

Archivage, gestion et préservation des données géospatiales

Rapport sommaire et recommandations

Noyau consultatif sur les politiques de GéoConnexions

Groupe de travail sur l'archivage et la préservation des données géospatiales

David L. Brown : Division de la cartographie de l'architecture et de la géomatique,
Bibliothèque et Archives Canada
Grace Welch : Université d'Ottawa
Christine Cullingworth : Chercheure

Février 2005

TABLE DES MATIÈRE

Rapport sommaire et recommandations	i
1. Introduction.....	1
2. Contexte	1
3. Contexte des activités de préservation au gouvernement	2
3.1 Environnement des politiques gouvernementales.....	2
3.2 Milieu de la recherche	4
3.3 Milieu institutionnel	5
4. Problèmes de préservation des données	5
4.1 Désuétude de la technologie	5
4.1.1 Représentation et formats des données	6
4.1.2 Technologies de stockage	6
4.1.3 Solutions technologiques à l'obsolescence	7
4.2 Gestion temporelle des données gouvernementales	8
4.3 Documentation et métadonnées	9
5. Conclusions et Recommandations	10
5.1 Recommandations pour une action au niveau des institutions	10
5.2 Recommandations d'action à l'échelle nationale	12
Références	14

Rapport sommaire et recommandations

Sous l'égide du Noyau sur les politiques de GéoConnexions, un groupe de travail a été chargé de faire le point sur les problèmes et de trouver des solutions liés à l'archivage et à la préservation à long terme des données géospatiales. Le présent rapport sommaire présente les points saillants et les recommandations d'une étude plus approfondie intitulée « La gestion et la préservation des données géospatiales ».

Des données géospatiales sont créées par tous les ordres de gouvernement et le secteur privé à une vitesse sans précédent. Cependant, l'accès à long terme à ces précieuses données risque d'être compromis si les dépositaires des données n'élaborent et mettent en place des politiques et procédures afin d'assurer la préservation des archives et l'accès aux données pour les décideurs, l'industrie et les chercheurs. De nombreuses recherches se penchent sur la question de la préservation de l'information électronique, mais bien peu se concentrent sur les défis que représentent l'archivage réussi et la préservation des données géospatiales.

Cette étude conclut qu'il n'existe pour l'instant aucune solution unique aux défis que nous réserve la préservation de l'information numérique. Les conclusions de cette étude concordent avec les études antérieures, comme l'a noté Au Yeung (2004) « ...la véritable solution de la préservation numérique pourrait être non pas technologique mais plutôt politique ». Il est cependant clair qu'une combinaison de stratégies doit être mise en place dans une structure de gestion de l'information efficace. La mise en œuvre dépendra des besoins des organisations en matière de législation, de politiques et de gestion de l'information et de leur aptitude à investir dans les activités de préservation à long terme.

Les recommandations qui guideront l'action et les recherches liées à l'archivage et à la préservation à long terme des données géospatiales peuvent se regrouper sous : 1) les activités mises en place par les institutions et 2) les activités nécessitant des politiques, des normes et des lignes directrices nationales qui pourraient être mises au point en collaboration entre des organisations

Recommandations pour une action au niveau des institutions

Pour préserver à long terme et sauvegarder de précieuses ressources, les producteurs de données doivent d'abord adopter un mode de gestion du cycle de vie de l'information de façon à s'assurer une gestion proactive de leurs données depuis leur création jusqu'à leur élimination.

Recommandation #1 : Les organisations doivent définir et mettre en application des politiques et des pratiques pour la création, l'utilisation, le maintien, la dissémination, la préservation et l'élimination des données géospatiales.

Méthode suggérée : Une étude de rentabilité doit être menée sur la création de données géospatiales de base afin de définir la valeur durable de l'information pour les activités d'une organisation et de déterminer les règles de gestion de cycle de vie qui peuvent s'appliquer à sa rétention et à son éventuelle élimination. Beagrie et Jones (2002) ont mis

au point un modèle d'arbre de décision qui peut servir de guide dans les activités de collecte des organisations.

(<http://www.dpconline.org/graphics/handbook/figure4.html>) . Le Groupe de travail sur les données historiques (1998) du Federal Geographic Data Committee (FGDC) a aussi publié une feuille de route afin d'aider les organisations à répondre à des questions telles que :

- Existe-t-il une exigence légale ou politique pour la conservation des données géospatiales?
- Les données sont-elles critiques pour la prestation des services de gestion centrales de (des) institution(s)?
- Quels sont les coûts opérationnels pour la collecte et la création des données? En cas de perte de données, quels seraient les coûts pour les recréer?
- Les données sont-elles uniques?
- Quelle est la qualité des données?
- Quelle est la valeur potentielle des données et quelle est la probabilité que certaines d'entre elles soient réutilisées au sein de l'institution ou ailleurs?

Une réponse positive à l'une ou plusieurs de ces questions pourrait automatiquement qualifier ces données pour la préservation à long terme.

Recommandation # 2 : Les organisations doivent établir des centres de responsabilité qui autorisent des individus à définir et à appliquer les principes de la gestion de l'information nécessaire pour assurer l'intégrité des bases de données géospatiales d'une organisation.

Méthode suggérée : Le modèle de dépôt représente un moyen de faciliter la gestion des données au nom des créateurs et des utilisateurs et d'assurer la continuité de livraison de l'infrastructure de données géospatiales. Dans une telle approche les producteurs de données devront :

- Créer une stratégie de gestion de données approuvée au sein de l'organisation et un plan qui détermine les ressources de données et établit les normes quant aux pratiques de collecte, de gestion et de préservation des données;
- Nommer une personne ou créer un comité qui agira comme dépositaire des données d'une organisation et qui travaillera au sein de l'organisation et veillera à ce que les politiques et procédures pour la gestion des données géospatiales soient définies, comprises et mises en œuvre;
- Créer un feuillet d'information sur l'évaluation des données pour aider à définir les données valables requérant la conservation à long terme;
- Déterminer la valeur de l'objet des données au moment de leur création à partir des besoins opérationnels et de préservation;
- Fournir des métadonnées détaillées basées sur des normes de métadonnées internationales et nationales approuvées. Les modèles de métadonnées et de données devraient être stockés avec les données;

- Inventorier les bases de données numériques et évaluer leur valeur pour la conservation et la préservation à long terme et définir un calendrier pour la migration et la modernisation des supports de données;
- Normaliser les formats de données avant de transmettre cette information au sein de l'organisation;
- Adopter des normes industrielles telles que ISO ou des solutions de source ouverte qui affranchissent les données des formats propriétaires.

