



68th IFLA Council and General Conference

August 18-24, 2002

Code Number: 102-157-F
Division Number: V
Professional Group: Government Information and Official Publications - Workshop
Joint Meeting with: -
Meeting Number: 157
Simultaneous Interpretation: -

Les systèmes d'information géographique (SIG) et le partage des données entre bibliothèques : thème de réflexion pour les bibliothèques numériques

Nancy C. Hyland

Albert R. Mann Library, Cornell University

Ithaca, USA

E-mail : nch9@cornell.edu

Introduction

Le département dépositaire d'informations à référence spatiale de l'université Cornell (CUGIR : Cornell University Geospatial Information Repository) permet, sur le *web*, la recherche, le balayage et le téléchargement de données et de métadonnées [mécanismes permettant aux producteurs de décrire les données qu'ils détiennent] à référence spatiale concernant l'Etat de New York. Des sujets tels que le relief et la topographie, les sols, l'hydrologie, les risques naturels, les activités agricoles, la faune, la gestion des ressources naturelles ont leur place dans le CUGIR. La plupart des informations contenues dans le CUGIR proviennent de sources officielles.

La bibliothèque Mann de l'université Cornell s'est investie dans les systèmes d'information géographique depuis 1994. La demande de données officielles SIG s'est accrue de manière exponentielle et il s'est souvent avéré difficile, pour les utilisateurs et les bibliothécaires, de les obtenir. Le CUGIR a été créé pour combler cette lacune. Depuis ses débuts à Mann dans le courant de l'automne 1998, plus de 100 000 structures de données géographiques ont été téléchargées. Le personnel du CUGIR continue de travailler avec les agences gouvernementales pour fournir des données aux chercheurs, aux urbanistes et à d'autres agences officielles.

Que sont les SIG?

Les systèmes d'information géographique (SIG) sont constitués de matériels, de logiciels et de contenus combinés pour créer une base de données relationnelle permettant le repérage et l'analyse de toute information ayant une composante spatiale. Bien qu'un SIG soit parfois conçu simplement comme un outil cartographique, c'est sa capacité à emmagasiner de l'information et à la relier à un simple point sur la carte qui fait sa force. De nombreux services d'urgence, par exemple, utilisent un SIG. En cas d'appel d'urgence, un régulateur, rien qu'en identifiant le numéro d'appel, localise l'endroit d'où l'appel a été émis, calcule la route la plus rapide pour s'y rendre, détermine s'il y a des éléments présentant un risque dans un rayon de deux pâtés de maisons et où se trouvent les secours les plus proches. Dans ces systèmes, les données utilisées peuvent provenir de nombreuses sources et organismes différents.

Un SIG peut aussi aider à l'analyse et à la compréhension de problèmes démographiques, environnementaux ou agricoles. De nombreuses agences gouvernementales prennent des décisions en matière d'urbanisme ou d'environnement en se fondant sur des analyses menées avec l'aide de SIG. On peut ainsi structurer de l'information à propos de l'écoulement des eaux dans une région agricole proche de marécages afin de déterminer quels secteurs doivent être soumis à une plus stricte réglementation de l'usage des pesticides. Les pouvoirs publics encouragent souvent les citoyens à s'impliquer et à participer à des enquêtes ou des réunions conduites par des spécialistes de SIG. Ces analystes établissent des cartes à l'aide de SIG dont les variables sont déterminées en fonction de la décision à prendre dans tel cas spécifique d'aménagement. Les cartes sont un instrument très efficace pour comprendre rapidement des ensembles complexes de données.

