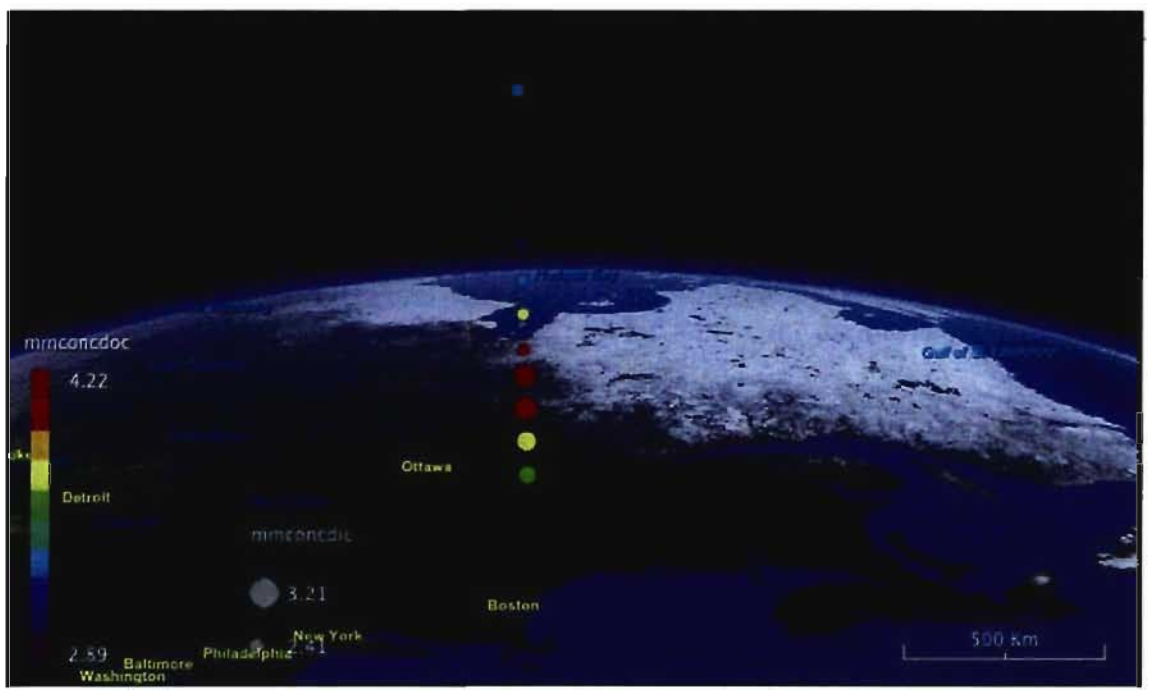

Figure 13a: Evolution in time of two variables on one site. The bottom icon is the oldest data point, while the highest one is the most recently sampled. Users can easily see trends of the two variables in time. The displayed data are fictionous and used only to demonstrate the visualizations capabilities of the application.