Recommandations d'action à l'échelle nationale

Recommandation # 1 : Les problèmes de préservation des données géospatiales s'inscrivent dans le cadre d'une politique sur l'information et d'une stratégie de gestion des données nationales. Travaillant en partenariat avec les communautés des bibliothèques et des archives, les producteurs de données gouvernementales doivent normaliser et adopter des politiques et des pratiques organisationnelles pour guider la création, l'utilisation, la rétention, la dissémination, la préservation et l'élimination des données géospatiales de manière à en assurer l'authenticité et l'intégrité aussi longtemps que le prévoient la loi, les arrêtés ministériels et autres lois et politiques.

Méthode suggérée : Le Canada produit des données géospatiales de grande qualité. Ses politiques et méthodes pour assurer leur préservation à long terme restent toutefois à développer.

Les agences et organismes nationaux tels que le Comité mixte des organismes intéressés à la géomatique (CMOIG), le Conseil canadien de la géomatique (COCG), les agences fédérales comme Bibliothèque et Archives Canada et les conseils nationaux tels que l'Office des normes générales du Canada devraient travailler de concert à l'élaboration des stratégies suivantes :

- Promouvoir l'utilisation des normes de métadonnées approuvées au sein de la communauté de géomatique, au moyen de publicité, d'ateliers et de création d'outils
- Promouvoir l'adoption de normes « non propriétaires » permettant la création et l'échange de données géospatiales telles que les spécifications du langage GML et celles de l'Open Geospatial Consortium
- Promouvoir l'utilisation des directives et des pratiques exemplaires établies par Bibliothèque et Archives Canada pour le stockage logique et physique de l'information numérique
- Établir des listes de contrôle et autres outils pratiques pour la sélection des données à préserver

- Définir et avaliser le concept de « conservateur de données ». Les organisations productrices de données devront définir leurs rôles et responsabilités de conservation pour la gestion des produits de données géospatiales collectifs. Le concept de conservation devrait comprendre les responsabilités liées à l'imputabilité, à la fiabilité, à l'authenticité, aux normes de métadonnées et de documentation, aux formats des fichiers et aux questions d'accès
- Rechercher et recommander les modèles opérationnels viables pour une préservation qui favorise le partage de la responsabilité et des coûts associés aux activités de préservation
- Travailler avec les communautés de la recherche, telles que InterPares, participant à des projets novateurs dans le domaine de la préservation, tout spécialement ceux liés aux données géospatiales, de façon à intégrer les nouveaux développements dans les politiques et directives de gestion de l'information.

Recommandation # 2 : Le forum conjoint (fédéral/provincial/territorial) du Conseil canadien de la géomatique devrait créer un groupe de travail et inviter les parties intéressées du milieu universitaire, du secteur privé et d'autres agences fédérales, provinciales ou territoriales à collaborer afin de développer les domaines prioritaires stratégiques définis précédemment.

1. Introduction

Sous l'égide du Noyau sur les politiques de GéoConnexions, un groupe de travail a été chargé de faire le point sur les problèmes et de trouver des solutions liés à l'archivage et à la préservation à long terme des données géospatiales. Par la suite, ce groupe a formulé des recommandations dans un rapport remis en juillet 2003. Le présent rapport sommaire présente les points saillants et les recommandations d'une étude plus approfondie intitulée « La gestion et la préservation des données géospatiales ».

2. Contexte

« La préservation et la réutilisation des données et de l'information géospatiales constituent à la fois la pierre angulaire de la future économie du savoir et l'assise de la mémoire ». (Ross 2000)

Les données géospatiales, qui précisent l'emplacement géographique des éléments ou phénomènes naturels et anthropiques à la surface de la Terre, représentent aujourd'hui une source d'information indispensable pour notre société. Elles stimuleront le marché international des produits et services basés sur la géomatique, générant des retombées de 45 à 67 milliards de dollars US d'ici 2004 (Hickling 2001). D'après l'enquête de Statistique Canada de 2000 sur les services en cartographie et en arpentage, il existe plus de 2 000 entreprises canadiennes générant des revenus annuels de 1,5 milliard à partir d'activités basées sur la géomatique. De la conception des routes que nous empruntons à l'emplacement de notre lieu de travail, les données géospatiales touchent une multitude de facettes de notre vie quotidienne.

Des données géospatiales sont créées par tous les ordres de gouvernement et le secteur privé à une vitesse sans précédent. Cependant, l'accès à long terme à ces précieuses données risque d'être compromis si les dépositaires des données n'élaborent et mettent en place des politiques et procédures afin d'assurer la préservation des archives et l'accès aux données pour les décideurs, l'industrie et les chercheurs. Les décisions économique, environnementales et sociétales ne peuvent reposer uniquement sur des données actuelles; elles nécessitent une analyse temporelle qui permet de déceler les tendances, d'évaluer les conséquences et de prendre des décisions informées. Les politiques gouvernementales actuelles de préservation des données, quand il y en a, sont incohérentes ou ne permettent pas d'examiner l'ensemble des problèmes que soulève la gestion de l'information dans un environnement numérique.

On a commencé à sonner l'alarme sur la perte possible de cette information inestimable. On emploie même les termes « catastrophe » et « richesse en péril » pour décrire ce qui pourrait arriver si on ne prend aucun moyen pour assurer la préservation à long terme de ces données. On trouve déjà de nombreux exemples où des données ont été perdues, comme celles compilées par l'État de New York pour compléter son inventaire de l'utilisation des terres et des ressources naturelles qui ne sont plus consultables, car le logiciel nécessaire pour lire ces données n'existe plus (Tristram 2002). Au Canada, il faut citer l'exemple du Système de données sur les terres du Canada, où d'importantes données sur l'utilisation des terres

recueillies pour l'Inventaire des terres du Canada (ITC) ont récemment été sauvées de justesse lorsque quatre ministères ont entrepris conjointement avec une entreprise privée du nom de Spatialanalysis un important projet de restauration (Brown 1999).

