ÉTUDE

*La gestion et l'archivage des sites Web institutionnels**

Bessem Khouaja
Carol Couture

*L*orsqu'on pense Web, on évoque la plus grande masse documentaire jamais produite et diffusée auparavant. Selon des statistiques datant de la fin de l'année 2000, le nombre des pages Web est estimé à plus de quatre milliards de pages Web en accès libre. Quand au Web invisible ou le Web profond[1] (*deep Web*), il contiendrait plus de 550 milliards de documents. Ces pages sont écrites en 220 langues différentes et constituent un volume se comparant à plus de 50 fois le volume de la collection de la Bibliothèque du Congrès. Plus de sept millions de pages Web s'ajoutent quotidiennement en même temps que d'autres disparaissent. Notons que la durée de vie moyenne d'une page Web est d'environ 44 jours seulement (Lyman et Varian 2000 cités par Lyman 2002). Dans cette masse grandissante, nous avons choisi de nous intéresser aux sites Web institutionnels.

Plusieurs avantages expliquent le recours des institutions publiques à l'utilisation du Web. Citons à titre d'exemple l'amélioration de la communication avec le public, la réduction du temps de réponse, le perfectionnement de la qualité de service, la diminution des coûts, la réduction de la bureaucratie et de la paperasse, l'élimination des files d'attente, une meilleure productivité, des meilleures commodités pour les clients et pour les agents. Tous ces avantages expliquent que le Web institutionnel gagne de plus en plus de terrain et prend de l'ampleur de jour en jour. Le Brown University Taubman Center for Public Policy, lors d'une récente enquête, a analysé 1197 sites gouvernementaux pour 198 pays (juin-juillet 2002). Les résultats démontrent que ce type de sites offre, dans une proportion de 12 %, des services entièrement exécutables en ligne. De plus, au delà des trois quarts des sites visités fournissent l'accès à des publications et 83 % ont des hyperliens vers des bases de données. L'enquête démontre également que 14 % des sites proposent des guichets uniques de services (*one-stop services «portal»*) ou possèdent des hyperliens menant à un portail gouvernemental (TCPP 2002).

* Ce texte est la version revue, corrigée et adaptée d'un travail universitaire réalisé dans le cadre du cours BLT 6059 - Archivistique et information documentaire – au cours de la session d'automne 2002 à l'École de bibliothéconomie et des sciences de l'information de l'Université de Montréal.

L'offre actuelle de plusieurs services dans maintes administrations nous encourage, en tant que bénéficiaires de ces services, à utiliser le Web comme médium possible pour mener à bien un certain nombre d'opérations : effectuer des transactions bancaires, remplir une demande d'emploi, remplir et envoyer sa déclaration annuelle d'impôts, s'inscrire à un programme d'études ou faire des demandes de prêts et bourses auprès du service de l'aide financière aux études. Tous ces services s'opèrent sans intermédiaire, ni papier dans un laps de temps relativement court.

Même si les sites Web institutionnels peuvent être considérés comme des pages qui font le lien entre un usager et un service offert à travers une base de données, il reste que ces sites génèrent aussi des documents qui constituent des traces de l'activité de plusieurs institutions gouvernementales, susceptibles de servir de preuves. Cela implique qu'au Canada par exemple ces documents doivent impérativement faire partie du système central de la gestion des archives et, par la même occasion, doivent être traités dès leur création et ce, jusqu'à leur disposition finale conformément d'une part, aux prescriptions de la loi sur les Archives (tant au niveau provincial qu'au niveau fédéral) et, d'autre part, à la politique du Conseil du Trésor du Canada relative à ce sujet (GTRII 1999a).

Malgré cette réalité, l'idée dominante est que l'utilité du Web cesse lorsque l'information recherchée est trouvée ou lorsque le service demandé est accompli. Cette vision quant à l'utilité du Web se trouve aussi dans les milieux institutionnels, car la plupart des intervenants agissant dans la création et la gestion des documents issus du Web (le producteur du document, le webmestre et le gestionnaire des documents) pensent que le Web n'est qu'une vitrine ou un pont entre les institutions et les clients. De ce fait, ces intervenants ont la certitude que les documents véhiculés par les sites Web de leurs institutions génèrent couramment de l'information d'actualité à caractère éphémère. Ainsi, ce qui existe sur le Web ne représente que des copies de travail (l'argument avancé par les gestionnaires étant que les originaux sont stockés sur support papier ou sur des supports de conservation électronique), ce qui rend légitime la destruction de ces sites Web.

Les questions qui se posent à ce niveau sont donc multiples : Perdons-nous de l'information en détruisant des pages Web ? Faut-il éliminer des traces significatives de l'évolution d'un organisme, des services qu'il donne, des informations qu'il génère ? Ne détruit-on pas à jamais des informations sur la structure des données, leurs créateurs, leur date d'apparition, les besoins pour lesquels ces documents sont mis en ligne ? Est-il justifié de déraciner les pages Web en leur enlevant dynamisme, liens hypertextuels et caractère multimédia, pour les enfermer dans un support qui les rend statiques et linéaires ?

Jusqu'ici, nous avons essayé d'exposer les raisons qui font que la gestion des sites Web institutionnels doit être perçue comme une nécessité. Le gouvernement fédéral, qui s'est penché sur la question de la gestion des réseaux Internet/Intranet en 1999, interpelle les organismes gouvernementaux pour « établir de saines pratiques de gestion de leurs sites Web » (GTRII 1999a) et les inclure dans leur stratégie globale de gestion de l'information. Aussi, faut-il bien identifier les principaux intervenants dans la gestion d'un site Web et définir leurs rôles respectifs afin de délimiter les responsabilités. Enfin, il faut prendre la décision de considérer le site Web soit comme une « entité » qui doit,

en aucun cas, être divisée, soit comme un ensemble de fichiers électroniques qui doivent être traités séparément comme les autres fichiers électroniques.

L'objectif premier de notre réflexion est de voir s'il est envisageable de gérer et d'archiver les documents issus des pages Web en appliquant des méthodes comparables à celles utilisées pour la gestion des documents administratifs sur support papier. En tenant compte de toutes les considérations évoquées précédemment et en prenant connaissance de certaines des pratiques actuelles exercées dans quelques services d'archives, nous nous questionnons sur maints aspects ayant trait à la gestion et à l'archivage des sites Web institutionnels : Est-ce qu'on doit traiter les pages Web comme des documents d'archives ? Est-ce qu'on doit traiter les sites Web comme de simples fichiers électroniques linéaires ? Étant donné le caractère dynamique d'un site Web, quelles sont les techniques à utiliser pour résoudre ce problème si l'on décide de conserver le site hors ligne ? Peut-on parler de cycle de vie pour un site Web : Web courant, Web intermédiaire et Web définitif ? Considérant le caractère éphémère des documents en ligne, a-t-on vraiment besoin d'une période intermédiaire ?

Pour répondre aux questions précitées, nous allons effectuer un tour d'horizon de quelques expériences réalisées dans ce domaine, en plus de nous inspirer des avancées effectuées par quelques bibliothèques nationales. En deuxième lieu, nous allons examiner les modes de gestion des sites Web et dégager des façons de faire. Nous préciserons alors quelques concepts clés dans cette recherche, nous distinguerons une typologie des sites Web institutionnels et nous présenterons les différents intervenants dans la gestion du Web. En dernière partie, nous formulerons quelques suggestions quant aux pratiques de gestion et d'archivage des sites Web en nous basant sur les fondements archivistiques et nous ferons quelques recommandations pour une meilleure gestion. Par la suite, nous proposerons des méthodes permettant de développer une politique de gestion des sites Web qui devrait être une partie intégrante de la politique générale de gestion des documents d'un organisme.

LA PRATIQUE EN MATIÈRE DE GESTION ET D'ARCHIVAGE DES SITES WEB

Qu'en ont dit les auteurs ?

La gestion et l'archivage des sites Web ont abondamment été étudiés dans la littérature. C'est le fait, entre autres des bibliothécaires qui ont attaché le grelot lorsqu'ils ont réalisé que plusieurs publications électroniques et une multitude d'informations sur le Web étaient en train de se perdre sans laisser de traces. Les archivistes de leur côté ont réfléchi sur la gestion et l'archivage des documents électroniques, des bases de données ainsi que du courrier électronique. La majorité des recherches s'intéresse au problème de l'intégrité et de l'authenticité des documents électroniques. Les questions de la préservation et de la conservation du médium électronique retiennent aussi l'attention des spécialistes de la discipline.

Le Web a été traité dans la littérature comme étant un moyen efficace de diffusion de l'information numérique (Lemay 1998). Une enquête a été réalisée pour étudier l'ampleur et le degré de l'utilisation des sites Web par les services d'archives pour la diffusion (Hamel 1998). Quelques spécialistes ont exploré les possibilités qu'offrent

le Web et les réseaux pour la collecte des métadonnées qui servent à la description (Bouchard et Piché 1997 ; Frost 2001) et qui facilitent, en outre, la tâche de l'archivage (Grange 1998, 2000 ; Motz 1998). Bertrand (1998), pour sa part, s'est questionné sur la nécessité de l'archivage avec la présence d'Internet, sur les documents qui devraient être archivés et sur la manière de le faire.