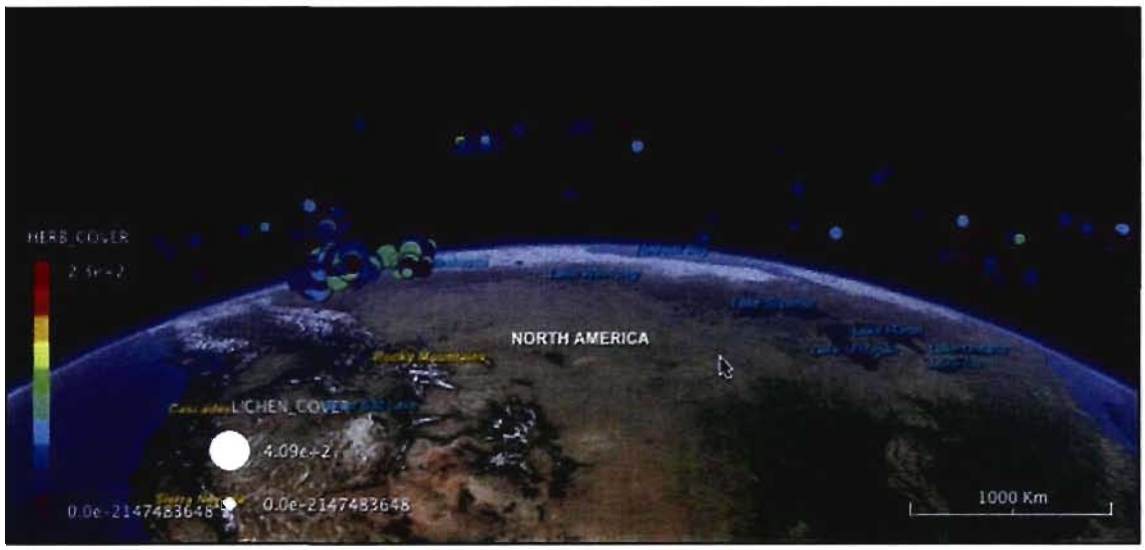

Figure 13b: Two variables have been sampled only once on each site. As with figure 13a, the bottom icon is the oldest data point, and the highest one is the most recently sampled. The difference in elevation quickly shows the chronology of the sampling process, and brings valuable information about a possible correlation between the values of displayed variables and their sampling date.

In total, these icons bear four dimensions (color, size, stroke thickness and half disc overlay transparency), which are to be added to the three spatiotemporal dimensions. The total of seven dimensions displayed at once does not match other multidimensional solutions (Chernoff 1973), but keeps visualizations simple enough to be understood at a glance. This, we believe, is an important criterion for a database geographic browsing tool such as this.

3.5. Discussion and annotation about datasets

Although essential, formal Metadata are rarely enough to describe all the details of original research, and informal communication is often very much necessary (Zimmerman 2007). There is theoretically no sufficient amount of

metadata to implement in order to fully describe a dataset. The extra information needed for the effective second-hand use of data will therefore usually have to come from the authors themselves. When scientists consider reusing second hand data in their own research, a few sensitive points arise (Zimmerman 2003). Among them is the essential need for the person about to reuse data to have a sufficient level of understanding of the dataset (data acquisition methods used, condition of sampling, etc.), in order to build confidence in the compatibility of the original fieldwork with the second-hand analysis. This confidence is mostly achieved by interpreting metadata in light of their own data acquisition experience. However, in some cases metadata is not enough and a personal communication with the authors of the original work, when possible, is a necessary step (Zimmerman 2007). To ease this communication, we implemented a module to discuss the datasets displayed by our system. For each visualization, a corresponding page showing a list of the datasets (with their corresponding geographic localizations) used to build the visualization can be accessed. Users can then select one of the datasets to open a corresponding discussion board and ask questions or exchange constructive considerations about it with other researchers and with the author of the dataset.

3.6. How it all works together

The modular architecture of SFMN GeoSearch makes its use very flexible: all the modules described in this paper are linked together, and actions in one module lead to either results or new opportunities to explore further the data in another. Therefore, describing formal sequences of actions to extract and display

data is difficult. Let us follow, to illustrate how the system can be interacted with, just one example of the use of the application for data discovery.

Displaying data on the virtual globe knowing the variables' names or site name

Once a variable name has been double-clicked in the tree menu (figure 6-C), its name is added to a list (figure 6-E) and the value of this variable is automatically shown on the virtual globe (figure 6-A) as an icon on the globe, everywhere it has been sampled. Also, the names of the authors of the studies from which the displayed values have been taken appear in the «Authors» list (figure 6-F).

If the values have been sampled several times on the same site, the application automatically computes and shows their means, as well as an indication of their standard deviations. As we discussed above, only two variables can be visualized at once (for one site at one date), so the variable list (figure 6-E) will not exceed two entries. If carefully selected, the icons can, with their different colors and/or size, help users to understand trends through the landscape. Once again, as dates of sampling are important to comprehend changes of value through space, users can, at all times, incline their view of the globe to turn the spatial visualization into the spatiotemporal visualization described above. Similarly, users can retrieve data by knowing the name of the study site, which is accessible through the hierarchical tree menu. Sites can be accessed by either alphabetical ordered lists or by rough toponymic classification.

Displaying variables on the virtual globe knowing their authors' names

Likewise, the data selection can be author oriented. Users can, in the «Social Network» tab, explore the social connections within the research network and select authors of interest. Back at the main page, the list of authors will be filled with the selected names. No data is displayed at this point, and the names appear in grey, showing that they are not yet connected to displayed values. To start using these names for data visualization, and to select data related to a particular researcher, users need to double-click on a name in this list. The hierarchical tree menu then expands to display the work of the selected author. As previously explained, variables can then be added to the «Variable» list. This action will update the «Sites» list, while the name of the author of the dataset will switch to white, showing that his/her data is actually shown, and the virtual globe will display the variable on the site studied by the researcher. Other researchers can be added by the same means.