De nombreuses recherches se penchent sur la question de la préservation de l'information électronique, mais bien peu se concentrent sur les défis que représentent l'archivage réussi et la préservation des données géospatiales.

3. Contexte des activités de préservation au gouvernement

3.1 Environnement des politiques gouvernementales

Le gouvernement du Canada (GDC) utilise de plus en plus les technologies de l'information dans sa prestation de services aux Canadiens de même que dans le suivi des affaires courantes, qui nécessite une information accessible et réutilisable au fil du temps malgré les avancées technologiques. Afin d'assurer une gestion efficace et efficiente de l'information gouvernementale au cours de son cycle de vie, le GDC a ratifié la « Politique sur la gestion de l'information gouvernementale » en mai 2003. Cette politique précise l'orientation des institutions, agences et ministères gouvernementaux dans la création, l'utilisation, la gérance et la préservation de façon globale et stratégique. La politique s'applique à toutes les institutions énumérées dans les annexes I, I.1 et II de la *Loi sur la gestion des finances publiques* (LGFP). La principale prémisse de la politique va en faveur de l'emploi du numérique dans les futurs dossiers gouvernementaux.

La politique préconise que les institutions :

- s'assurent de mettre en place les structures de gouvernance et de responsabilité pour la gestion coordonnée et efficiente de l'information dont elles disposent, en faveur d'un processus décisionnel éclairé et d'une prestation des services et des programmes efficaces.
- Fournissent l'infrastructure pour la gestion efficace et rentable de l'information, quel que soit le médium ou le format, afin d'assurer l'authenticité et l'intégrité aussi longtemps que la loi, les arrêtés ministériels et toute autre loi et politique le prescrit.
- Gèrent l'information afin d'en faciliter l'accès universel et d'en optimiser le partage et la réutilisation selon les obligations légales et stratégiques.
- Étayent les processus décisionnels au fil de l'évolution des politiques, des programmes et de la prestation de services.
- Préservent l'information de valeur historique pour le gouvernement du Canada ou pour les Canadiens.
- Établissent une méthode coordonnée et exhaustive pour décrire l'information qui leur est propre.

- Entretiennent une (des) structure(s) de classification complètes et à jour, comprenant des métadonnées.

L'autorité qui veillera à la réalisation des objectifs de la politique sera exercée par l'entremise du Conseil du Trésor et de Bibliothèque et Archives Canada. Ces institutions seront responsables de maintenir une entente globale sur l'état des pratiques de gestion de l'information et des mécanismes de contrôle appropriés dans l'ensemble du gouvernement. Elles travailleront avec d'autres institutions gouvernementales dans le but de résoudre les problèmes de gestion de l'information et mèneront les initiatives mises en place pour améliorer la gestion de l'information pour tout le gouvernement. La *Loi sur les Archives nationale du Canada* et la *Loi sur la Bibliothèque nationale* ont été harmonisées en une seule et même structure légale pour former Bibliothèque et Archives Canada le 21 mai 2004.

Voici d'autres lois fédérales qui ont des répercussions sur la préservation des données géospatiales : la *Loi sur le droit d'auteur*, la *Loi sur l'accès à l'information*, la *Loi sur la preuve au Canada*, la *Loi sur la Protection des renseignements personnels et les documents électroniques*, la *Loi sur la protection des renseignements personnels* et la *Loi sur la statistique*.

En ce qui concerne les données géospatiales au fédéral, le Comité mixte des organismes intéressés à la géomatique (CMOIG) est un comité supérieur créé dans le but d'assurer la coordination des activités de géomatique du gouvernement du Canada. Le CMOIG joue un rôle prépondérant dans l'élaboration et l'appui des politiques liées à la préservation et à la gestion de l'information géospatiale. Le programme GéoConnexions représente une autre initiative majeure financée par le gouvernement du Canada pour mettre sur pied l'Infrastructure canadienne de données géospatiales (ICDG) dans le but d'harmoniser les bases de données canadiennes et d'en ouvrir l'accès par Internet. Par le biais de partenariats entre les gouvernements fédéral, provinciaux et locaux, le secteur privé et le milieu universitaire, le programme GéoConnexions promouvoit l'utilisation de normes et de protocoles pour faciliter l'accès aux données géospatiales. Les activités de GéoConnexions sont orientées selon les recommandations émises par le Conseil canadien de géomatique (COCG), un comité consultatif fédéral provincial sur la géomatique, et l'Association canadienne des entreprises de géomatique (ACEG). Le Noyau consultatif sur les politiques de GéoConnexions s'emploie à réviser et à harmoniser les politiques actuelles de manière à créer un cadre stratégique favorable au partage et à la distribution des données. Le Noyau sur les politiques a formulé plusieurs recommandations incitant le gouvernement à changer ses politiques de tarifications et à harmoniser ses politiques de distribution et de licence d'utilisation.

Au niveau des provinces et des territoires, il existe un fondement législatif pour préserver et archiver l'information électronique. Bien que toutes les provinces et territoires aient établi un cadre législatif pour les archives, dont certains appuyés par une politique de gestion de l'information pour l'information numérique, il semble qu'à ce jour (jusqu'en 2003), nul n'acquière de l'information numérique.

Quant aux municipalités, il est difficile, en l'absence de recherches approfondies sur le sujet, d'établir quelles politiques et procédures sont en place présentement. On soupçonne cependant qu'il existe très peu d'activités liées à la préservation des données géospatiales.