La gestion du Web institutionnel devenait par ailleurs une nécessité avec l'apparition d'une nouvelle notion de « gouvernance électronique » ou « e-gouvernance » qui se résume en l'utilisation des technologies de l'information pour mieux servir le citoyen et lui fournir une meilleure prestation de services. L'objectif est de moderniser le pouvoir public pour faire face au mécontentement des citoyens. Cette notion est bien développée en Europe. À la deuxième rencontre de DLM (Données lisibles par machine) Forum Europeen, qui s'est tenue à Bruxelles en 1999, a été abordé le thème : « Le citoyen européen et l'information électronique : La mémoire de la Société d'Information ». Durant cette réunion, on s'est intéressé à la gestion des documents, aux archives numériques, au développement des standards et à la récupération et la gestion des documents électroniques provenant des sites Web ou échangés par la voie des réseaux dans l'environnement des administrations (Harries 1999).

La gouvernance électronique est une notion en développement en Amérique du nord. Quelques études sont menées pour mesurer le besoin et l'impact de cette notion sur la société. Nous pouvons citer l'enquête effectuée par le Brown University Taubman Center for Public Policy qui a étudié l'ampleur de l'utilisation du Web par les institutions publiques aux États-Unis et au Canada dans le cadre de la « e-gouvernance ». Michel Audet (1999) a évalué l'influence des technologies de l'information au Québec. Liette D'Amours (2001) a traité de l'impact et des bénéfices de ce type de gouvernance sur la société. Toutes ces études ainsi que d'autres sont regroupées sur le site http://www.cefrio.qc.ca du Centre francophone d'information des organisations (CEFRIO).

Les études marquantes dans le domaine de la gestion et de l'archivage du Web ont surtout été réalisées par des archivistes américains qui relatent leurs expériences dans le domaine de la gestion et de l'archivage des sites Web institutionnels. Charles Dollar a fait une étude pour le Smithsonian Institution Archives à Washington, D.C dans laquelle il propose une stratégie pour la gestion des ressources Web (Dollar 2001). À travers son expérience, Dollar évoque quelques modèles pour la sélection, la gestion et la préservation des pages Web de l'Institution. Par ailleurs, l'Archives and records management division du South Carolina Department of Archives and History a traité la question de la gestion des archives publiques issues du Web dans sa revue *Public records information leaflet*. Cette étude avait pour objectif d'aider les départements de l'État de la Caroline du Sud à mettre en place une politique pour la gestion des sites Web. De leur côté, McClure et Sprehe (1998) ont essayé de mettre en place un guide pour la gestion des documents administratifs des sites Web pour l'État fédéral et les départements d'États américains. Cette étude vise à développer une stratégie de gestion et de préservation des archives du Web en donnant des indications sur le repérage de ces documents, les principaux intervenants dans ce processus, les méthodes de gestion selon le type de site Web (dynamique ou stable) et selon sa vulnérabilité. Le même sujet a été traité, entre autres, par LeFurgy (2001) qui considère que la conservation hors ligne des documents du Web diminue la valeur des sites Web. Il pense aussi que la

conservation d'une partie d'un site Web institutionnel n'est pas une solution, puisqu'on n'est pas certain de conserver la partie qui a la plus grande valeur. Sur cette question, la Library of Congress a publié une étude pour la mise en place d'une stratégie nationale pour la préservation des documents publiés en format électronique. Le chapitre 3 de cette étude traite des documents issus du Web (Lyman 2002).

Par ailleurs, la consultation de deux guides s'impose. Le premier, *Archiving Web resources : Guidelines for keeping records of Web-based activity in the Commonwealth Government* a été élaboré par les National Archives of Australia (2001). Le deuxième, du Public Records Office of U.K., a pour titre *Managing Web resources. Management of electronic records on websites and Intranet : An ERM toolki* (2001). Ces deux organismes ont été les premiers à mettre en œuvre un guide pour la gestion des sites Web institutionnels dans leur pays. Les deux ouvrages répondent fort bien aux besoins des services d'archives et permettent de prendre en main la gestion des sites Web et d'harmoniser toutes les politiques existantes à l'échelle du pays. Au Canada, le Groupe de travail sur les réseaux Internet/Intranet a élaboré un guide qui répond « . . . à l'orientation pangouvernementale en matière de tenue des dossiers et de gestion des publications dans les installations ministérielles logées sur réseaux. » (GTRII 1999a) Ce guide traite à la fois des publications électroniques des ministères qui doivent être déposées à la Bibliothèque Nationale conformément aux dispositions de la Loi sur le dépôt légal et des documents administratifs qui doivent être gérés en vertu de la Loi sur les archives nationales du Canada.

L'intérêt que ces auteurs portent à la question de la gestion et de l'archivage des sites Web s'est manifesté suite à une sensibilisation faite auprès de quelques pays qui ont compris l'importance que prend Internet et le rôle majeur qu'il joue dans la diffusion d'information dans tous les secteurs de la connaissance. Ces pays ont alors entrepris quelques expériences dans le domaine de la gestion et la préservation de leur patrimoine sur le Net. Nous présentons dans la partie qui suit quelques initiatives archivistiques dans ce domaine.

Quelques expériences de préservation du contenu des sites Web

Les initiatives prises concernant la préservation du contenu du Web datent de 1996, c'est-à-dire trois ans après la prolifération à grande échelle du Web comme média accessible au grand public. Ces initiatives ont été basées sur trois approches qui répondent aux besoins de chaque pays concerné.

Approche maximaliste ou exhaustive

Cette approche consiste à archiver et à préserver tout le contenu du Web en s'appuyant sur l'utilisation d'un robot automatisé qui permet de capter tous les sites selon une présélection déjà établie (ex. : un pays, un domaine, une discipline). Cette méthode élimine le risque de perte de documents historiques due au jugement de valeur porté sur des documents. Elle se révèle aussi être la moins coûteuse par rapport aux autres approches. Parmi les résultats les plus connus dans cette approche nous pouvons citer :

« Wayback machine »

Le concept d'archivage du Web est apparu en 1996 avec Brewster Khale, spécialiste des ordinateurs parallèles. B. Khale, l'un des pionniers en la matière, a anticipé

l'importance de garder des traces des sites Web. Pour ce faire, Khale a commencé à stocker les pages Web à intervalle régulier et à héberger des sites Web dans un espace disque personnel. Son objectif était de ressusciter la Bibliothèque d'Alexandrie et de bâtir une bibliothèque virtuelle en conservant toute la connaissance accumulée sur le Web. Pour la réalisation de ce rêve, Khale s'est donné comme mission de conserver des pages Web. La masse documentaire accumulée en 2001 était estimée à dix tétrabits, l'équivalent de cinq fois la collection de la Library of Congress qui compte vingt millions d'ouvrages. Grâce au site *www.archives.org*, nous pouvons retracer aujourd'hui des pages Web qui ont été mises hors ligne depuis longtemps. Ainsi, peut-on voir les premières versions d'un site Web qui demeure encore actif. Les spécialistes estiment que ce projet est un nouveau moyen pour contourner le message d'erreur « 404 not found ». Ce projet de taille a été baptisé « The Wayback Machine » ou encore « la machine qui remonte le temps » (Roumieux 2001 ; Morin 2002b ; Notess 2002).

L'expérience suédoise

Le projet *Kulturarw3* a été mis en place en 1996 par la Royal Library. La Suède a adopté une approche globale qui part du principe que la sélection des sites Web à conserver est une opération pénible, coûteuse et subjective et que l'accessibilité des supports de stockage et de conservation des documents électroniques et leur performance ne cessent d'augmenter. Cette approche veut recueillir tout le contenu du domaine *« .se »* du Web, toutes les pages hébergées sur des serveurs suédois, tous les sites des producteurs suédois hébergés à l'extérieur ainsi que toutes les pages Web en rapport avec la Suède. Cette cueillette est entièrement automatisée et se fait par l'intermédiaire d'un robot qui capte continuellement les pages Web depuis 1997. Jusqu'en 2001, ce robot a effectué 7 fois le tour du Web. Le dernier a duré 8 mois et a traité 50 400 000 pages, soit 2 395 gigaoctets. Plusieurs supports de stockage ont été testés pour la sauvegarde des données. Actuellement, l'objectif de ce projet vise à garder tous les fichiers capturés en ligne et tente de rassembler ces pages en fonction d'une arborescence qui dépend de leur adresse URL (Perrin 2001 ; NDL 2002 ; BNQ 2001).

À ces expériences s'ajoutent celles de la Finlande, de l'Autriche et du Danemark.