Narrowing down a selection

All these queries can be narrowed down to a particular region; users can either zoom into the 3D environment of the virtual globe or specifically choose which sites to display in the tree menu. Moreover, users can narrow down their selection to periods of time of interest by selecting series of dates in the timescale (figure 6-G).

3.7. Future work

There are many ways in which our tool could be improved. Future work falls into four main areas: (1) The consistency of the database. Gathering ecological variables collected by different researchers for different goals, with various

methods and equipment, leads to heterogeneity in the database. Likewise, similar variables are often named through different terms. Ontology and metadata management can address these essential problems. These are active domains of research providing new solutions for formal description and effective management of ecological data. We believe that systems such as Metacat (Jones 2001) could greatly improve the overall quality of SFMN GeoSearch's database; (2) Better geographic exploration. Users should be able to use the virtual globe as much for visualizing as for building queries. The exploration of the database will therefore be pushed towards a deeper integration of the geographic tool as an exploration tool. For instance, drawing a rectangle on the globe, or simply zooming toward a smaller portion of territory, would narrow the query to the specified spatial extent; (3) Data visualizations. For now, our system can only display punctual quantitative values. An important next step would be to implement ways to display qualitative data, as well as integrating additional means to visualize scientific data through, for example, raster maps or polygons; and (4) Finally, we need to polish the user interface and improve the general user experience by collaborating with a sample of researchers from the network.

3.8. Discussion

Today's environmental challenges often involve phenomenon acting across spatial and temporal scales, and ecological researchers require broad datasets to work from. These data will necessarily come from multidisciplinary sources to cover a wide spectrum of phenomenon spread across large spatial and temporal scales. Working towards addressing this fundamental issue, the organization of science has progressively shifted towards larger-scale scientific collaborations, and

led to the emergence of knowledge networks (Geuna 2003). Networks such as LTER, which has been grouping research sites across the USA for almost 30 years, providing scientists with large amounts of spatially and temporally broad ecological data, are a good example of what has been accomplished in the last decades. Comprehensive technologies to manage them (Jones 2007), as well applications dedicated to exploring and visualizing the wealth of organized data made available, are taking shape. They use the Internet as a medium to connect users to the meta-database, and are most often distributed via web portals (Halpin 2006, Flemons 2007). SFMN GeoSearch builds on some aspects of these systems to provide novel shared data exploration and spatiotemporal visualizations.

Space being an important dimension in discovering patterns in ecological data, most of these exploratory systems use some kind of cartographic plotting tools to display the results of user's queries. For the most part, these cartographic components are two-dimensional, which implies limitations in terms of multidimensional data visualization. At least with regards to the addition of the third dimension, virtual globes are a promising alternative. While some projects already use external applications such as Google Earth to display the results of queries made via a web portal (Gemmell 2007), few or none integrate three-dimensional cartography within the interface of a data sharing system. The interaction between the geographic component and the query is then slowed down by the constant need for downloading and handling visualization files between the web portal and the visualization environment. One of the ameliorations provided by SFMN GeoSearch is precisely the tight integration of a three-dimensional cartographic module within the query window. Users can seamlessly build queries and see their result instantly on the globe, within the same window. This reactivity allows users to

try many combinations of queries in the least amount of time, making the exploration of potentially large databases much more efficient.

Other than providing a more realistic environment to display data on, virtual globes make space for displaying more information through their third dimension. In ecology, time is another sensitive dimension, and cartographic methods have been designed to display time through the third spatial dimension (Hägerstrand 1967, Kraak 2003). The interactive nature of three-dimensional globes such as WW makes it an ideal support for visualizing the relative time value of any given sampled variable. The KML file format can carry size, color, and elevation information about any icons, and can therefore display values in all the dimensions our system allows. Building from this, we introduce with SFMN GeoSearch a flexible way to dynamically alternate visualizations according to the point of view on the virtual globe at any given moment: the time values of variables sampled on a site are shown when the user chooses to tilt the globe, alternatively, the mean and standard deviation is displayed when the point of view is directly vertical to the site. Once again, we believe the sense of interaction, as well as the reactivity of the visualizations on the virtual globe allows for a better connection between the users and the database.