3. 2 Milieu de la recherche

La plupart des recherches qui portent sur la préservation de l'information numérique se fait dans d'autres domaines que celui de la technologie des SIG. La gestion des documents informatiques constitue l'un des défis majeurs dans le milieu des bibliothèques et des archives et un grand nombre de projets et programmes sont en cours. Bien que d'importantes recherches aient été entreprises, aucune solution définitive n'a encore été trouvée. Tout développement dans ce milieu profitera à la communauté de la géomatique, car les problèmes sont fondamentalement les mêmes. Voici les principaux programmes en cours :

- Le *Cedars Digital Preservation Project* est une initiative du Consortium of University Research Libraries du Royaume-Uni et de l'Irlande. Ce projet a publié une série de manuels sur les méthodes techniques pour la préservation des données numériques et leur accès.
- Créée en 2001 pour assurer la préservation des ressources numériques du Royaume-Uni, la *Digital Preservation Coalition* s'emploie à sauvegarder la mémoire numérique mondiale et tout le savoir qui s'y rattache. Elle compte parmi ses membres les plus influents la British Library, le Public Record Office, le Consortium of University Research Libraries, le Joint Information Systems Committee of the Higher et les Further Education Funding Councils (JISC).
- *The Electronic Resource Preservation and Access NETWORK Project (ERPANET)* est une initiative financée par l'Union Européenne qui a produit des guides de pratiques exemplaires, du matériel d'atelier et des rapports sur la préservation numérique du patrimoine culturel et d'éléments scientifiques.
- *InterPARES* est une initiative multinationale de recherche en coopération visant à déterminer les exigences archivistiques afin d'assurer l'authenticité des différents types d'enregistrement électroniques; établir les principes et pratiques pouvant s'appliquer à la préservation effective des documents informatiques et concevoir les cadres des politiques et normes de préservation.
- Géré par la National Library of Australia, *PADI (Preserving Access to Digital Information)* est un vaste portail dédié à la préservation des ressources et activités numériques à l'échelle internationale. PADI vise à faciliter la mise au point de stratégies et de directives pour la préservation de l'accès à l'information numérique.

Mis à part l'étude de Bleakley (2002) qui expose plusieurs problèmes liés à la préservation à long terme des données géospatiales et un rapport de Zaslavsky (2001) portant sur le problème particulier de l'archivage des données spatiales, il existe bien peu d'études spécifiques à la préservation des données géospatiales.

En dépit de l'absence de solutions définitives, on observe dans la littérature plusieurs thèmes pouvant s'appliquer à toute activité de préservation d'information numérique. Par exemple, un article sur les pratiques exemplaires dans les musées (Au Yeung 2004) résume les principaux points liés à la préservation numérique : la préservation doit prendre place au

moment de la création; la préservation est mieux servie par une approche distribuée où la coopération, l'adoption de normes et l'interopérabilité sont essentielles; la préservation des métadonnées est cruciale; et finalement, les activités de préservation doivent s'inspirer du modèle des systèmes OAIS [Open Archival Information Systems].

Il existe malheureusement bien peu d'études sur les coûts de la préservation et sur le type de support que les organisations devraient accorder aux activités de préservation. L'une des rares études sur le sujet a publié récemment un rapport sur les aspects économiques associés à la préservation à long terme des ressources numériques, notamment les mesures incitatives en faveur de la préservation (Lavoie 2003).

3.3 Milieu institutionnel

Toutes les activités d'une organisation ou institution qui touchent la collecte, la gestion et la préservation des données de même que leur accès sont motivées par la nécessité de prendre des décisions d'affaires et de livrer des produits et services basés sur des données fiables et précises. Bon nombre d'organisations ne conservent les données et l'information que le temps de leur valeur courante, cessant de les maintenir de façon adéquate une fois l'utilité immédiate passée. La valeur opérationnelle des données et de l'information d'une organisation connaît diverses phases, augmentant ou diminuant au fil du temps.

Afin de maintenir la précision et la valeur à long terme des données, les organisations doivent élaborer des plans de gestion et adopter des stratégies de préservation basées sur un modèle de gestion de cycle de vie de l'information. La gestion de l'information (GI) vise à offrir l'accès à une *information* qui a été créée et gérée au sein d'un cadre de gestion de l'information qui en assure la fiabilité, l'intégrité et l'authenticité au fil du temps. Comme le constate Au Yeung (2004), « un consensus général existe aussi sur l'archivage ou la préservation d'information numérique en tant qu'activité continue sous la forme de services des activités gérées... ou comme méthode de gestion du cycle de vie ».

L'information numérique est par nature fragile et éphémère et deviendra vite obsolète si on ne spécifie pas dès le départ le contexte dans lequel elle a été créée et utilisée pour ensuite l'intégrer à l'environnement qui assurera sa préservation dans le temps. La préservation constitue un des multiples aspects dont doivent se préoccuper les organisations détentrices de données géospatiales dans les processus de gestion du cycle de vie de l'information. Cependant, elle représente un enjeu majeur, tout spécialement dans la gestion des ressources numériques lisibles par les ordinateurs. La préservation vise à assurer la maintenance et la protection du corpus d'information de sorte qu'il soit accessible aux générations actuelles et futures.

4. Problèmes de préservation des données

4.1 Désuétude de la technologie

Le défi de la gestion et de la préservation des données numériques comprend la mise au point d'une stratégie de préservation efficace qui libère les données des formats de fichier propriétaires qui sont liés à un logiciel ou à un équipement spécifique. La création de

procédures de sauvegarde de bases de données reposant sur l'utilisation d'un logiciel de restauration de système d'opération ne peut être considérée comme une stratégie de préservation à long terme même si cette méthode peut combler les besoins opérationnels à court terme. De plus, une stratégie de préservation doit prendre en compte le caractère éphémère du support physique sur lequel les données sont placées pour des besoins de stockage à court, moyen et long termes.

4.1.1 Représentation et formats des données

Actuellement, les activités de collecte et de gestion de données au Canada sont dictées par les besoins des organisations ou institutions particulières. Il en résulte un manque de cohérence dans l'utilisation de structures de données homogènes entre les diverses organisations, en particulier à l'échelle nationale. Les formats les plus fréquemment utilisés pour stocker et échanger les données géospatiales sont les fichiers de couverture thématique d'ESRI, les fichiers de transfert (E00), les fichiers de forme (shp) et le format ArcSDE. Les autres formats reconnus sont les fichiers AutoDesk DXF, CARIS NTX et ASC; les fichiers PIX (PCI); GeoTIFF; TIFF; JPEG; COGIF; GIF; de même que les logiciels d'encodage XML/GML. Il est intéressant de noter que la majorité de ces formats de données et d'échange reposent sur des normes de l'industrie plutôt que sur des normes nationales ou ISO dûment adoptées et mises en application.