Approche minimaliste ou sélective

L'approche sélective se base sur une sélection minutieuse des pages Web à archiver selon des critères prédéfinis. Elle est pratiquée par des spécialistes qui jugent la pertinence des ressources Web à conserver. Cette méthode est celle qui génère le moins de bruit. Elle est cependant la plus coûteuse en temps et en argent des trois méthodes de sélection. Cette approche est adoptée par :

L'Australie

Nous lui devons les premières démarches pour la mise en place de procédures pour l'archivage des publications diffusées en réseau en 1996. Ce projet a été baptisé *PANDORA*, pour *Preserving and Accessing Networked Documentary Resources of Australia*. La stratégie adoptée était très sélective : elle consistait à ne conserver que les publications sous format électronique qui n'avaient pas d'équivalent papier et qui étaient jugées d'importance nationale. Le but visé était de créer une collection représentative

plutôt qu'exhaustive (BNC 2001). La collection numérique actuelle compte 2000 sites Web, soit 320 gigaoctets ou 11 millions de fichiers dont la majorité des sites Web qui ont été créés pour les jeux olympiques de Sydney et qui n'existent plus sur le Web. Chaque année, 500 sites sont ajoutés en moyenne (NDL 2002). Un dispositif électronique prend régulièrement des extraits de quelques sites après avoir effectué un balayage pour les stocker par la suite sur des serveurs dédiés à cet effet. L'indexation s'effectue grâce à un robot *Harvesting*. Les publications ont un identificateur *PURL*[2] (*Persistent Universal Resource Locator*) qui précise la version du document dans chacune de ses éditions. Notons que beaucoup d'efforts sont déployés pour préserver cette collection. D'ailleurs, une coopération internationale au sein de groupes de travail, tels le NEDLIB (Networked European Deposit Library) ou le Cedars (CURL Exemplars in Digital Archives) supportés par CURL (Consortium of University Research Libraries), veille à la préservation de cette collection. Après six ans d'expérience, 127 fichiers ont subi avec succès une migration vers d'autres supports. (Perrin 2001 ; NDL 2002).

Les États-Unis

La Library of Congress s'est donné comme mission de colliger et de préserver les documents du Web pour en permettre l'accès aux membres du Congrès américain et à tous les Américains en général. Le projet *MINERVA* (*Mapping the Internet Electronic Resources Virtual Archive*) date de l'année 2000. Un comité, le Web-Archivist, s'occupe de la sélection et de la diffusion des sites Web en se basant sur plusieurs critères. Le site Web sélectionné sera téléchargé en utilisant un système de cueillette à l'aide d'un robot Web, *Web-Crawler*. Une copie *snapshot* [3] sera conservée aux archives. D'autres copies seront effectuées à intervalle régulier. Toutes ces copies seront cataloguées en utilisant le standard *CORC* (*Cooperative Online Resource Catalog*), géré par l'OCLC (Online Computer Library Corporation). La recherche de ces sites peut s'effectuer par titre, par sujet ou par adresse URL (NDL 2002 ; BNQ 2001).

Le Canada

Dès 1997, la Bibliothèque nationale du Canada (BNC) a pris en charge l'archivage des documents nationaux officiels. Une politique relative aux publications électroniques diffusées en réseau a été adoptée en octobre 1998 (BNC 1998). Dès lors, la BNC s'est chargée d'identifier les publications électroniques sur le réseau, de les colliger et d'y donner accès par la voie son site Web. La recherche de ces publications électroniques s'effectue à travers le catalogue AMICUS. Pour ce qui est des documents administratifs, un groupe de travail sur les réseaux Internet/Intranets a été créé pour proposer des procédures de gestion des documents issus des réseaux. Ce groupe fait partie du Forum sur la gestion de l'information, coprésidé par le Secrétariat du Conseil du Trésor et les Archives nationales du Canada (ANC). Deux documents de travail cités précédemment présentent ce projet (GTRII 1999). Le système semble être lourd à mettre en œuvre et, jusqu'à maintenant, il n'y a pas vraiment de procédure pour l'archivage du Web. Il importe toutefois de noter que la Politique de communication du gouvernement du Canada (voir le point 18 de la Politique qui concerne l'Internet et la communication électronique), entrée en vigueur le 1er avril 2002, fait référence à l'enregistrement et à l'archivage pour la conservation à long terme de l'information diffusée sur les sites Web (Secrétariat du Conseil du Trésor du Canada 2002).

Le Québec

Selon Mme Danielle Léger, responsable du dépôt légal à la Bibliothèque nationale du Québec (BNQ), 1700 titres de publications électroniques provenant d'une vingtaine de ministères et organismes gouvernementaux québécois sont déposés à la BNQ. L'objectif est d'aller chercher les publications de l'ensemble des institutions gouvernementales, soit 30 000 titres produits par près de 120 institutions. D'après la BNQ, toutes les publications mises en ligne par les institutions québécoises et qui s'apparentent à une monographie ou à une publication en série doivent être déposées. Par ailleurs, la collection gouvernementale, sujette à ce type de dépôt, constitue 17% de la collection patrimoniale.

Approche intermédiaire ou intégrative[4]

Cette approche, la plus récente des trois, est encore en expérimentation et est principalement adoptée par la France. Elle consiste en une collecte automatique doublée d'un suivi individuel et manuel pour assurer à la fois un repérage large, un échantillonnage représentatif du Web d'aujourd'hui et un suivi précis de certains sites inaccessibles aux robots, ce que l'on appelle le *Web profond* (Castier 2002). Selon Julien Masanes, conservateur en charge de ces projets expérimentaux à la Bibliothèque nationale de France (BnF), la philosophie qui fonde cette approche veut que les robots ne puissent pas atteindre les zones grises du Web ou encore le Web profond et ce, à cause de raisons techniques, de restrictions de type site payant ou restreint pour une catégorie d'utilisateurs bien définie. Ce type de contenu, en général plus riche que le reste du Web, est difficile à archiver en utilisant une approche globale. En contrepartie, une sélection minimaliste, purement manuelle demande plus de temps et risque de ne pas suivre l'évolution de la masse d'information qui ne cesse d'augmenter (Masanes 2002b). La solution selon Masanes est d'assurer une collecte automatique, de négocier les droits de dépôt légal des sites Web avec les institutions mères pour s'assurer de la collecte des contenus des sites Web dits invisibles (Web profond). Toutefois, une vérification de la valeur de l'information capturée automatiquement par les robots ne peut assurer la validité de l'information archivée.

L'éparpillement de ces expériences, leur importance, leurs différences, ainsi que l'intérêt que le Web acquiert de jour en jour ont poussé la communauté internationale à unir ses efforts en échangeant des expériences à travers des manifestations scientifiques aux niveaux international, régional et local afin de débattre ce sujet et d'identifier des solutions pratiques.

Les manifestations scientifiques

L'archivage des sites Web est un sujet qui gagne du terrain et qui occupe plusieurs chercheurs et professionnels à travers le monde. Au cours des trois dernières années, dans sept manifestations scientifiques, le sujet de l'archivage du Web a été traité.

Le symposium international *Archives Online : moving into the digital era*, organisé par les Archives nationales néerlandaises à Den-Haag le 17 avril 2003, a traité la question de la préservation des nouvelles archives produites par les différents services gouvernementaux et a proposé des solutions. Le 2ème atelier de la ECDL[5] (European Conference on Digital Libraries) tenu à Rome le 19 septembre 2002 a eu pour objectif de créer une interaction entre les bibliothécaires, les archivistes, les chercheurs du monde

académique et les chercheurs industriels intéressés aux expériences de l'archivage du Web, afin d'établir des méthodes efficaces et de développer des solutions pour améliorer ces opérations. L'atelier a présenté la situation actuelle de la recherche et les différentes expériences réalisées dans ce domaine.

La 68e Conférence annuelle de l'IFLA, tenue à Glasgow en Écosse entre le 18 et 24 août 2002, a traité, entre autres, du rôle que doit jouer la bibliothèque à l'ère de la démocratie électronique (Paré 2002). Durant cette conférence, un cadre juridique pour les publications issues du Web a été présenté par Alenka Kavcic-Colic (2002). Par ailleurs, le DPC (Digital Preservation Coalition) Forum, qui a eu lieu le 25 mars 2002 à Londres, s'est intéressé à la question de la gestion et de l'archivage des documents et des archives en ligne (Redfern 2002). Aussi, l'International Symposium on Web Archiving a regroupé à Tokyo le 30 janvier 2002 des spécialistes de la Library of Congress, de la Bibliothèque nationale de l'Australie, de la Royal Library of Danemark et de la National Diet Library of Japan. Chaque intervenant a exposé un projet sur la conservation des documents issus du Web tel que mené dans son pays (NDL 2002). En outre, la 5th European Conference on Research and Advanced Technology for Digital Libraries qui s'est tenue à Darmstadt en Allemagne le 8 septembre 2001, a eu pour thème : *What's next for Digital Deposit Libraries? Preservation Online Content for Future Generation*. Des spécialistes des bibliothèques nationales de l'Allemagne, de la France, du Danemark, de l'Australie, de la Finlande, de la Suède, de l'Autriche et des États-Unis (Library of Congress) ont présenté chacun l'expérience de leur pays dans le domaine du dépôt légal des documents électroniques issus des réseaux. Enfin, la conférence tenue à Copenhagen les 18-19 juin 2001 par la Royal Library a traité le thème *Preserving the present for the future. Strategies for the Internet*. L'objectif de cette conférence était de dresser de nouvelles stratégies et méthodologies pour préserver le contenu du Web pour les générations futures.

Tous ces écrits, ces expériences, ces approches et ces manifestations scientifiques nous ont permis de constater l'intérêt que portent les spécialistes de différents domaines de recherche quant à l'importance de gérer et d'archiver le contenu des sites Web et d'assurer pour la postérité la sauvegarde de ce riche patrimoine. Mais, qu'en sera-t-il alors de la bonne gestion de ces ressources particulières?