3.9. Conclusion

The system we have designed aims at helping members of scientific networks to share, discover and reuse ecological field data. SFMN GeoSearch is able to receive georeferenced datasets relative to field or modeled ecological data, to organize them, to allow users to explore and to visualize them intuitively, and to download relevant datasets. Once logged into the system, members of the network

can access datasets for their own use. These data can be explored through: (1) a hierarchical menu giving access to the whole database, or (2) a network visualization interface that provides clear and fast ways to discover data through their author's social connections. The latter can efficiently address problems related to the lack of social connection awareness within the social network. Our system also helps facilitate discussions between members of the network (and hopefully between the authors and the potential second-hand users) about datasets contained in the database. While being built for the SFM Network, our system is open source and can be adapted to most other ecological networks.

Ecological variables being tightly connected to their spatial and temporal contexts (Kreuseler 2000), we believe visualizing aggregated datasets along these spatiotemporal dimensions is important for data discovery in large meta-databases. When displayed at the same time, ecological data produced and managed by different people at different dates for research relative to different sites can then be understood: (1) Within their geographical and temporal contexts, (2) Across different scales, since variable evolutions can be visualized just as easily at continental or site level, and (3) In comparison to each other. Therefore, a tool such as SFMN GeoSearch, leveraging novel and interactive database exploration and spatiotemporal visualizations, can facilitate our understanding of trends through space and time in such a large amount of shared data, and effectively assist users in spotting the right datasets to reuse in new research.

3.10. Acknowledgments

We thank The Sustainable Forest Management Network and the Centre d'Étude de la Forêt (Forest Study Centre) for their financial support. We thank Jim

Fyles for believing in this project and for providing regular feedback and support during the development of the SFMN Geosearch system.

3.11. Extra material

A demo of SFMN GeoSearch can be found at:

http://meta.geog.umontreal.ca/currentprojects/forests/

It requires a recent personal computer (running either Microsoft Windows, Max OS X, or Linux/Unix). A good amount of memory and a 3D accelerated graphic card are necessary hardware. On the software side, an up-to-date web browser such as Mozilla Firefox, and a recent Java runtime environment are needed (the latter two are free and widely available).

4. Conclusion générale

Les avancées de ces dernières décennies, dans les domaines de l'informatique et des technologies de l'information, constituent la fondation technique à un changement dans les usages liés à la gestion des données de recherche. Alors que jusqu'à récemment, ces données restaient principalement cantonnées au bureau de leurs auteurs (ou, au mieux, profitaient à leurs étudiants ou collègues proches), il est désormais techniquement possible de les partager rapidement, efficacement et en grand nombre, sous forme de fichiers informatiques. Ces informations, cataloguées dans de grandes bases de données, peuvent ainsi contribuer aux recherches d'autres scientifiques dont les besoins en données surpassent, souvent de loin, leur capacité à les échantillonner dans un temps et dans un budget raisonnables.

Si ces bases technologiques sont aujourd'hui bien établies, les outils capables de référencer, d'explorer et de mettre à disposition toutes ces données hétérogènes sont encore relativement rares. Le plus souvent, les organismes gouvernementaux ou les grands réseaux de chercheurs mettent leurs bases de données scientifiques en ligne, offrant aux chercheurs une interface textuelle sous la forme d'un site Internet. L'utilisateur peut alors explorer les données disponibles grâce à un système de requête par mots clefs exploitant les informations contenues dans les métadonnées des jeux disponibles. Depuis quelques années toutefois, certains répertoires offrent des systèmes plus sophistiqués qui intègrent un module de représentation spatiale. Ceux-ci permettent à l'utilisateur de visualiser rapidement les données qui l'intéressent sur un support géographique sommaire. En ce qui concerne les études environnementales, cette option apporte

une valeur ajoutée considérable : l'espace fait partie des dimensions primordiales à la compréhension de nombreux phénomènes, et une représentation cartographique, même sommaire, peut aider à l'identification de jeux de données potentiellement utiles à de nouvelles recherches. Aux côtés de la dimension spatiale, la dimension temporelle a également une grande importance. Malheureusement, les procédés couramment utilisés pour visualiser l'espace offrent peu de place à la visualisation du temps. Dans ce contexte, les globes virtuels ont un potentiel à faire valoir : de par les avantages qu'ils ont sur les cartes statiques traditionnelles, ils pourraient être utilisés dans de nouveaux outils de partage de données scientifiques.