Plusieurs dérivés du langage XML (Extensible Markup Language), un métalangage qui permet de créer des étiquettes descriptives pour différents objets numériques, paraissent prometteurs pour le futur. Par exemple, le langage de balisage géographique (GML), un dialecte du langage XML, a été mis au point pour la manipulation de l'information géographique. Le GML a été conçu pour répondre à une multitude d'objectifs, dont certains dépassent la portée du langage XML. Le langage GML fournit les règles de l'encodage géospatial tant pour le transport que pour le stockage des données et ces règles sont particulièrement adaptées pour l'environnement des SIG. Suffisamment extensible pour reconnaître une grande variété de fonctions et de tâches géospatiales, le format permet d'encoder efficacement la géométrie géospatiale, de séparer le contenu spatial du contenu non spatial et de définir un ensemble de paramètres pour les objets géographiques qui rend possible l'interopérabilité des données entre des applications conçues séparément.

L'utilisation de plateformes diverses provenant de fabricants différents ont également des répercussions sur la possibilité d'échange de données géospatiales entre les organisations.

4.1. 2 Technologies de stockage

Le support physique sur lequel les données sont stockées doit être soigneusement pris en compte. On voit sur le marché un grand nombre de dispositifs de mémoire à disques ou à bandes et des systèmes de gestion de données utilisant une variété de supports de données optiques, métalliques ou de polyuréthane. Malheureusement, plusieurs de ces systèmes Intran/Extran sont de nature propriétaire, ce qui complique l'échange de données. On peut avoir les meilleures intentions en sauvegardant les données géospatiales dans les formats logiques standard, mais à moins d'étendre les pratiques de normes à l'environnement physique, il en résultera assurément des données inutilisables.

Dans la plupart des environnements autonomes, le cycle de vie du support physique dépasse habituellement celle de l'appareil qui a servi à copier les données. Bien que le support de données puisse théoriquement durer des centaines d'années (p. ex., bande et disques) la vie des appareils de lecture et d'écriture utilisés pour copier et restaurer les données varie plutôt de trois à cinq ans. Au cours des trente dernières années, on a vu nombre d'exemples de ce type de dépréciation dans les domaines de la géomatique et de l'archivage. Pour prévenir cette obsolescence, il suffit de contrôler la manipulation et le stockage du support physique dans lequel les données sont placées et de mettre en place des procédures adéquates de mise à jour et de migration des données. Plusieurs organisations en ont fait l'expérience, les coûts pour maintenir des conditions adéquates de manipulation et de stockage sont bien inférieurs à ceux du remplacement des données ou de la recréation des données perdues.

4.1. 3 Solutions technologiques à l'obsolescence

La communauté internationale de l'archivage propose tout un ensemble de solutions pour contrer l'obsolescence technologique. Cependant, il semble qu'aucune solution n'existe pour surmonter tous les défis d'archivage associés à la préservation de l'information numérique. En fait, la plupart des institutions et les ministères gouvernementaux qui ont la responsabilité d'archives utilisent une combinaison de solutions évolutives. Cette tendance devrait prévaloir du moins dans un avenir prévisible.

Toutes les stratégies de préservation doivent soupeser les coûts et avantages de la préservation de chaque élément et de la définition de règles des pertes acceptables si ces stratégies s'avèrent financièrement restrictives. Les stratégies suivantes semblent être les méthodes les plus populaires pour neutraliser les effets de l'interdépendance des technologies et des données :

- Préservation de la technologie : une démarche qui maintient les vieilles technologies matérielles et logicielles pour faciliter l'accès aux versions antérieures des formats de stockage logiques et physiques.
- Émulation de la technologie : une technique qui permet aux logiciels et aux équipements actuels de reproduire la fonctionnalité ainsi que l'apparence et la sensation des objets et applications numériques historiques.
- Migration de l'information : une technique qui permet de transférer l'information numérique d'un format de stockage logique à un autre, de façon à utiliser l'information à partir des nouvelles générations d'ordinateurs et de logiciels.
- Encapsulation de l'information : une technique utilisée pour regrouper des objets numériques et des métadonnées qui contiennent l'information nécessaire pour y accéder de nouveau éventuellement.

Le tableau ci-dessous illustre les avantages et désavantages liées à chacune des méthodes de préservation.

TECHNIQUE	AVANTAGES	DÉSAVANTAGES	PÉRIODE DE PRÉSERVATION
Préservation de	L'information est sauvegardée	Le maintien de la technologie	Les solutions à court

la technologie	dans son format original.	originale est à la fois difficile et coûteux. L'information n'est accessible qu'au moyen de l'équipement et du logiciel originaux.	terme sont impossibles à mettre en application à long terme.
Émulation de la technologie	Préserve la fonctionnalité originale d'une application de même que son apparence et la sensation qui lui sont propres	Exige l'accès au code de la source originale pour accéder à l'application et reproduire la fonctionnalité dans un nouvel environnement technologique. La reproduction d'une fonctionnalité « émulée » est coûteuse et difficile à réaliser.	Durée de court à moyen terme
Migration de l'information	Le logiciel d'application original n'a pas besoin d'être maintenu. Il y a peu de risque de perte d'information.	Il y a un risque d'altérer les fichiers au fil du temps, ce qui entraînerait probablement la perte de l'intégrité des données. La méthodologie est un processus continu et dispendieux au fil du temps. Des données administratives doivent être mises au point afin de documenter les processus de migration.	Durée de court à moyen terme
Encapsulation de l'information	Préserve tant les fichiers que l'information sur les fichiers. Ne change pas le format original du fichier.	Un temps considérable doit être investi dans la mise au point de formats et de systèmes d'encapsulation.	Solution à long terme.

D'un point de vue opérationnel, toutes les organisations seront appelées à migrer leurs bases de données afin de diminuer le risque de perte d'information. La plupart des agences ou ministères responsables d'archives appuient la migration de l'information comme la stratégie de préservation la plus convaincante. Dans le but de réduire les risques d'obsolescence les activités de migration de l'information devraient être intégrées dans les opérations quotidiennes d'une organisation et reposer sur des normes et pratiques de stockage de l'information logiques et collectivement acceptées afin d'en assurer le succès.

4.2 Gestion temporelle des données gouvernementales

La gestion temporelle des données géospatiales comporte de nombreux défis. Plusieurs méthodologies peuvent s'appliquer dont la création de versions, d'éditions ou d'instantanés de bases de données. La méthodologie choisie dépendra des besoins juridiques et opérationnels de l'organisation de produire et de maintenir des produits ou objets géospatiaux spécifiques.