L'information est créée à partir de mesures prises sur le terrain par levé géodésique ou à distance avec les photographies aériennes et les images satellites, ou encore à partir de la numérisation de cartes existantes. Une fois que les cartes numériques ont été dressées, on peut créer des liens entre elles et toutes sortes de données à référence spatiale. L'établissement de telles cartes revient cher, en argent et en temps. Il n'est donc pas étonnant que ce soient les pays aux infrastructures les plus développées qui aient justement développé le plus de SIG et par là même, produit le plus d'information. Ceci dit, les bases de données fondamentales, modèles d'élévation numérisée, hydrographie, topographie, existent quasiment pour le monde entier. Sans données, un SIG n'est d'aucune utilité. De plus en plus de données deviennent disponibles, mais l'utilisateur final peut encore éprouver quelque difficulté à les localiser.

En raison de la nature des données des SIG, les administrations ne les ont pas diffusées aussi largement que d'autres documents officiels. Il a fallu mettre en place de nouveaux modes de diffusion. Le comité fédéral américain des données géographiques (FGDC : Federal Geographic Spatial Data Committee) a prôné une meilleure distribution des données des SIG à la fois à l'intérieur des Etats-Unis et sur le plan international par le truchement de son infrastructure pour les données géographiques nationales (NSDI : National Spatial Data Infrastructure). Il s'agit d'une série de nœuds de partage de données reliés entre eux par une norme d'échange et une interface de recherche¹. La NSDI compte ainsi plus de 250 nœuds à l'échelle nationale et internationale. En mettant en place des normes de données géographiques et de mécanismes pour les décrire (metadata) par le biais de la NSDI, le FGDC a fait progresser de manière significative la mondialisation de l'échange des données des SIG. Les bibliothèques étant faites pour conserver, organiser et fournir l'information, elles sont idéalement placées pour héberger de tels points de connexion.

¹ National Research Council (1994). Promoting the National Spatial Data Infrastructure Through Partnerships. Washington, D.C., National Academy Press.

Les SIG dans les bibliothèques

Lorsqu'une bibliothèque envisage d'ajouter un service SIG à sa palette documentaire, elle doit d'abord mesurer le degré d'implication possible de son personnel. Il y a trois types de services avec les SIG : les services réduits à l'essentiel, une collection physique, l'accès à un réseau. Les recommandations suivantes sont fondées sur les huit ans d'expérience des SIG à la bibliothèque Mann.

Le premier niveau demande seulement la disponibilité du bibliothécaire assurant la permanence de service public, qui peut aider les utilisateurs à naviguer entre les différents nœuds de données et qui doit être familier des logiciels SIG accessibles gratuitement tel ESRI (<http://www.esri.com>). N'importe quel bibliothécaire ayant un peu d'intérêt et d'aptitude pour la chose sera en mesure d'apporter une première assistance SIG au bout d'une vingtaine d'heures de formation, soit reçue en stage ou par autoapprentissage, et en repérant quels types d'information sont disponibles sur des sites tels que :

The GIS Data Depot <http://www.gisdatadepot.com>

http://www.gisnet.com/notebook/GIS_Resources.htm

US based data <http://www.cast.uark.edu/local/hunt/index.html>

<http://www.fgdc.gov/nsdi/nsdi.html>

<http://www.gsdi.org/main2.html>

<http://www.spatial.maine.edu/~onsrud/GSDI/htm>

<http://www.gisdevelopment.net/tutorials/>

<http://www.csc.noaa.gov/products/sccoasts/html/gistut.htm>

A ce stade minimal, il n'est pas possible de guider les utilisateurs jusqu'à la création d'un SIG entièrement intégré. La bibliothèque servira plutôt de simple intermédiaire permettant à l'utilisateur de trouver ses renseignements en utilisant les cartes infographiques de base.

Entretenir une collection physique demande un plus gros effort en termes de budget et de personnel. En plus des coûts de matériels et de logiciels, il faut prévoir un temps considérable pour former le personnel au catalogage des disques compacts, au renseignement du public et à l'enseignement du maniement des logiciels, ainsi qu'aux acquisitions mêmes des disques de données. Un grand nombre d'articles ont paru dans la littérature professionnelle sur la mise en place de ce type de service SIG. Je recommande notamment la lecture de Dean Jue : « Implementing GIS in the Public Library Arena² » et Jurg Buhler : « Digital Maps in Map Collections³ ».