LA GESTION ET L'ARCHIVAGE DES SITES WEB INSTITUTIONNELS

Nombreuses sont les expériences de gestion et d'archivage des sites Web dans le monde. Ces expériences prennent soit la forme d'initiative personnelle telle que celle de Brewster Khale avec sa Wayback Machine, soit celle d'une initiative prise par des associations à but non lucratif ou par quelques universités. Parmi ces initiatives, citons : a) AIR (Association of Internet Researchers), une association académique qui se consacre à l'avancement du champ des études interdisciplinaires sur Internet, b) ERPANET (Electronic Resource Preservation and Access Network) qui représente un projet instauré par la Commission Européenne afin d'améliorer les pratiques d'archivage et de développer les qualifications dans le secteur de la préservation numérique de l'héritage culturel et la recherche scientifique, c) les efforts de la Wellcome Library, un organisme anglais spécialisé dans le domaine de la médecine et le JISC (Joint Information Systems

Committee*)*, comité du Ministère de l'enseignement supérieur en Grande-Bretagne qui assiste les organismes à vocation scientifique dans la promotion et l'utilisation des technologies de l'information et de la communication pour soutenir la recherche scientifique et l'enseignement. Wellcome Library et JISC ont entrepris une étude de faisabilité sur l'archivage des ressources Internet s'intéressant au domaine médical. Les objectifs de cette étude étaient de cerner ce qui existe déjà comme pratiques dans l'archivage du Web et de recommander les pratiques jugées intéressantes pour leurs deux institutions. Pour ce qui est des universités, l'Université de Glasgow et l'Université du Michigan mènent actuellement des projets en ce sens. Par ailleurs, on peut aussi ajouter l'intérêt qui se manifeste auprès de certains comités de rédaction de revues électroniques dans le domaine scientifique au sujet de la gestion de leurs sites et la constitution d'une rubrique « archives » qui donne accès au contenu de la revue.

D'autres initiatives sont à considérer comme celles entreprises au niveau national (voir la section Quelques expériences de préservation du contenu des sites Web) ou au niveau régional telles que NEDLIB (Networked European Deposit Library) en Europe et NWA (The Nordic Web Archive) dans les pays scandinaves. Au plan international, signalons les efforts de normalisation du W3C (World Wide Web Consortium).

Au Canada, par exemple, l'utilisation d'Internet et des communications électroniques est fortement encouragée par le gouvernement. D'ailleurs, dans sa politique de communication, entrée en vigueur le 1er avril 2002, le Secrétariat du Conseil du Trésor du Canada considère le réseau mondial comme :

> un important outil pour fournir de l'information et des services au public. Internet facilite la communication interactive et bidirectionnelle ainsi que la rétroaction. Il offre des possibilités de joindre les Canadiens peu importe où ils habitent et de leur fournir des services personnalisés (Secrétariat du Conseil du Trésor du Canada 2002).

Ainsi, chaque institution doit :

> Veiller à ce que l'information diffusée sur les sites Web soit enregistrée et archivée pour la conservation à long terme et préserver la mémoire institutionnelle avant de faire des modifications ou des mises à jour des sites. Des processus uniformes étant mis en place en temps opportun à cet effet, de concert avec les gestionnaires des fonds de renseignements d'une institution (Secrétariat du Conseil du Trésor du Canada 2002).

Afin de respecter les dispositions de la Loi sur les archives et d'appliquer ces directives, chaque institution est appelée à intégrer la gestion et l'archivage de son site Web dans sa politique générale de gestion de l'information (PGGI). Tout comme c'est le cas pour les archives électroniques, les documents issus du site Web de l'institution doivent subir un traitement qui ne diffère pas des autres documents produits et reçus par l'organisme, et ce, en tenant compte de leurs spécificités.

Typologie des sites Web institutionnels

Pour gérer les pages Web institutionnelles, il est impératif de commencer par étudier les différentes formes et les types de sites Web qui existent actuellement sur le Net. Pour répondre davantage aux attentes, la catégorisation des sites Web s'avère un outil essentiel pour le choix et la sélection de la stratégie la mieux adaptée.

Les sites Web statiques ou informatifs

Un site peut avoir la forme d'un ensemble de fichiers qui sont regroupés dans un ou plusieurs dossiers. Ces dossiers ou fichiers sont hébergés sur un même serveur et sont associés par des hyperliens. Ce genre de site est recommandé pour les institutions de petite envergure comme c'est le cas de la grande majorité des sites des compagnies actuellement disponibles sur le Web. Ces sites servent principalement à faire la promotion de l'institution, à informer les clients au sujet des produits et services offerts, et à afficher des renseignements généraux en établissant des liens entre les fichiers. Un site statique ou informatif n'est pas nécessairement volumineux ou complexe. Il comporte des informations élémentaires. Dans certains cas, une seule page contenant quelques paragraphes et images suffit pour la consultation.

Les sites Web interactifs

Ce type de sites comporte une information plus détaillée et actualisée et il forme une source quasi inépuisable d'informations sur plusieurs facettes de l'institution. Les sites peuvent aussi présenter une structure statique, mais ils doivent obligatoirement offrir un service interactif entre les clients et l'institution mère. Par exemple, ce service peut prendre la forme de formulaires à remplir par les utilisateurs du site Web pour recueillir des données ou pour avoir une rétroaction directe sur les renseignements diffusés.

Les sites Web dynamiques

Ce sont des sites qui changent souvent de contenu tout en étant reliés avec un élément extérieur comme une base de données constamment mise à jour. La principale caractéristique de ces sites est que « les pages telles qu'elles sont visualisées côté client, n'existent pas sur le serveur. Elle sont générées à la volée lorsque la requête est envoyée aux serveurs. » (Masanes 2002a) Les internautes peuvent aussi bien les interroger que les alimenter. Dans la majorité des cas, les documents constituant le site sont des éléments d'une base de données où chaque document a un identifiant unique, souvent visualisé avec l'adresse URL du site pour que l'utilisateur puisse retracer cette information en tapant directement l'adresse complète du document (NAA 2001).

Les sites dynamiquement générés

Plusieurs sites récents sont bâtis grâce à la génération instantanée « on the fly » du contenu de leurs pages. Cette technique consiste à séparer le contenu d'un document Web, sa structuration et sa présentation. L'affichage des pages Web dépend alors de la combinaison des données et de la feuille de style utilisée. Dans ce cas, on peut considérer qu'il n'y a pas une seule représentation de la page Web, mais plusieurs selon la feuille de style associée. Chaque utilisateur peut ainsi visualiser le contenu du site selon ses préférences et ses besoins (NAA 2001).

Gestion et traitement des dossiers issus du Web

Identification des responsables de la gestion d'un site Web

Selon le Groupe de travail sur les réseaux Internet / Intranets (GTRII 1998a) et les Archives nationales d'Australie (NAA 2001), les quatre principaux intervenants dans la gestion d'un site Web d'une institution sont :

Les créateurs des documents : Ils constituent un élément primordial dans le processus de diffusion de l'information. Leur rôle consiste à donner leur autorisation pour diffuser le contenu de leurs documents de travail.

Le gestionnaire du contenu : C'est le responsable du contenu véhiculé par le site Web de l'institution. Il doit sélectionner les documents qui donnent l'information la plus exhaustive et la plus complète sur les activités de l'institution. Il doit donc toujours avoir en main les versions les plus complètes et les plus récentes. Ce gestionnaire peut intervenir pour s'assurer que chaque document affiché sur le Web comporte des métadonnées.

L'administrateur du site Web (le Webmestre) : C'est lui qui s'occupe des aspects informatiques et techniques du site. Il est responsable de la conception, de l'hébergement et de la mise à jour des pages Web dans l'institution. Il peut intervenir pour favoriser un format d'affichage qui peut différer de celui de l'original sans pour autant en altérer le contenu.

Le gestionnaire des documents : Son rôle est primordial du fait qu'il doit s'assurer que chaque document diffusé sur le Web reçoit le même traitement que les autres documents de l'organisme. Il doit aussi veiller à ce que les documents soient identifiés, aient une cote de classification et qu'une règle de conservation en encadre le cycle de vie.

Évaluation des risques d'un site Web institutionnel

S'il est un facteur essentiel dont il faut tenir compte dans la mise en place d'une politique générale de gestion de l'information (PGGI), c'est l'évaluation des facteurs de risque que le site Web présente pour l'organisme. Selon les Archives nationales d'Australie, les principaux facteurs à prendre en considération sont (NAA 2001) :

La visibilité et l'image de marque de l'institution : Le site Web est une image qui reflète les activités de l'institution, la nature de ses transactions et le degré de confiance qui s'établit entre l'institution et ses clients. Le taux d'utilisation des sites Web institutionnels varie donc d'un site à un autre selon le champ d'activité concerné.

L'objectif du site Web : Un site peut répondre à divers besoins de l'institution. Son objectif varie selon le but à atteindre. L'objectif peut se limiter à un simple partage d'information ou à une diffusion plus efficace. Il peut aussi aider au partage des tâches du travail (notamment dans le cas d'un extranet) ou aux transactions commerciales. Par ailleurs, le but du site Web peut se limiter à des fins de communication avec les clients pour vérifier l'efficacité des services et recueillir leurs points de vue sur des sujets particuliers ou sur un nouveau produit ou service.