Le travail présenté dans ce mémoire a mené à la création de SFMN GeoSearch, un outil géographique exploitant les nouvelles possibilités offertes par la technologie des globes virtuels. Après presque un an de programmation pour deux personnes, les concepts développés lors des deux années de recherche de cette maîtrise sont en bonne partie intégrés à l'outil. Certains aspects tels que la visualisation spatio-temporelle et l'exploration hiérarchique de la base de données sont robustes et exploitables immédiatement. D'autres, tels que l'intégration de l'outil de sélection par auteurs ont été implémentés de manière expérimentale et nécessitent un travail d'optimisation. D'autres, encore, qui ne sont pas directement liés au corps de la recherche, nécessitent des travaux complémentaires de développement. Plus particulièrement, il est nécessaire de considérer les importantes et délicates questions de la gestion de l'hétérogénéité du contenu de la base de données et de la description des jeux de données qui le constitue. En ce qui concerne l'interface spatio-temporelle, SFMN GeoSearch pourrait bénéficier

dans le futur de certaines améliorations intégrables relativement facilement sur la base de ce qui est déjà développé. Pour ne citer que quelques exemples :

— le globe virtuel est un outil particulièrement adapté à l'exploration de contenu géographique. Une amélioration importante au système consisterait à utiliser ce globe non seulement pour visualiser le résultat de requêtes, mais également pour les construire. Le champ de vision choisi par l'utilisateur pourrait, par exemple, constituer le bornage spatial de la requête. La sélection d'une zone du globe à l'aide de la souris pourrait sélectionner les sites qu'elle contient. etc. ;

— le module de visualisation ne traite actuellement que des données ponctuelles, une évolution possible serait d'offrir la possibilité de superposer à l'imagerie fournie par WW des données sous forme de couches matricielles. Il sera également possible de visualiser ces données dans le temps selon le même principe que ce qui a été développé pour les icônes : par un étagement en altitude des différentes couches ;

— la forte concentration de sites dans certaines zones géographiques mène, lorsque l'utilisateur choisit un point de vu trop éloigné, à des empilements d'icônes qui nuisent à la lisibilité. Un remède à ce problème consisterait à agréger automatiquement les icônes qui se superposent au moins partiellement. Une seule icône représenterait, par la moyenne des valeurs (avec la même indication d'écart type que ce qui est décrit dans la section 3.4.1), plusieurs icônes visuellement trop proches les unes des autres ;

— certains ajustements visuels sont nécessaires à une compréhension plus rapide des données. Par exemple, il est parfois difficile de déterminer à quels

lieux correspondent certaines icônes lorsque de nombreuses données sont affichées simultanément en vue inclinée. Une croix au sol ainsi qu'une ligne verticale liant tous les points issus du même site pourraient clarifier certaines visualisations.

SFMN GeoSearch met en œuvre, à partir de technologies de pointe, des méthodes modernes d'exploration, de visualisation dynamique et de partage de données scientifiques géoréférencées. Parmi ces nouvelles technologies, deux en particulier sont utilisées de manière innovante dans ce projet.

La première est celle de la cartographie des réseaux. La bibliothèque de programmation Java *prefuse* permet la visualisation dynamique des interconnexions au sein de réseaux sociaux. Dans le cadre de ce projet, cette visualisation est utilisée à la fois pour permettre à l'utilisateur de comprendre les cercles professionnels dans lesquels il évolue (par une cartographie de son environnement social) et pour offrir une méthode d'exploration, plus efficace dans un certain contexte, de la base de données : si l'arbre hiérarchique est adapté à la sélection de données catégorisées, il est trop rigide pour une recherche efficace quand celle-ci est liée aux personnes. La visualisation et la sélection dans un graphe permettent d'affranchir l'utilisateur des listes prédéfinies en le laissant explorer la base de données partagée selon les interconnexions sociales de leurs auteurs.