A la bibliothèque Mann, le gros de nos efforts en matière de SIG est porté sur le troisième type de service : l'accès à distance à une bibliothèque numérique ou 'géobibliothèque'. Fournir l'accès aux données des SIG dans une bibliothèque numérique suppose la participation de tous les départements de la bibliothèque. Le CUGIR est un nœud de la NSDI depuis 1998. A Mann, des bibliothécaires de différents

² Jue, D. K. (1996). Implementing GIS in the Public Arena. 1995 Clinic on Library applications of Data Processing, Urbana, Ill., University of Illinois.

³ Buhler, J. (1999). « Digital Maps in Map collections - Presenting new electronic media. » LIBER Quarterly, the journal of the European research libraries 9 (2) : 228-234.

départements (services techniques, développement des collections, nouvelles technologies, service public) ont participé à l'entreprise. Ceux des services techniques ont installé les métadonnées, très nombreuses pour un service SIG, à l'intérieur du CUGIR. Actuellement, ces mécanismes de description emploient la norme FGDC. A l'été 2002, la norme FGDC fusionnera avec la norme ISO. Le processus de conversion est en cours. Nos métadonnées sont aussi converties en xml, sgml, html, DublinCore et format MARC de manière à en permettre le plus large accès possible⁴. En outre, un autre genre de métadonnées a vu le jour, intitulées « métadonnées de service ». Elles décrivent les capacités de serveurs cartographiques expérimentaux à partager des données sans canevas. Notre collègue du service des nouvelles technologies assure la maintenance du serveur hôte du CUGIR, conçoit une nouvelle base relationnelle pour un meilleur accès et s'occupe de la programmation nécessaire à l'interface entre le web et la norme Z39.50. Le service du développement des collections apporte son aide pour la conservation et la politique documentaire. Et enfin, le responsable du service public a en charge l'aide apportée à l'utilisateur final, c'est lui aussi qui prend contact et négocie avec les partenaires qui fournissent les données.

La négociation avec les partenaires fournisseurs de données

Nous avons commencé à mettre en place le CUGIR en partenariat avec le département de protection de l'environnement de l'Etat de New York (NYSDEC) et le laboratoire des systèmes d'information sur les sols (SISL). La plupart des données du CUGIR proviennent soit directement de sources officielles, soit en sont dérivées. D'autres données à l'échelon local ont également été ajoutées en partenariat avec des agences locales d'urbanisme. Ces partenariats se sont avérés extrêmement profitables ; néanmoins un certain nombre de problèmes doivent être abordés et négociés dès le départ lorsqu'on s'engage avec des partenaires gouvernementaux.

Ce qu'on en attend ; ce qu'ils en attendent ?

Comme dans toute négociation, il est utile de désigner clairement les bénéfices que chaque partie tirera de l'accord. En un sens, le CUGIR fonctionne pour les administrations comme un intranet à distance. Les données d'un SIG ont un très vaste champ d'application et peuvent consister en centaines de fichiers différents. Certains employés des agences gouvernementales racontent qu'avant ils devaient conserver les données sur leurs disques durs ou appeler un collègue, localisé dans un autre building, qui avait les données voulues sur un CD. Depuis que les données sont accessibles sur le CUGIR, l'accès à leurs propres données est devenu plus aisé et plus rapide. De plus, la plupart des services qui produisent des données SIG passent beaucoup de temps à les fabriquer et à les améliorer. Ils sont mal équipés ou n'ont simplement pas le temps de répondre à l'attente du public en fournissant les données sur CD. Autoriser la bibliothèque à fournir l'accès aux données leur évite d'avoir à régler la question de l'utilisateur final.