Les créateurs et les utilisateurs de données géospatiales veulent habituellement l'information la plus à jour et exacte possible. C'est pourquoi, les créateurs de données remplacent régulièrement les produits réalisés, perdant ainsi à l'occasion la trace de l'évolution des objets géographiques. Il arrive parfois que les créateurs et les utilisateurs s'intéressent à l'analyse temporelle des changements dans les entités géographiques afin de comprendre les répercussions des décisions politiques sur la distribution géographique actuelle de ces entités. Ces différentes perspectives de recherche requièrent des modes de gestion de données différentes.

Un de ces modes de gestion consiste à utiliser le modèle de base de données par version pour symboliser les caractéristiques associées à un segment de données géographique à un moment particulier. Au fil du temps, une base de données par version représente une « collection d'états » qui sont des instantanés discrets de chaque entité géographique inscrite dans la base de données. La conception de la base de données permet à une organisation de gérer des représentations spatiales *à-un-moment-donné* multiples pour différents types de données géographiques. La conception de la base de données permet aussi d'explorer au fil du temps des séries de relations spatiales complexes. La base de données par version a l'avantage de permettre de recréer les caractéristiques spatiales des entités géographiques à des moments spécifiques dans le temps. Sur le plan historique, cette méthode permet de reproduire des éditions sur demande.

Une autre méthode consiste à utiliser le modèle de base de données instantané. Ce modèle est utilisé pour représenter des entités géographiques multiples à des moments donnés spécifiques. Les instantanés de base de données sont créés à des intervalles réguliers et peuvent être considérés comme une édition ou une image de l'état de la base de données à un moment précis. Chaque instantané est généralement utilisé comme une image publique ou officielle de la base de données jusqu'à ce qu'elle soit remplacée par une autre. Au fil du temps, les instantanés statiques peuvent servir à analyser les changements dans la configuration spatiale des entités géographiques.

Un grand nombre d'entreprises de SIG ont créé une diversité de solutions pour aider les organisations dans la gestion des données géospatiales et favoriser l'archivage, la préservation et l'accès. La communauté des archives participe à l'effort international pour développer le modèle de référence pour un système ouvert d'information d'archivage appelé OAIS (Model Open Archival Information System) qui définit les composantes fonctionnelles et techniques des systèmes utilisés pour gérer et préserver l'information numérique au fil du temps (Consultative Committee for Space Data Systems, 2002). Conçu pour répondre aux besoins de gestion des données de la communauté spatiale, le modèle OAIS peut aussi s'appliquer à la gestion temporelle des données géospatiales.

La National Archives and Records Administration (NARA) des États-Unis et le San Diego Supercomputer Center (SDSC) tentent présentement d'adapter le modèle de référence OAIS dans une application pratique d'une architecture de gestion de l'information reposant sur une pratique de préservation durable des objets (Thibodeau, 2001). Le projet vise à constituer des archives durables pour appuyer la collecte, le stockage d'archives, la découverte d'information et la préservation des collections numériques. Les résultats de cette recherche, combinés à la recherche présentement menée par l'équipe de l'InterPARES (International Research on Permanent Authentic Records Electronic Systems) pour l'« Atlas cybercartographique de l'Antarctique », pourrait s'avérer une solution pour la gestion et la préservation des objets de l'information géospatiale.

4.3 Documentation et métadonnées

Les métadonnées jouent un rôle central dans la stratégie de préservation à long terme. Aucune organisation ne peut réussir à préserver les données géospatiales sans une documentation adéquate. Les métadonnées, qui sont des données sur les données, fournissent l'information nécessaire pour trouver et échanger l'information numérique entre

les producteurs et les utilisateurs de données géospatiales. Cette information est reconnue comme l'un des principaux éléments de toute stratégie de préservation des données géospatiales.

Dans un environnement numérique, les métadonnées doivent être suffisamment précises pour survivre aux changements de mode d'enregistrement des diverses générations d'ordinateurs et de logiciels, permettre la reconstruction des processus de décision, et fournir une liste des modifications apportées tout au long du cycle de vie d'un fichier. Il est nécessaire de créer dans la communauté de la géomatique une infrastructure de métadonnées qui complète les activités de préservation face à une demande toujours croissante de partage et de distribution de produits géospatiaux fiables entre les créateurs et les usagers à un coût modique.

La norme de contenu pour les métadonnées géospatiales numériques (Content Standard for Digital Geospatial Metadata) du FGDC constitue la norme de métadonnées la plus souvent en usage au fédéral. Cette norme est en voie d'être remplacée par la nouvelle norme internationale ISO 19115. L'harmonisation et la mise au point d'un profil nord-américain sont présentement en cours.

5. Conclusions et Recommandations

Malgré le grand nombre de recherches présentement en cours, il semble qu'il n'existe pour l'instant aucune solution unique aux défis que nous réserve la préservation de l'information numérique. Comme l'a noté Au Yeung (2004) « ...la véritable solution de la préservation numérique pourrait être non pas technologique mais plutôt politique ». Il est cependant clair qu'une combinaison de stratégies doit être mises en place dans une structure de gestion de l'information efficace. La mise en œuvre dépendra des besoins des organisations en matière de législation, de politiques et de gestion de l'information et de leur aptitude à investir dans les activités de préservation à long terme.

Les recommandations qui guideront l'action et les recherches liées à l'archivage et à la préservation à long terme des données géospatiales peuvent se regrouper sous : 1) les activités mises en place par les institutions et 2) les activités nécessitant des politiques, des normes et des lignes directrices nationales qui pourraient être mises au point en collaboration entre des organisations telles que Bibliothèque et Archives Canada, le Comité mixte des organismes intéressés à la géomatique avec des conseils nationaux tels que (CMOIG) et l'Office des normes générales du Canada.

5.1 Recommandations pour une action au niveau des institutions

Pour préserver à long terme et sauvegarder de précieuses ressources, les producteurs de données doivent d'abord adopter un mode de gestion du cycle de vie de l'information de façon à s'assurer une gestion proactive de leurs données depuis leur création jusqu'à leur élimination.

Recommandation #1 : Les organisations doivent définir et mettre en application des politiques et des pratiques pour la création, l'utilisation, le maintien, la dissémination, la préservation et l'élimination des données géospatiales.