La seconde, qui est à la base de ce projet, est celle des globes virtuels. Elle offre une perception nouvelle de l'espace cartographié en permettant à l'utilisateur de parcourir une représentation plus fidèle du monde de manière continue, souple, intuitive et interactive. L'utilisateur a la liberté d'explorer l'espace géographique à différentes échelles (du village, dans lequel on peut parfois distinguer routes et

chemins, à la planète dans son ensemble) tout en conservant une idée claire du contexte spatial. Les globes virtuels permettent ainsi une compréhension rapide et claire de l'espace et constituent de ce fait un support de choix pour la visualisation de données environnementales. Construite sur les bases de cet outil puissant, la méthode de visualisation temporelle mise au point dans le cadre de ce travail apporte à l'utilisateur une dimension supplémentaire pour interpréter les données affichées. Cette visualisation spatio-temporelle peut, dans de nombreux cas, s'avérer être une aide efficace à la découverte de patrons dans les données rassemblées.

Ce travail a ainsi permis la mise au point d'un outil de partage où des données, auparavant cataloguées de manière textuelle, peuvent désormais être liées les unes aux autres sur un support spatio-temporel. SFMN GeoSearch place la géographie au cœur de la problématique du partage de données environnementales et permet, en mutualisant des travaux accumulés individuellement à travers les années, d'aider à la création de nouvelles connaissances scientifiques.

Références

Abrams, J., et Hall, P. (2006). *Else/where: Mapping new cartographies of networks and territories.* Minneapolis, MN: University of Minnesota Design Institute.

Andelman, S.J., Bowles, C.M., Willig, M.R., et Waide, R.B. (2004). Understanding environmental complexity through a distributed knowledge network. *BioScience*, 54, 240-246.

Andrienko, N., Andrienko, G., et Gatalsky, P. (2003). Exploratory spatio-temporal visualization: An analytical review. *Journal of Visual Languages and Computing*, 14, 503-541.

Arzberger, P., Schroeder, P., Beaulieu, A., Bowker, G., Casey, K., Laaksonen, L., et coll. (2004). Science and government: An international framework to promote access to data. *Science*, 303, 1777-1778.

Barabàsi, A. (2003). *Linked: How everything is connected to everything else and what it means for business, science, and everyday life.* New York: Plume Books.

Beddow, J. (1990). Proceedings from the 1st Conference on Visualization '90: *Shape coding of multidimensional data on a microcomputer display.* Los Alamitos, CA: IEEE Computer Society Press, 238-246.

Bucy, E. (2000). Social access to the internet. *Harvard International Journal of Press/Politics*, 5, 50-61.

Butler, D. (2006). Virtual globes: The web-wide world. *Nature*, 439, 776-778.

Campbell, E., Clarridge, B., Gokhale, M., Birenbaum, L., Hilgartner, S., Holtzman, N., et coll. (2002). Data withholding in academic genetics: Evidence from a national survey. *JAMA*, 287, 473-480.

Chernoff, H. (1973). The use of faces to represent points in k-dimensional space graphically. *Journal of the American Statistical Association*, 68, 361-367.

Ferreira de Oliveira, M.C., et Levkowitz, H. (2003). From visual data exploration to visual data mining: A survey. *IEEE Transaction on Visualization and Computer Graphics*, 9, 378-393.

Fienberg, S., Martin, M., et Straf, M. (1985). *Sharing research data.* Washington, DC: National Academy Press.

Flemons, P., Guralnick, R., Krieger, J., Ranipeta, A., et Neufeld, D. (2007). A web-based GIS tool for exploring the world's biodiversity: The global biodiversity information facility mapping and analysis portal application (GBIF-MAPA). *Ecological Informatics*, 2, 49-60.

Foresman, TW. (2008). Evolution and implementation of the digital earth vision, technology and society. *International Journal of Digital Earth*, 1, 4-16.

Foucault, M. (1980). *Power/knowledge: Selected interviews and other writings, 1972-1977.* New York: Pantheon Books.

Gemmell, A.L., Blower, J., Haines, K., et Smith, G. (2007). Proceedings from the American Geophysical Union, Fall Meeting 2007: *Using virtual globes and a Java web application to visualize and compare ocean observations and model data.* San Francisco: American Geophysical Union.