La nature des données des SIG les rend difficiles et parfois impossibles à exploiter sans les métadonnées adéquates. De nombreux professionnels des SIG regardent la mise en place de ces mécanismes descriptifs comme une tâche vraiment onéreuse. Cela aide considérablement le public général désireux d'utiliser ces données, mais ce peut aussi être la seule documentation laissée par un professionnel. La mobilité des personnels peut se traduire par une somme assez conséquente de travail à refaire à chaque fois si les métadonnées sont incomplètes. D'un autre côté, les catalogueurs sont tout à fait aptes à décrire le travail et les données et les convertissent aisément dans la norme FGDC ou dans les autres normes de métadonnées. A la différence des catalogues traditionnels, les métadonnées ne peuvent être créées si les données n'ont pas été entrées. Le personnel du CUGIR travaille en étroite collaboration avec nos partenaires et assurent la formation et le conseil pour la création des métadonnées. En dernier ressort,

⁴ Chandler, A. & E. Westbrooks (2002). « Distributing Non-MARC Metadata : The CUGIR Metadata Sharing Project. » Library Collections, Acquisitions & Technical Services. A paraître.

nous sommes les éditeurs des métadonnées et assurons leur compatibilité avec les normes les plus courantes. Nos partenariats avec les agences gouvernementales nous permettent de nous tenir au courant des tendances de la communauté des SIG et nous aident ainsi à mieux répondre à l'attente de nos usagers.

Conservation, versions et mises à jour

Une autre question à prendre en compte lors des négociations est celle de la conservation des données. L'avènement de l'information numérique et en ligne a d'abord suscité l'espoir que les problèmes de conservation allaient s'alléger. Une fois le document numérisé, étant facile à reproduire, sa perte serait moins à craindre. Certes de multiples efforts de conservation au cours des dix dernières années ont entraîné la mise sur support numérique de nombreux documents. La durée de vie des documents numérisés, cependant, est bien plus courte que celle de n'importe quel autre équivalent, en partie à cause de la dégradation du support. Il y a presque une relation inversement proportionnelle entre le temps passé à produire un document et la longévité du support sur lequel il est conservé⁵. Un plan précis de sauvegarde des versions plus anciennes et périmées des données doit être établi et négocié avec les partenaires.

Nombreuses sont les données constamment revues et corrigées, ce qui a des incidences sur la conservation. Quand des données sont mises à jour, il est important de s'assurer si l'on a affaire à une simple mise à jour ou à une nouvelle version. Dans le premier cas, il peut juste s'agir de la corrections d'erreurs : le bon sens voudrait alors que l'on détruise purement et simplement la version antérieure. En réalité, les producteurs de données souhaitent souvent que les versions antérieures d'une série soient détruites dès qu'une nouvelle version paraît. On peut toutefois objecter que, si un projet est en cours ou même achevé, il peut être important de connaître l'état d'avancement des données au fur et à mesure de la réalisation du projet car elles en constituent la mémoire.

Dans certains cas, les données sont mises à jour uniquement parce que l'agence a décidé de changer le système de projection ou le système de coordonnées. Ce sont les concepts géographiques de base employé dans les SIG. Pour faire vite, ce sont des modèles mathématiques qui permettent de faire apparaître en deux dimensions sur l'écran ou le papier les trois dimensions de la terre. Pour en savoir plus long, se reporter à : « Datums and projections : a brief guide » sur <http://biology.usgs.gov/geotech/documents/datum.html>. Si l'utilisateur final consulte des données de différentes sources, les trois modèles peuvent se mélanger. Si, par exemple, une couche d'information sur les sols est recouverte par une couche d'information hydrographique et qu'elles proviennent de systèmes différents, le sol du lit de la rivière pourrait apparaître à trente mètres du cours d'eau lui-même.