La complexité du site Web : Comme nous l'avons vu auparavant, cette complexité dépend de la structure du Web, qu'il soit statique, interactif, dynamique ou mixte.

La fréquence et la régularité de la mise à jour : Elle variera selon la taille de l'institution, ses activités, la population desservie : la fréquence des mises à jour diffère d'un site à l'autre.

Toutes ces considérations ainsi que celles développées dans la section Typologie des sites Web institutionnels doivent aussi être prises en compte dans le choix d'une stratégie adéquate de gestion d'un site Web institutionnel et pour que la PGGI réponde davantage aux attentes et aux résultats escomptés.

Création des documents d'archives issus du Web

McClure et Sprehe (1998) déterminent deux façons différentes de traiter les documents d'un site Web institutionnel.

Selon la première méthode, qui est la plus conventionnelle, un administrateur crée lui-même ses fichiers électroniques à l'aide des logiciels bureautiques. Ces fichiers subissent des changements purement techniques de la part du Webmestre sans que celui-ci change le contenu afin que ces fichiers soient diffusés sur le Web. Cette méthode appelée « purpose-prepared Website posting » présente l'avantage de garder le document original tel qu'il a été créé par son auteur. Nous pouvons donc considérer le fichier sur le Web comme une copie de diffusion qui peut être détruite immédiatement après usage.

Le deuxième cas de figure, plus compliqué et de plus en plus répandu actuellement, se présente quand le document est directement créé sur le Web. Citons comme exemples les rapports administratifs, les formulaires d'inscription, les commentaires et les suggestions des citoyens envoyés sur les sites des ministères, la soumission des requêtes spécifiques, la soumission des demandes d'emploi. De plus, le Web véhicule des documents qui peuvent être considérés comme des originaux, notamment ceux qui nécessitent une mise à jour fréquente.

> A state agency reported that the official version of a document was still the printed paper version. In fact, the most current, up-to-date, and reliable version was the one on the Website. The paper version was out-to-date by the time it appeared. The Website version was what all state employees used as their authoritative source (McClure et Sprehe 1998).

Donc, pour certains documents, il est souvent plus logique de se fier à la seule version électronique disponible sur le Web. Ces documents électroniques doivent faire partie de la PGGI et subir un traitement adéquat qui tient compte de leurs particularités que nous abordons dans la partie « Capture » et préservation des sites Web institutionnels.

Identification des documents d'archives

Le Web peut véhiculer de nombreux types de documents autres que des archives. La première opération à entreprendre sera de distinguer les documents d'archives de ceux qui n'en sont pas et qui ne présentent pas une valeur administrative, légale ou financière. Citons à titre d'exemple les publications que les institutions diffusent. Ces publications doivent impérativement être déposées à la Bibliothèque nationale conformément à la loi du dépôt légal.

Par exemple, en Australie, dans de tels cas, quand une version papier existe, les Archives nationales obligent les ministères à en déposer une copie imprimée à la Bibliothèque. Si la copie électronique diffère de celle qui existe sur support papier ou s'il n'y a tout simplement pas d'équivalent sur papier, le ministère est obligé par la loi de déposer une copie de cette publication électronique dans le système *PANDORA* (NAA 2001).

Autre exemple, en Angleterre, le Public Record Office rappelle qu'il faut identifier les documents à archiver :

> It is essential to clarify which website content has the status of corporate record and which does not ... Some useful questions to ask are : is the website content a unique instance? if so, what is its importance? or If not, is the website version of business importance in its own right (although also held elsewhere) (PRO 2001, 13).

La description des documents issus du Web : les métadonnées

Les métadonnées jouent un rôle prépondérant dans la description du contenu des documents du Web et dans la description des documents électroniques hors ligne. En outre, les pages Web demandent un effort supplémentaire pour leur description vu la diversité des versions du contenu, leur fréquente mise à jour, la gestion des hyperliens, la dépendance de quelques spécificités techniques comme le langage de balisage utilisé (tels HTML, XML, XHTML), le navigateur (et la version) employé (Internet Explorer, Netscape ou autres).

L'utilisation d'un des multiples systèmes de standards pour la description des documents Web est recommandée dans la majorité des cas. Nous pouvons citer comme exemples le *CORC*, utilisé par la Library of Congress ou le Dublin Core supporté par l'OCLC pour la description des ressources électroniques sur le Web. Il y a aussi le RDF (Resource Description Framework) qui est recommandé par les membres du W3C (World Wide Web Consortium) et autres groupes intéressés pour les documents XML (W3C 1999). Le problème principal vient du fait qu'il n'y a pas actuellement unanimité pour l'utilisation d'un standard unique.

Les éléments de base à prendre en considération pour la description d'un site Web sont le titre de l'affichage, la date d'affichage, le numéro de la version d'affichage, l'auteur du document et ses coordonnées, la date de modification, les dates de retrait ou de remplacement de l'affichage, les hyperliens menant à l'affichage et la disposition finale après retrait définitif (GTRII 1999a). En plus de ces éléments, il importe d'ajouter le contexte de création, la langue du document, un court résumé du contenu, une cote de classification, la règle de conservation qui s'applique et les informations sur le balisage et les logiciels à utiliser pour la lecture.

« Capture » et préservation des sites Web institutionnels

Stratégies et modèles de « capture » des sites Web institutionnels

Le modèle de capture et de gestion des sites Web dépend pour beaucoup du type et de la complexité du site Web, du secteur d'activité dans lequel l'institution œuvre, du degré de risque et de vulnérabilité que le site représente et des choix faits par l'institution elle-même pour la récupération. Cette stratégie peut combiner plus qu'une technique de capture et doit être flexible pour s'adapter aux changements technologiques lorsque c'est nécessaire.

Dans la littérature, les techniques de « capture » les plus recommandées sont les suivantes :

La gestion du contenu du Web (*object-driven approach and event-driven approach*) : Cette technique se concentre sur la capture du contenu du site au détriment de sa forme. Elle est suggérée pour les sites qui utilisent les langages neutres de balisage (ex. le HTML) et qui ne sont pas trop complexes. L'application de cette méthode nécessite l'usage de métadonnées plus développées pour pouvoir faire le lien entre les différents objets. Cependant, cette méthode risque de faire perdre des informations comme le lien entre les objets, le contexte de création, les transactions, l'interaction avec les clients.

Les instantanés : appelés aussi les strates, mais surtout les snapshots. « C'est le fait de prendre une photographie du site Web à un moment bien déterminé. Chaque strate

représente une image du Web à un instant « t ». On navigue à l'intérieur d'une époque comme sur le Web actuel et on ajoute une fonctionnalité au navigateur pour passer d'une strate à une autre. » (Gharsallah 2001, 40) Cette technique est recommandée pour les sites Web qui ne sont pas trop interactifs ou dynamiques. L'inconvénient que présente cette technique a été souligné par les Australiens : « If snapshots are captured in the absence of other records of web-based activity, it will be impossible to reconstruct the site together with its functionality at any other point in time. » (NAA 2001, 26)

Maintien des sites Web en ligne : Cette méthode tente de combiner la diffusion de documents actifs et de documents considérés comme des archives sur le même site Web. Ce modèle requiert une gestion plus complexe du site et un investissement plus important pour assurer le maintien de ces deux types de documents ensemble sur un même serveur. L'institution peut aussi louer les services d'un serveur dédié à cet effet. Cette approche est recommandée pour les sites complexes et très développés.

Préservation des sites Web institutionnels

La garantie d'accéder à long terme aux documents du Web préoccupe aussi les spécialistes. À quoi sert-il donc, en effet, de prendre des instantanés d'un site Web si nous ne pouvons pas garantir de pouvoir retracer ce site ultérieurement et le consulter ?

Le site Web, comme c'est le cas des archives électroniques, dépend du matériel informatique et des logiciels (hardware et sofware) qui le soutiennent. Pour retracer une page Web, il faut disposer des applications permettant la lecture des fichiers (ex. Adobe Acrobate Reader, Power point, Word). La question de l'obsolescence des technologies est un autre élément qu'il faut solutionner dans l'archivage des pages Web : « Electronic materiel created under older systems becomes unreadable in the original form after relatively short periods of time. Agencies taking website snapshots for online or offline storage need to plan for technology obsolescence. » (NAA 2001, 32)

Lors des prises des instantanés, l'institution doit absolument tenir compte de l'évolution de la nature des standards (les différentes versions du langage HTML, la tendance actuelle vers XML). Elle doit favoriser l'utilisation des logiciels non propriétaires qui sont les plus répandus aujourd'hui. Avantager les technologies neutres et les formats standards échangeables constitue une solution à explorer.

Plusieurs projets en ce sens sont en cours d'expérimentation pour la préservation des pages Web. Citons le projet InterPARES[6] (International Research on Permanent Authentic Records in Electronic Systems) qui est une entreprise d'envergure internationale regroupant des spécialistes de différents secteurs comme des institutions nationales d'archives, des ingénieurs informatiques et quelques industries privées représentatives du milieu. Tous ces spécialistes œuvrent au développement de la méthode la plus adéquate pour garantir la préservation et l'authenticité des archives électroniques. Un autre projet qui mérite d'être mentionné est le NEDLIB[7] (Networked European deposit Library) qui regroupe les bibliothèques nationales de plusieurs pays d'Europe. L'objectif de ce projet est de proposer des solutions pour la gestion et la préservation des publications électroniques. Parmi les projets déjà testés, mentionnons la capture des pages Web pour l'archivage et la gestion des publications hors ligne. Des tests ont été réalisés pour mesurer la faisabilité de l'émulation en tant que méthode de garantie

d'accès à long terme. Les résultats ont été convaincants et plusieurs autres projets en cours s'intéressent aux publications en ligne.