Geuna, A., Salter, A.J., et Steinmueller, W.E. (2003). *Science and innovation: Rethinking the rationales for funding and governance.* Cheltenham, UK: Edward Elgar Publishing.

Golle, P., Leyton-Brown, K., et Mironov, I. (2001). Proceedings from the 3rd ACM conference on Electronic Commerce: *Incentives for sharing in peer-to-peer networks.* Heidelberg, GE: Springer Berlin.

Goodchild, M.F. (2008). The use cases of digital earth. *International Journal of Digital Earth*, 1, 31-42.

Ohazama, C. (2008, 11 Février). Truly global. Message posted to: http://google-latlong.blogspot.com.

Grinstein, G., Pickett, R., et Williams, M. (1989). Proceedings from Graphics Interface'89: *EXVIS: An exploratory visualization environment.* London, ON.

Guralnick, R.P., Hill, A.W., et Lane, M. (2007). Towards a collaborative, global infrastructure for biodiversity assessment. *Ecology Letters*, 10, 663-672.

Hägerstrand, T., Pred, A., et Haag, G. (1967). *Innovation diffusion as a spatial process*. Chicago: University of Chicago Press.

Hägerstrand, T. (1970). What about people in regional science? *Papers in Regional Science*, 24, 6-21.

Halpin, P.N., Read, A.J., Best, B.D., Hyrenbach, K.D., Fujioka, E., Coyne, M.S., et coll. (2006). OBIS-SEAMAP: Developing a biogeographic research data commons for the ecological studies of marine mammals, seabirds, and sea turtles. *Marine Ecology Progress Series*, 316, 239-246.

Heer, J. (2004). Prefuse: A software framework for interactive information visualization. (Masters dissertation, University of California, 2004).

Heer, J., et Boyd, D. (2005). Proceedings from the 2005 IEEE Symposium on Information Visualization: *Vizster: Visualizing online social networks*. Need Location/Publisher. Washington, DC: IEEE Computer Society Press.

Inselberg, A., Dimsdale, B., Center, I., et Los Angeles, C. (1990). Proceedings from the 1st Conference on Visualization '90: *Parallel coordinates: A tool for visualizing multi-dimensional geometry*. Los Alamitos, CA: IEEE Computer Society Press

Jameson, F. (1991). *Postmodernism, or, the cultural logic of late capitalism*. Durham, NC: Duke University Press.

Jones, M.B., Berkley, C., Bojilova, J., et Schildhauer, M. (2001). Managing scientific metadata. *Internet Computing IEEE*, 5, 59-68.

Jones, M. (2007). Meta-information systems and ontologies: A special feature from ISEI '06. *Ecological Informatics*, 2, 193-194.

Kraak, M.J. (n.d.). *What has ITC done with Minard's map?* Retrieved from http:// www.itc.nl/personal/kraak/1812/minard-itc.htm.

Kraak, M.J. (2003). Geovisualization illustrated. *ISPRS Journal of Photogrammetry and Remote Sensing*, 57, 390-399.

Kreuseler, M. (2000). Visualization of geographically related multidimensional data in virtual 3D scenes. *Computers and Geosciences*, 26, 101-108.

Kwan, M. (2000). Interactive geovisualization of activity-travel patterns using three-dimensional geographical information systems: A methodological exploration with a large data set. *Transportation Research Part C*, 8, 185-203.

Latour, B. (1999). Circulating reference: Sampling the soil in the amazon forest. In B. Latour, *Pandoras hope: Essays on the reality of science studies* (pp. 24-79). Cambridge, MA: Harvard University Press.

LeBlanc, J., Ward, M., et Wittels, N. (1990). Proceedings from the 1st Conference on Visualization '90: *Exploring n-dimensional databases*. Los Alamitos, CA: IEEE Computer Society Press.

Madin, J., Bowers, S., Schildhauer, M., Krivov, S., Pennington, D., et Villa, F. (2007). An ontology for describing and synthesizing ecological observation data. *Ecological Informatics*, 2, 279-296.

Michener, W., Brunt, J.W., Helly, J.J., Kirchner, T.B., et Stafford, S.G. (1997). Nongeospatial metadata for the ecological sciences. *Ecological Applications*, 7, 330-342.

Michener, W. (1998). Ecological metadata. In W.K. Mitchner, J.H. Porter, et S.G. Stafford (Eds.), *Data and information management in the ecological sciences: A resource guide.* Albuquerque, NM: LTER Network Office, University of New Mexico.