Lorsqu'une administration décide de changer de modèle et met à jour ou corrige ses données en conséquence, elle peut souhaiter détruire les données antérieures. Ce peut être pour des raisons d'ordre juridique, mais nous nous sommes aussi rendu compte que c'est souvent parce qu'elle craint que l'utilisateur final ne confonde les données anciennes et actuelles. Ce qui peut en effet arriver si deux bases de données se trouvent sous le même nom dans la bibliothèque numérique. Bien que l'information au sujet de la date et de la version soit enregistrée dans les métadonnées, les producteurs de données doutent, peut-être à juste titre, que les utilisateurs y fasse attention. Quant à nous, nous avons créé une zone d'archivage. L'utilisateur ne pourra pas télécharger sans avoir été pris connaissance de la date des données. C'est parce que nous avons créé cette zone que certaines organisations nous permettent de continuer à offrir l'accès aux anciennes données.

⁵ Conway, P? (1996). « Preservation in the digital world. » Commission on Preservation & Access, Council on Library and Information Resources : 24 p.

Il est possible de transférer les anciennes données d'un support sur un autre sans les rendre accessibles en ligne. Cette stratégie peut fonctionner mais doit être mise en oeuvre avec précaution. Si, par exemple, je devais copier des données périmées du CUGIR sur des CD, je ne devrais pas seulement préciser très scrupuleusement ce que sont ces données et inclure toutes les métadonnées pertinentes, mais je devrais aussi garder trace de la manière dont elles ont été transférées sur CD, avec quel matériel et avec quel logiciel. Il conviendrait aussi de programmer le moment où il faudra à nouveau les transférer. Se contenter de mettre les CD dans une boîte et les y oublier signifierait en définitive la perte des données. S'il est bien compris par l'institution que les données doivent toutes les x années être réenregistrées sans omettre les précisions nécessaires, il y a plus de chance qu'elles ne se perdent pas. Conserver les données accessibles en ligne augmentent leur chance de ne pas être oubliées.

Accès aux données et propriété intellectuelle

En plus des problèmes de conservation, plusieurs questions touchant à l'accessibilité et à la propriété intellectuelle doivent être débattues avec les partenaires :

1. A qui appartiennent les données et à quel prix?

Une fois que les données sont disponibles dans la géobibliothèque, est-ce qu'elles appartiennent à cette dernière ou à l'agence gouvernementale? Ce point doit être précisé même si la bibliothèque les acquiert à titre gratuit. Des événements imprévus peuvent rendre cette question de propriété cruciale pour la continuité de l'accès.

2. Comment peut-on exploiter les données?

D'une manière générale, toutes les données du CUGIR peuvent être exploitées pour toute application, y compris commerciale. Cela est spécifié dans les métadonnées et l'on n'a pas à y revenir à chaque fois au niveau du réseau. S'il y a des restrictions de communication pour certaines données, les appliquer ne nous pose pas de problème technique, et nous pouvons prendre une telle décision au cas par cas.

3. Qui peut télécharger les données?

De nombreux réseaux d'information géographique ont mis en place des mécanismes permettant de ne donner l'accès à tout ou partie des données qu'à certains publics. Au CUGIR, il n'y a pas de restrictions significatives. Certaines agences gouvernementales ont décidé de ne pas verser leurs données au CUGIR à cause de cela. Il convient de se mettre d'accord au commencement de la négociation.

Au CUGIR nous avons voulu expressément que toutes les données soient accessibles sans restrictions. Cela nous a privé de certaines données géographiques, mais nous a aussi permis d'offrir une interface plus rapide et conviviale. Il nous a semblé que le CUGIR fournissait l'accès à des données ne présentant aucun risque pour l'ordre public et qu'au contraire leur large diffusion était dans l'intérêt général. En février 2002, le Bureau de la sécurité publique de l'Etat de New York a demandé la fermeture du CUGIR afin de vérifier qu'il ne fournissait l'accès à aucune donnée présentant un risque pour la sécurité nationale. Toute référence aux ponts, aux aéroports pouvait être considérée comme une menace. Le CUGIR a procédé lui-même à une évaluation des risques et a maintenu son réseau ouvert en dépit de la demande⁶. Nous avons informé nos partenaires de notre décision : l'un d'entre eux nous a demandé de retirer ses données du réseau le temps qu'il procède lui-même à une évaluation des risques, ce que nous avons fait.