Enfin, en plus des projets précités, il importe de signaler qu'il faut prendre en considération la mise en place des systèmes de sécurité pour maintenir un contrôle d'accès aux documents, un contrôle sur l'intégrité de l'information et encourager l'usage des adresses Internet de type PURL.

Stockage des sites Web

La spécificité des pages Web et leurs différents fichiers électroniques obligent les gestionnaires de ces documents à prendre la décision d'assurer une conservation sur des supports de stockage connus (disquette, CD-ROM, CD-R) ou d'opter pour le stockage sur un serveur dédié à la conservation. Le choix répond à un besoin spécifique : « option for offline storage include optical disk or magnetic tape. In contrast, online storage provides instantaneous access in the form of a hard drive. » (NAA 2001, 34) La décision à prendre dépend des spécificités des sites Web développés auparavant, des besoins réels de l'institution, de la quantité des données que le site génère et surtout des moyens humains, financiers et matériels disponibles.

PROPOSITIONS POUR UNE MEILLEURE GESTION DES SITES WEB INSTITUTIONNELS

L'objectif de cette partie est de présenter quelques pistes de solution et de recommander quelques pratiques concernant la gestion et l'archivage des sites Web et ce, en s'appuyant sur les notions de base en archivistique et en essayant de les adapter aux caractéristiques du Web. Nous proposons aussi quelques façons de faire pour aider le gestionnaire des documents à mener convenablement et aisément sa tâche en ce qui a trait aux sites Web institutionnels. Les solutions proposées doivent impérativement coller à la mission, aux spécificités et aux besoins réels de l'institution. Comme les Archives nationales d'Australie l'ont souligné : « There is no generic solution for creating and maintaining records of web-based activity. The best option will depend on the outcome of an analysis of the particular circumstance. » (NAA 2001, 36)

Parmi les facteurs que chaque institution devrait prendre en considération, citons : la typologie de son site Web, le caractère des informations véhiculées par le site, la complexité des transactions et la fréquence de mise à jour du site. La détermination de toutes ces caractéristiques est fondamentale pour l'élaboration de la politique de gestion du site Web qui fait partie de la PGGI de l'institution.

Principes généraux de gestion des archives imprimées appliqués aux sites Web

Le cycle de vie et la théorie des trois âges

L'unique caractéristique des sites Web sur laquelle on peut se baser pour la distinction entre un site Web courant, intermédiaire ou définitif est la notion d'accessibilité ou d'activité du site. Un site Web courant est constitué de l'ensemble de ses fichiers électroniques qui sont accessibles au grand public ou pour une catégorie d'utilisateurs bien déterminée (dans le cas d'un Intranet, d'un Extranet, des sites payants ou avec

abonnement). Ce site doit être impérativement sur un serveur actif qui en assure l'accès. La notion du Web intermédiaire disparaît puisqu'elle se confond avec celle du Web courant. Donc, le stade de « semi-activité » se voit supprimé pour les documents du Web. En conséquence, aussitôt qu'un site n'est plus accessible et qu'il est retiré du serveur, nous recommandons qu'il devienne automatiquement un site définitif. Notons qu'à ce niveau, il n'existe pas d'élimination au sens archivistique du terme. À tout le moins, les doublons seraient éliminés.

La notion de fonds d'archives

En se référant à l'article de Michel Duchein (1977), il importe d'appliquer la notion de fonds aux sites Web. Selon ce principe, la délimitation des frontières d'un site Web est vitale pour sa gestion comme fonds d'archives. Une approche minimaliste de cette notion est privilégiée, surtout par les organismes de grande envergure et ce, pour responsabiliser chaque service de la gestion de son site Web. Les différentes versions d'un site Web peuvent être considérées comme un élément séparateur qui permet d'éclater le site Web en sous-fonds ou en séries pour en faciliter la gestion[8].

La gestion continue d'un site Web

Il est fortement recommandé aux gestionnaires des documents de procéder à un archivage instantané du site Web de leurs institutions comme c'est le cas pour les archives imprimées. Ils doivent déterminer une période de temps moyenne qu'ils jugent adéquate pour en faire des copies. Cette période dépend étroitement du volume du site, de son degré de vulnérabilité et de la fréquence de sa mise à jour. Chaque fois qu'il y a un changement important, le gestionnaire devrait prendre une copie des fichiers du site et les conserver pour en constituer une copie complète. Ainsi, l'ensemble des copies reflétera avec exactitude le contenu du site Web de l'institution durant une période de temps précise.

Il est déconseillé d'attendre la fin d'année pour prendre une copie puisque une grande masse d'informations risque d'être déjà retirée du serveur, ce qui peut causer un risque considérable de perte d'informations numériques pour l'institution.

La description et le classement des archives issues du Web

La description des archives issues du Web se présente comme une nécessité pour le repérage de ces fichiers d'archives surtout si l'on considère la vitesse à laquelle le Web se développe. L'utilisation des métadonnées dans les pages Web, telles que celles spécifiées dans le schéma descriptif du Dublin Core, est *fortement* recommandée. Un moteur de recherche intégré au site peut, par ailleurs, faciliter énormément la recherche dans le contenu des fichiers eux-mêmes.

Le cadre de classement de ces fichiers ne doit en aucun cas être séparé du cadre général de classement des fonds d'archives de l'institution. Une séparation au niveau du traitement intellectuel n'est pas permise dans ce cas, car les archives issues du Web constituent une partie intégrante des archives institutionnelles, tandis qu'une distinction physique est souhaitable pour la bonne conservation des fichiers nécessitant des mesures de préservation plus poussées.

Quelques bonnes pratiques en matière de gestion des sites Web

Dans cette partie, nous reprenons quelques points déjà traités dans les parties précédentes de ce texte. Aussi, nous avons sélectionné quelques pratiques recommandées par la Smithsonian pour la conception et le montage de sites Web et des pages HTML en vue d'un accès à long terme (Dollar 2001).

La garantie d'un accès immédiat et de longue durée

Lorsque nous pensons archives électroniques, nous passons souvent sous silence l'importance de préserver le matériel et le logiciel de lecture qui assureront l'exploitation de ces données plus tard. Pour contrer le problème de l'obsolescence des technologies, il serait primordial de choisir une technologie standardisée et supportée par des organismes internationaux. Citons entre autres les formats PDF (Portable Document Format) et RTF (Rich Text Format). Pour ce qui est du balisage des sites Web, nous recommandons l'utilisation de la norme XML (eXentensible Markup Language) qui permet l'indépendance des documents des infrastructures technologiques actuelles. Les bases des données, quant à elles, doivent aussi être supportées par XML pour assurer leur exploitation ultérieure, indépendamment des logiciels qui les créent. Par ailleurs, il faut éviter autant que possible l'emploi d'un moteur de recherche exclusif à un tiers qui ne soit pas sous le contrôle du webmestre de l'institution et, naturellement, ce moteur doit accompagner le site Web une fois qu'il est archivé (Dollar 2001).

L'utilisation des liens hypertextuels relatifs

Un site Web est apprécié pour son caractère dynamique et surtout pour l'utilisation qu'il permet des liens hypertextuels. Jusqu'à présent, il n'existe pas de solution pour conserver un site Web en maintenant ses liens hypertextes actifs. D'ailleurs, même certains sites en ligne connaissent le problème des liens morts ou inactifs (erreur 404). Et ce problème est accentué pour les sites archivés. Pour contrer cette difficulté, il pourrait être avantageux d'utiliser des liens hypertextes relatifs[9] plutôt qu'absolus[10] quand cela est possible. Ainsi, l'on peut maintenir actifs certains liens internes (si les noms des fichiers restent inchangés) dont la configuration des répertoires sur le serveur risquent de changer.

D'un autre côté, les liens inactifs absolus peuvent ne pas constituer une perte d'informations dans le cas où ils pointent vers des documents organiques d'autres institutions (liens externes). Ces liens externes inactifs ne risquent pas d'altérer la préservation du « fonds électronique » constitué par les fichiers du site.

L'analyse de la « vulnérabilité » du site Web institutionnel

Cette vulnérabilité touche la nature des informations véhiculées (sensibles, essentielles ou autres) et la fréquence de la mise à jour du site. L'analyse de la vulnérabilité pour un site Web institutionnel est recommandée afin de permettre un choix judicieux du modèle de gestion. D'ailleurs,

> pour chaque site Web, l'administrateur Web, le gestionnaire de contenu et l'agent de dossier doivent ensemble établir la vulnérabilité du site, en obtenant (s'il y a lieu) l'avis des conseillers juridiques et des spécialistes des affaires publiques du ministère […]. Les sites Web peuvent être classés selon un niveau de risque faible, moyen ou élevé (GRTII 1998a, 29).

Cette analyse va permettre de choisir la stratégie la mieux adaptée et la plus adéquate pour la gestion du site Web (voir la partie « Capture » et préservation des sites Web institutionnels).