Michener, W. (2006). Meta-information concepts for ecological data management. *Ecological Informatics*, 1, 3-7.

Milgram, S. (1967). The small world problem. *Psychology Today*, 2, 60-67.

Monmonier, M. (1990). Strategies for the visualization of geographic time-series data. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 27, 30-45

Nourbakhsh, I., Sargent, R., Wright, A., Cramer, K., McClendon, B., et Jones, M. (2006). Mapping disaster zones. *Nature*, 439, 787-788.

Olson, R., Voorhees, L., Field, J., et Gentry, M. (1996). Proceedings from Eco-Informa '96: *Packaging and distributing ecological data from multisite studies.*

Parr, C., et Cummings, M. (2005). Data sharing in ecology and evolution. *Trends in ecology and evolution*, 20, 362-363.

Pickett, R.M. et Grinstein, G.G. (1988). Proceedings of the 1988 IEEE International Conference: Iconographic Displays For Visualizing Multidimensional Data. *Systems, Man, and Cybernetics*, 1, 514-519.

Porter, J., et Callahan, J. (1994). Circumventing a dilemma: Historical approaches to data sharing in ecological research. Environmental Information Management and Analysis. CRC Press

Porter, T. (1995). *Trust in numbers: The pursuit of objectivity in science and public life.* Princeton, NJ: Princeton University Press.

Ranganathan, K., Ripeanu, M., Sarin, A., et Foster, I. (2004). Proceedings from the 2004 IEEE International Symposium on Cluster Computing: *Incentive mechanisms for large collaborative resource sharing.* Washington, DC: IEEE Computer Society Press.

Schelling, T. (1978). *Micromotives and macrobehavior.* New York: Norton and Company, Inc.

Schöning, J., Hecht, B., Raubal, M., Krüger, A., Marsh, M., et Rohs, M. (2007). Proceedings from the 13th International Conference on Intelligent User Interfaces: *Improving interaction with virtual globes through spatial thinking: Helping users ask "Why?".* New York: ACM.

Schwartz, M.F., et Wood, D.C.M. (1993). Discovering shared interests using graph analysis. *Communications of the ACM*, 36, 78-89.

Scott, J. (2000). *Social network analysis: A handbook.* Thousand Oaks, CA: Sage Publications.

Servilla, M., Costa, D., Laney, C., San Gil, I., et Brunt J. (2008). Proceedings from the 2008 Environmental Information Management Conference: *The EcoTrends web portal: An architecture for data discovery and exploration.*

Smith, P., Falloon, P., Kîrschens, M., Shevtsova, L., Franko, U., et Romanenkov, V. (2002). EuroSOMNET - a European database of long-term experiments on soil organic matter: The WWW metadatabase. *The Journal of Agricultural Science*, 138, 123-134.

Stanley, B., et Stanley, M. (1988). Data sharing. *Law and Human Behavior*, 12, 173-180.

Star, S. (1985). Scientific work and uncertainty. *Social Studies of Science,* 15, 391.

Tooth, S. (2006). Virtual globes: A catalyst for the re-enchantment of geomorphology? *Earth Surface Processes and Landforms*, 31, 1192-1194.

Vasiliev, I. (1997). Mapping time. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 34, 1-51.

Watts, D. (1999). Networks, dynamics, and the small-world phenomenon. *American Journal of Sociology*, 105, 493-527.

Watts, D. (2003). *Six degrees: The science of a connected age.* New York: W.W. Norton & Co.

Westbrooks, E. (2004). Proceedings from the 2003 International Conference on Dublin Core and Metadata Applications: *Distributing and synchronizing heterogeneous metadata for the management of geospatial information repositories.* Singapore: Dublin Core Metadata Initiative.

Zimmerman, A. (2003). *Data sharing and secondary use of scientific data: Experiences of ecologists.* (Doctoral Dissertation, University of Michigan, 2003).

Zimmerman, A. (2007). Not by metadata alone: The use of diverse forms of knowledge to locate data for reuse. *International Journal on Digital Libraries*, 7, 5-16.

Zimmerman, A. (2008). New knowledge from old data: The role of standards in the sharing and reuse of ecological data. *Science, Technology, & Human Values*, 33, 631-652.