Evaluation des risques

⁶ Knezo, G. J. (2002). « Possible impacts of Major Counter Terrorism Security Actions on Research, Development and Higher Education. » Congressional Research Service.

Nous avons partiellement obtempéré à la demande qui nous a été faite en février en passant en revue toutes les données disponibles sur le CUGIR. Deux membres du personnel de la bibliothèque Mann ont mené à bien cette étude. Ils ont défini deux principaux critères pour l'évaluation : le degré de risque et la disponibilité. S'agissant du niveau de risque, ils ont passé en revue chaque structure de données en envisageant quel usage des criminels ou des terroristes pourraient faire de l'information. Le second critère reposait sur la possibilité d'obtenir la même information par d'autres sources. Cette question de la diffusion a été considérée comme un facteur important parce que, si l'on peut penser que la localisation d'un aéroport constitue un risque pour la sécurité, et si d'autre part cette information est disponible de mille autre manières, on ne peut, dans la pratique, la classer parmi les données à risque. Les données ont été notées de 1 à 5, 1 pour le degré moindre de risque, 5 pour le plus élevé⁷. L'analyse a montré que nos seules données de type 5 étaient largement diffusées sur d'autres supports. Notre méthode d'analyse était fondée sur l'évaluation des risques présentés par les données numériques auxquelles la bibliothèque Mann fournit accès en collaboration avec le ministère américain de l'Agriculture⁸. L'enquête a été menée conjointement avec l'agence qui a retiré toutes ses données. A l'heure où j'écris cet article, toutes les données ont été remises dans le CUGIR, à l'exception de trois séries. Il est souhaitable qu'une telle évaluation des risques soit conduite préalablement par la géobibliothèque, avant que les problèmes ne se posent. Chaque donnée présentant des caractéristiques uniques doit être évaluée spécifiquement.

Un autre risque avec les informations à référence spatiale concerne la responsabilité civile. Avec une telle masse d'informations et un si grande précision dans le détail, les producteurs de données commettent presque inévitablement des erreurs. De nombreuses agences hésitent à donner au public l'accès à leurs données de peur de poursuites judiciaires en cas d'erreur. Toutes les données contenues dans le CUGIR sont précédées d'une mention déchargeant leur auteur de toute responsabilité indiquant entre autres qu'« il appartient entièrement à l'utilisateur de déterminer le bien-fondé de l'usage » qu'il en fera. De telles précautions devraient être de règle dans toute bibliothèque numérique.

Conclusion

La nature des données des SIG suppose une certaine souplesse dans leur mise en place au sein de la bibliothèque numérique. Il faut préparer très attentivement les partenariats avec les producteurs de données et satisfaire au mieux leurs exigences afin d'assurer ensuite le fonctionnement le plus fluide possible de la procédure. Les bibliothèques sont le lieu idéal pour fournir sur le long terme un accès convivial aux données. Les partenariats avec les organisations gouvernementales s'avèrent profitables tant pour la bibliothèque que pour les administrations. Grâce à une planification soigneuse, nous espérons quant à nous, à la bibliothèque Mann, continuer à offrir un service de qualité dans les années à venir grâce au CUGIR.

⁷ Martindale, J. (2002). National Security and Access to GIS Data via the Internet : Cornell University Geospatial Information Repository (CUGIR). ESRI Education User Conference, San Diego, CA. A paraître.

⁸ Lawrence, G., W. Kehoe, et al. (1999). Risk Management of Digital Information : A File Format Investigation : A Report Prepared for the Council of Library and Information Resources. Ithaca, NY, Cornell University : 1-19.