Le stockage des instantanés

Nous conseillons aux praticiens d'utiliser les supports de conservation optique de la famille des disques optiques numériques comme les CD-R (Recordable), CD-WORM (Write Once Read Many) et surtout ceux qui sont dédiés à la conservation de longue durée. Actuellement, même en l'absence d'un consensus quant à la fiabilité de ces supports, il est conseillé de les utiliser car ils demeurent les mieux adaptés, en termes d'espace de stockage et de facilité de manipulation, pour conserver les sites Web en espérant que la technologie nous fournisse le support plus fiable tant attendu. Nous suggérons, en plus, de dater les copies des sites Web archivés en indiquant la période couverte, la date de réalisation de la copie et toute autre information pouvant servir à la gestion de celle-ci. Nous conseillons aussi d'effectuer de temps en temps des copies récapitulatives sur des supports fiables qui garantissent une conservation de longue durée.

Un site Web n'est pas encombrant et constitue un document très riche en contenu. Aussi, reflète-t-il bien l'évolution de l'institution durant son activité. Une conservation permanente des copies récapitulatives qui regroupent tout le contenu du site dans une période de temps bien déterminée est une stratégie recommandée pour garantir un accès à long terme aux fichiers du site. Cette solution peut être considérée comme provisoire jusqu'à ce que la technologie Web se stabilise davantage et que le progrès technologique fournisse d'autres solutions plus adéquates.

L'élaboration et le maintien à jour d'un registre du site Web institutionnel

La création d'un registre du site Web est fortement recommandée et peut être d'une utilité extrême pour suivre l'évolution, les changements, les modifications ainsi que chaque intervention qu'un site a subis. Ce registre peut être tenu par le gestionnaire des documents en coordination avec le webmestre de l'institution. Chaque fois qu'un document est placé sur le serveur de l'institution ou qu'il en est retiré, il doit être mentionné dans ledit registre qui peut prendre une forme imprimée ou électronique. Toutes révisions majeures du site Web devraient être soigneusement documentées.

Selon le *Guide de mise en œuvre de la gestion de l'information sur Internet et Intranet* (GRTII 1998a), ce registre peut comprendre les éléments suivants : le titre de l'affichage, son numéro séquentiel, le nom de l'auteur du document affiché ainsi que ses coordonnées, son contenu et son étendue, la date du premier affichage, les dates de modifications de celui-ci, son adresse complète sur le serveur, les liens que cet affichage comporte, la date de retrait ou du remplacement de celui-ci et les raisons de ce retrait, la dernière disposition de cet affichage ainsi que son sort final.

Cette pratique va permettre de suivre l'évolution du site Web à travers le temps. Aussi, va-t-elle nous aider à retracer un document à partir de sa date d'affichage jusqu'à son retrait.

La création d'un espace dédié pour le regroupement de tous les affichages du site Web institutionnel

Ce service peut être loué chez un fournisseur privé ou être géré par le webmestre si toutes les conditions sont favorables à cette opération. Elle consiste à sauvegarder

la totalité du contenu du site hors ligne dans cet espace loué (serveur informatique). Cette solution est recommandée pour les institutions dont le site présente un risque élevé de vulnérabilité.

Le serveur dédié peut être considéré comme un relais ou une copie de sécurité qui peut être mise en place au cas où tout le système tomberait en panne. D'un point de vue archivistique, ce système peut être considéré comme un dépôt où sont centralisés les sites Web courants et intermédiaires. Le gestionnaire des documents peut considérer ce relais comme un stade intermédiaire avant de prendre la décision d'archiver.

Nous le concédons, ces solutions demeurent très partielles et ne sauraient résoudre tous les problèmes rencontrés par ce nouveau média qu'est le Web. L'interrelation entre les sites, la difficulté de définir les limites de chaque site ainsi que les caractéristiques de ce nouveau média ne sont qu'une infime partie des problèmes que rencontre ce média.

CONCLUSION

Comme nous le mentionnions en introduction, nous croyons pouvoir affirmer sans grand risque de nous tromper que le Web réunit actuellement les ensembles d'informations les plus volumineux, qui ont probablement été les plus largement diffusés dans l'histoire du genre humain. En fait, le Web est un lieu inépuisable d'entreposage d'informations, un véhicule révolutionnaire de diffusion de ces informations et une fenêtre ouverte sur le monde pour nombre d'institutions qui autrement demeureraient totalement inconnues. Mais qu'en est-il des préoccupations plus terre à terre de gestion et d'organisation de toute cette information ? Plus précisément, pour peu qu'on pense à la réalité de l'utilisation qu'en font les institutions, quelles considérations doivent être prises en compte pour gérer efficacement un site Web institutionnel ? Et comment, d'un point de vue archivistique, faut-il aborder la problématique de la gestion d'un tel site ? Que conserver ? Pendant combien de temps conserver ? Comment organiser les informations ? Comment les rendre disponibles et pour qui ? Tels étaient nos principaux questionnements quand nous avons entrepris la présente réflexion.

Après une revue de littérature et un tour d'horizon d'expériences pertinentes réalisées dans plusieurs pays, nous avons proposé une typologie des sites Web institutionnels et présenté rapidement le rôle des différents intervenants dans la gestion de ce genre de site ; le tout ayant pour but de dégager des pratiques simples et efficaces de gestion et d'archivage de ces banques d'informations. Le lecteur aura compris par ailleurs que ce texte se veut, en dernière analyse, une amorce de réflexion sur un sujet hautement d'actualité – les sites Web institutionnels – et que de ce fait, il ne faut absolument pas y voir des propositions définitives reposant sur une étude exhaustive et finale. Ces limites étant clairement énoncées, nous croyons toutefois que notre contribution puisse mener à la réflexion et mettre le lecteur sur la piste de solutions réalistes et applicables dans son institution pour assurer une meilleure gestion de ces ensembles d'informations qui, s'ils ne sont pas apprivoisés, auront vite fait de devenir absolument monstrueux, incontrôlables, voire inutilisables.

Naturellement, il s'impose que d'autres après nous poursuivent le travail de recherche et de réflexion. Il nous paraît primordial que l'on continue d'être à l'affût

des développements technologiques et disciplinaires pour s'assurer que les préoccupations de gestion et plus précisément celles qui, au plan archivistique, caractérisent le traitement de ce type d'information soient toujours présentes à l'esprit des gestionnaires des institutions qui sont aux prises avec de telles problématiques. Car, il n'y a aucun doute dans notre esprit, nous sommes convaincus que le Web, et plus précisément le Web institutionnel, est un outil essentiel à la poursuite de l'évolution de nos institutions dans la société.

Bessem Khouaja Étudiant à la maîtrise en sciences de l'information. EBSI, Université de Montréal

Carol Couture Professeur titulaire. Directeur de l'EBSI, Université de Montréal

NOTES

1. *Web profond,* aussi appelé *net invisible, internet invisible* ou *Web caché* ou encore *Web invisible :* Partie du Web correspondant à l'ensemble des documents Web qui ne sont pas indexés par les outils traditionnels de recherche. Les données relatives à ces documents constituant le Web invisible peuvent être dynamiques (non localisables), non référencées (volontairement ou non) ou de nature non indexable (ex. : les animations). Les ressources du Web invisible comprennent, entre autres, les sites Web construits autour d'une base de données (interrogeable uniquement par un moteur de recherche interne), les pages accessibles par un formulaire de recherche, les pages protégées par un mot de passe, les pages interdites aux robots d'indexation, les pages écrites dans des formats propriétaires (Word, Flash, PDF, etc.), les intranets et les extranets. Actuellement, on estime que la taille du Web invisible atteint environ 40 % du contenu total (Vocabulaire d'Internet - Banque de terminologie du Québec).
2. *PURL :* Abréviation de *Persistent Uniform Resource Locator* qui forme un type de URL qui agit en tant qu'intermédiaire pour le véritable URL d'un site Web. Quand on entre une requête dans un moteur de recherche, ce dernier envoie la demande de la page Web, objet de recherche à un serveur PURL qui renvoie alors la véritable adresse URL de la page. Les adresses PURL sont persistantes parce qu'une fois qu'elles sont établies, il n'est pas besoin de les changer. La véritable adresse d'une page Web peut changer mais l'adresse PURL demeure. Ces adresses sont gérées et contrôlées par OCLC (Online Computer Library Center*). [*Internet.com PC Web/ Opaedia.*]*
3. *Snapshot :* Copie, reproduction de l'état de tout ou partie d'un système, d'un disque dur, d'une disquette, d'une partie de mémoire, d'un fichier, d'un programme ou de données quelconques, à un instant déterminé. Par exemple, un programme de mémoire virtuelle permet de créer une image d'une partie de la mémoire principale sur un disque (Office de la langue française 2002).
4. Terme emprunté de Castier (2002).
5. Les prochaines activités de la ECDL (troisième séance) se tiendront à Trondheim, en Norvège du 17 et 22 août 2003. L'archivage du Web en sera le thème principal.
6. Site Web du projet : http ://www.interpares.org/.
7. Site Web du projet : http ://www.kb.nl/coop/nedlib/.
8. Prenant comme exemple l'Université de Montréal, ses sites Web ont subi deux grands changements depuis leur création en 1995 : celui de 1998 où les sites ont changé de serveur et celui de 2001 où chaque département a été appelé à coordonner la structure du site et son aspect graphique pour que les internautes puissent identifier l'Université de Montréal chaque fois qu'ils sont sur un de ses sites. En outre, il a été convenu que tous les sites de l'Université permettent dans toutes les situations de retourner à la page d'accueil de l'Université, de consulter son bottin ou encore de consulter les ressources de ses bibliothèques. Selon cette façon de faire, nous pouvons considérer le site de

l'Université de Montréal comme un fonds et les versions de 1995, de 1998 et celle de 2001 comme des sous-fonds ou des séries selon le cadre de classement. Pour des raisons pratiques, il est plus facile pour le gestionnaire des documents de traiter le site Web comme une entité.