Méthode suggérée : Une étude de rentabilité doit être menée sur la création de données géospatiales de base afin de définir la valeur durable de l'information pour les activités d'une organisation et de déterminer les règles de gestion de cycle de vie qui peuvent s'appliquer à sa rétention et à son éventuelle élimination. Beagrie et Jones (2002) ont mis au point un modèle d'arbre de décision qui peut servir de guide dans les activités de collecte des organisations. Le Groupe de travail sur les données historiques (1998) du Federal Geographic Data Committee (FGDC) a aussi publié une feuille de route afin d'aider les organisations à répondre à des questions telles que :

- Existe-t-il une exigence légale ou politique pour la conservation des données géospatiales?
- Les données sont-elles critiques pour la prestation des services de gestion centrales de (des) institution(s)?
- Quels sont les coûts opérationnels pour la collecte et la création des données? En cas de perte de données, quels seraient les coûts pour les recréer?
- Les données sont-elles uniques?
- Quelle est la qualité des données?
- Quelle est la valeur potentielle des données et quelle est la probabilité que certaines d'entre elles soient réutilisées au sein de l'institution ou ailleurs?

Une réponse positive à l'une ou plusieurs de ces questions pourrait automatiquement qualifier ces données pour la préservation à long terme.

Recommandation # 2 : Les organisations doivent établir des centres de responsabilité qui autorisent des individus à définir et à appliquer les principes de la gestion de l'information nécessaire pour assurer l'intégrité des bases de données géospatiales d'une organisation.

Méthode suggérée : Le modèle de dépôt représente un moyen de faciliter la gestion des données au nom des créateurs et des utilisateurs et d'assurer la continuité de livraison de l'infrastructure de données géospatiales. Dans une telle approche les producteurs de données devront :

- Créer une stratégie de gestion de données approuvée au sein de l'organisation et un plan qui détermine les ressources de données et établit les normes quant aux pratiques de collecte, de gestion et de préservation des données;
- Nommer une personne ou créer un comité qui agira comme dépositaire des données d'une organisation et qui travaillera au sein de l'organisation et veillera à ce que les politiques et procédures pour la gestion des données géospatiales soient définies, comprises et mises en œuvre;
- Créer un feuillet d'information sur l'évaluation des données pour aider à définir les données valables requérant la conservation à long terme;
- Déterminer la valeur de l'objet des données au moment de leur création à partir des besoins opérationnels et de préservation;

- Fournir des métadonnées détaillées basées sur des normes de métadonnées internationales et nationales approuvées. Les modèles de métadonnées et de données devraient être stockés avec les données;
- Inventorier les bases de données numériques et évaluer leur valeur pour la conservation et la préservation à long terme et définir un calendrier pour la migration et la modernisation des supports de données;
- Normaliser les formats de données avant de transmettre cette information au sein de l'organisation;
- Adopter des normes industrielles telles que ISO ou des solutions de source ouverte qui affranchissent les données des formats propriétaires.

5.2 Recommandations d'action à l'échelle nationale

Recommandation # 1 : Les problèmes de préservation des données géospatiales s'inscrivent dans le cadre d'une politique sur l'information et d'une stratégie de gestion des données nationales. Travaillant en partenariat avec les communautés des bibliothèques et des archives, les producteurs de données gouvernementales doivent normaliser et adopter des politiques et des pratiques organisationnelles pour guider la création, l'utilisation, la rétention, la dissémination, la préservation et l'élimination des données géospatiales de manière à en assurer l'authenticité et l'intégrité aussi longtemps que le prévoient la loi, les arrêtés ministériels et autres lois et politiques.

Méthode suggérée : Le Canada produit des données géospatiales de grande qualité. Ses politiques et méthodes pour assurer leur préservation à long terme restent toutefois à développer.

Les agences et organismes nationaux tels que le Comité mixte des organismes intéressés à la géomatique (CMOIG), le Conseil canadien de la géomatique (COCG), les agences fédérales comme Bibliothèque et Archives Canada et les conseils nationaux tels que l'Office des normes générales du Canada devraient travailler de concert à l'élaboration des stratégies suivantes :

- Promouvoir l'utilisation des normes de métadonnées approuvées au sein de la communauté de géomatique, au moyen de publicité, d'ateliers et de création d'outils
- Promouvoir l'adoption de normes « non propriétaires » permettant la création et l'échange de données géospatiales telles que les spécifications du langage GML et celles de l'Open Geospatial Consortium
- Promouvoir l'utilisation des directives et des pratiques exemplaires établies par Bibliothèque et Archives Canada pour le stockage logique et physique de l'information numérique
- Établir des listes de contrôle et autres outils pratiques pour la sélection des données à préserver

- Définir et avaliser le concept de « conservateur de données ». Les organisations productrices de données devront définir leurs rôles et responsabilités de conservation pour la gestion des produits de données géospatiales collectifs. Le concept de conservation devrait comprendre les responsabilités liées à l'imputabilité, à la fiabilité, à l'authenticité, aux normes de métadonnées et de documentation, aux formats des fichiers et aux questions d'accès
- Rechercher et recommander les modèles opérationnels viables pour une préservation qui favorise le partage de la responsabilité et des coûts associés aux activités de préservation
- Travailler avec les communautés de la recherche, telles que InterPares, participant à des projets novateurs dans le domaine de la préservation, tout spécialement ceux liés aux données géospatiales, de façon à intégrer les nouveaux développements dans les politiques et directives de gestion de l'information.

Recommandation # 3 : Le forum conjoint (fédéral/provincial/territorial) du Conseil canadien de la géomatique devrait créer un groupe de travail et inviter les parties intéressées du milieu universitaire, du secteur privé et d'autres agences fédérales, provinciales ou territoriales à collaborer afin de développer les domaines prioritaires stratégiques définis précédemment.

Références

1. ANZLIC the Spatial Information Council. (2001). *ANZLIC Metadata Guidelines: Core metadata elements for geographic data in Australia and New Zealand, Version 2*. Février, 95 p.
2. Beagrie, Neil and Jones, Maggie. (2001). *The Preservation Management of Digital Material - A Handbook*. London: The British Library.
3. Bleakly, Denise R. (2002). *Long-Term Spatial Data Preservation and Archiving: What are the Issues?*, Sand Report, SAND 2002-0107. Albuquerque, New Mexico : Sandia National Laboratories.
4. Brown, David L. & Comeau, Mike. (1999). "Restoration of the Canada Land Data System". *Association des Cartothèques et Archives cartographiques du Canada*, n° 106, Automne, p. 42-52.
5. Brown, David L. & Comeau, Mike. (2001). *Restoration of the Canada Land Inventory*. Paper presented at the Social Science History Association Conference, Historical Geography Network, Archives in the Digital Age, Chicago, Illinois, November 15-18.
6. Brown, Richard & Brown, David L. (2001). *Information Management in the Government of Canada: Improving Service Today and Planning for Tomorrow*. Rapport non publié. Ottawa: Archives nationales du Canada.
7. Chen, Su-Shing. (2003). *Digital Library Preservation: Organizational Commitment, Archival Stability, and Technology Continuity*, Department of CISE, Gainesville, FL : University of Florida.
8. Consultative Committee for Space Data Systems (CCSDS). (2002). *Reference Model for an Open Archival Information System (OAIS)*. Document CCSDS 650.0-B-1, Blue Book, Issue 1.
9. Council on Library and Information Resources and Library of Congress, (2002). *Building a National Strategy for Preservation: Issues in Digital Media Archiving*, Washington, DC : CLIR. Extrait (mai 2002) de : <http://www.clir.org/pubs/abstract/pub106abst.html>.
10. Data and Information Systems Panel, Canadian Global Change Program. (1996). *Data Policy and Barriers to Data Access Across Canada: Issues for Global Change Research*. Extrait (février 2002) de : http://www.globalcentres.org/cgcp/english/html_documents/publications/data/toc.htm.
11. FGDC Historical Data Working Group (1998). *Managing historical geospatial data records: a guide for federal agencies* Extrait (février 2002) du site : <http://www.fgdc.gov/nara/hdwgfsht.html>
12. Hedstrom, Margaret & Montgomery, Sheon. (1998). *Digital Preservation Needs and Requirements in RLG Member Institutions*. Mountain View, CA. : Research Libraries Group. Extrait (avril 2002) de : <http://www.rlg.org/preserv/digpres.html>

13. Hedstrom, Margaret. (1995). "Digital Preservation: A Time Bomb for Digital Data". In *U.S. Geological Survey: EROS Data Center Collections Policy Manual*. Index révisé en 2002 par Denny Thurman et Diane Krell. Extrait (août 2002) de : <http://www.uky.edu/~kiernan/DL/hedstrom.html>.
14. Hickling Arthurs Low (HAL) Corporation. (2001). *Geomatics Sector Human Resources Study*. CD-ROM. Ottawa : Conseil canadien des Arpenteurs-Géomètres, de anadian Council of Land Surveyors Association canadienne des sciences géomatiques et Association canadienne des entreprises de géomatique.
15. Lee, Kyong-Ho, et. al. (2002). The State of the Art and Practice in Digital Preservation, *Journal of Research of the National Institute of Standards and Technology*, vol. 107, n° 1.
16. Lavoie, Brian F. (2003). The Incentives to Preserve Digital Materials: Roles, Scenarios, and Economic Decision-Making. White paper published electronically by OCLC Research. Extrait (juin 2004) de : <http://www.oclc.org/research/projects/digipres/incentives-dp.pdf>
17. Bibliothèque et Archives Canada. (2004). *Lignes directrices relatives aux types de fichiers informatiques, aux formats de transfert et aux normes d'information*. Version 1.0, février. Ottawa : Bibliothèque et Archives Canada.
18. McDonald, John. (2000). *Gestion de l'information au sein du gouvernement du Canada - Analyse de la situation*. Rapport non publié pour le bureau du Dirigeant principal de l'information, le Secrétariat du Conseil du Trésor et l'Archiviste national du Canada. Ottawa : Archives nationales du Canada.
19. Moore, R.W., & Mersky, A. (2002). *Persistent Archive Basic Components – Draft 1.0*, Persistent Archive Research Group, Global Grid Forum. Extrait (août 2002) de : <http://www.sdsc.edu/NARA/Publications/Web/PA-cap.doc>.
20. Archives nationales du Canada. (2003). *Modèle d'évaluation des capacité en matière de gestion de l'information : méthodologie et directives*. Ottawa : Archives nationales du Canada.
21. Consultation sur les archives nationales de données. (2002). *Infrastructure pour l'archivage et l'accessibilité des données de recherche du Canada*. Rapport non publié présenté par le Groupe de travail CAND au Conseil de recherches en sciences humaines du Canada et à l'archiviste national du Canada. Ottawa : le Groupe de travail CAND.
22. Open GIS Consortium. (2001). *Geography Markup Language*. Open GIS Implementation specification. OGC Document Number: 01-029. Extrait (juillet 2002) de : <http://www.opengis.net/gml/01-029/GML2.pdf>
23. Parer, Dagmar, éd. (2001). *Archival Legislation for Commonwealth Countries*. Cairns, Australia : Association of Commonwealth Archivists and Records Managers (ACARM).
24. Ross, Seamus. (2000). *Changing Trains at Wigan: Digital Preservation and the Future of Scholarship*. London: National Preservation Office. Extrait (août 2002) de : <http://www.bl.uk/services/preservation/occpaper.pdf>.

25. Statistique Canada. (2000). *Enquête de l'an 2000 sur les industries de services : prospection, arpentage et cartographie*, Systèmes de classification des industries de l'Amérique du Nord (SCIAN) codes 54136 et 54137. Ottawa : Statistique Canada.
26. Thibodeau, Kenneth. (2001). "Building the Archives of the Future: Advances in Preserving Electronic Records at the National Archives and Records Administration". *D-Lib Magazine*, vol. 7, n°. 2, février.
27. Secrétariat du Conseil du Trésor du Canada. (2001). *IM Ready Report*. Version 2, Rapport non publié. Ottawa : Division du Dirigeant de l'information.
28. Tristram, Claire. (2002). "Data Extinction". *Technology Review*, Octobre 2002, p. 37-42.
29. Zaslavsky, Ilya. (2001). *Archiving Spatial Data: Research Issues*. San Diego Supercomputer Center Technical Report, TR-2001-6. Extrait (juin 2002) de : <http://www.sdsc.edu/TR/TR-2001-06.doc.pdf>