9. *Lien hypertexte relatif* : Lien hypertexte dont l'ancre possède un attribut HREF constitué de l'adresse Web relative aux données vers lesquelles il dirige l'internaute et qui sont généralement situées dans le même serveur ou le même répertoire que le document HTML qui contient l'ancre (Vocabulaire d'Internet - Banque de terminologie du Québec).
10. *Lien hypertexte absolu* : Lien hypertexte dont l'ancre possède un attribut HREF constitué de l'adresse Web complète des données vers lesquelles il dirige l'internaute et qui sont généralement situées dans un serveur différent de celui du document HTML qui contient l'ancre (Vocabulaire d'Internet - Banque de terminologie du Québec).

BIBLIOGRAPHIE

BERTRAND, Guy. 1998. L'Internet : anarchiste ou anarchiviste ? In *Actes du XXVI^e congrès de l'Association des archivistes du Québec, Magog-Orford, 28-30 mai 1998*, Magog-Orford, AAQ : 19-26.

BIBLIOTHÈQUE NATIONALE DU CANADA (BNC) ; Martin, Elizabeth. 2001. Gestion des publications électroniques diffusées en réseau : état de la question dans divers pays, novembre 2001. [En ligne]. http://www.nlc-bnc.ca/obj/r7/f2/r7-100-f.pdf (Page consultée le 01 mai 2003).

BIBLIOTHÈQUE NATIONALE DU CANADA (BNC). Groupe de coordination des collections électroniques. 1998. Politiques et directives relatives aux publications électroniques diffusées en réseau. Octobre 1998. [En ligne]. http://www.nlc-bnc.ca/publications/8/index-f.html#i (Page consultée le 01 mai 2003).

BOUCHARD, Carl et Jean-Stéphane PICHÉ. 1997. L'Internet et les archives. In *Actes du XXVI^e congrès de l'Association des archivistes du Québec, Alymer, 28-31 mai 1997*, Alymer, AAQ : 223-234.

CASTIER, Claire-Hélène. 2002. Plaidoyer pour une mémoire digitale. *Le Monde informatique* 960, 22 novembre 2002. [En ligne]. http://www.weblmi.com/articles_store/960_36/Article_view (Page consultée le 01 mai 2003).

CENTRE D'EXPERTISE ET DE VEILLE INFOROUTE ET LANGUES. 1998. *Internet, Intranet, Extranet : Comment en tirer profit*. Montréal, Éditions Transcontinental.

D'AMOURS, Liette. 2001. Gouvernance électronique : freins, impacts et bénéfices. *La Presse*, 18 décembre 2001.

DAUKANTAS, Patricia. 2002. What on Web merits saving? *Government Computer News* 21, 2, mai 2002. [En ligne]. http://www.gcn.com/21_12/news/18754-1.html (Page consultée le 01 mai 2003).

DAY, Michael. 2003. Collecting and preserving the World Wide Web : a feasability study undertaken for the JISC and Wellcome Trust. [En ligne]. http://library.wellcome.ac.uk/projects/archiving_feasibility.pdf (Page consultée le 01 mai 2003).

DIGITAL PRESERVATION COALITION (DPC). 2002. Web-archiving : managing and archiving online documents and records, mars 2002. [En ligne]. http://www.jisc.ac.uk/dner/preservation/webforum.html (Page consultée le 01 mai 2003).

DOLLAR, Charles. 2001. *Archival Preservation of Smithsonian Web Resources : Strategies, Principles, and Best Practices*. July 2001. [En ligne]. http://www.si.edu/archives/archives/dollar%20report.html (Page consultée le 01 mai 2003).

DUCHEIN, Michel. 1977. Le respect des fonds en archivistique : Principes théoriques et problèmes pratiques. *La Gazette des Archives* 97, 2e trimestre 1977 : 71-96.

FROST, John. 2001. Web technologies for information management. *The Information Management Journal* 35, 4 : 34-37.

GHARSALLAH, Mehdi. 2001. Pour que la mémoire ne flanche pas. *Archimag* 145, juin 2001 : 37-40.

GROS, Marie-Joëlle. 2001. La BNF cultive la mémoire du réseau. *Libération*, 9 juillet 2001.

GTRII (Groupe de travail sur les réseaux Internet / Intranets, Forum sur la gestion de l'information). 1999a. *Guide de mise en œuvre à la gestion de l'information sur Internet et Intranet pour assurer l'accès à long terme et la reddition de comptes*. [En ligne]. http://www.imforumgi.gc.ca/consult/inter-intra/implement2_f.pdf (Page consultée le 01 mai 2003).

GTRII (Groupe de travail sur les réseaux Internet / Intranets, Forum sur la gestion de l'information). 1999b. *Approche à la gestion de l'information sur Internet et Intranet pour assurer l'accès à long terme et la reddition de comptes*. [En ligne]. http://www.imforumgi.gc.ca/iapproach2_f.html (Page consultée le 01 mai 2003).

HAMEL, Michel. 1998. Enquête sur l'utilisation du Web pour la diffusion des archives. *Archives*. 30, 2 : 43-82.

HARRIES, Stephen. 1999. Capturing and managing electronic records from websites and Intranets in the government environment. In D*LM-Forum- European citizens and electronic information : the memory of the Information Society, Bruxelles, 18-20 octobre, 1999*. [En ligne]. http://europa.eu.int/ISPO/dlm/fulltext/full_harr_en.htm (Page consultée le 01 mai 2003).

HARRIES, Stephen. 2002. *The end of print to paper*. Update : Library information, juin, 2002. [En ligne]. http://www.cilip.org.uk/update/issues/june02/article4june.html (Page consultée le 01 mai 2003).

KAVCIC-COLIC, Alenka. 2002. Archiver le Web : Quelques perspectives juridiques. In *68th IFLA Concil and general Conference, Glasgow, August 18-24, 2002*. [En ligne]. http://www.ifla.org/IV/ifla68/papers/116-163f.pdf. (Page consultée le 01 mai 2003).

LÉGER, Danielle. 2001. Legal Deposit and the Internet : Reconciling Two Worlds. In *5th European Conference on Research and Advanced Technology for Digital Libraries, Darmstadt, Allemagne, le 8 septembre 2001*. [En ligne]. http://www.bnf.fr/pages/infopro/dli_ECDL2001.htm (Page consultée le 01 mai 2003).

LEFURGY, William G. 2001. *Records and archival management of World Wide Web sites*. [En ligne] http://www.mybestdocs.com/lefurgy-w-grn0104.htm#_ftn1 (Page consultée le 01 mai 2003).

LEMAY, Yvon. 1998. Les sites Web des services d'archives universitaires au Canada et la diffusion. *Archives* 30, 1 : 3-24.

LYMAN, Peter. 2002. Archiving the World Wide Web. In *Building a National strategy for digital preservation : issues in digital media archiving*. Washington : Library of Congress, 2002, p.38-51. [En ligne]. http://www.clir.org/pubs/reports/pub106/pub106.pdf. (Page consultée le 01 mai 2003).

MASANES, Julien. 2002a. Préserver les contenus du Web. [En ligne]. http://bibnum.bnf.fr/conservation/migration_web.pdf (Page consultée le 01 mai 2003).

MASANES, Julien. 2002b. Towards Continuous Web Archiving : First Results and an Agenda for the Future. *D-Lib Magazine*, vol. 8, no 12, décembre 2002. [En ligne]. http://www.dlib.org/dlib/december02/masanes/12masanes.html (Page consultée le 01 mai 2003).

MCCLURE, Charles R. et J. Timothy SPREHE. 1998. *Guidelines for electronic records management on state federal agency Websites*. [En ligne] http://istweb.syr.edu/~mcclure/guidelines.html (Page consultée le 01 mai 2003).

MORIN, Hervé. 2002a. Le dépôt légal du Web, terrain de compétition à la française. *Le Monde*, 06 avril 2002.

MORIN, Hervé. 2002b. Internet cherche à se préserver de l'amnésie. *Le Monde*, 06 avril 2002.

MOTZ, Arlene. 1998. Intranets : an opportunity for records managers. *Records management quarterly* 32, 3 : 14-16.

NATIONAL ARCHIVES OF AUSTRALIA (NAA). 2001. *Archiving Web resources : Guidelines for keeping records of Web-based activity in the Commonwealth Government*, mars 2001. [En ligne]. http://www.naa.gov.au/recordkeeping/er/web_records/archweb_guide.pdf (Page consultée le 10 octobre 2002).

NATIONAL DIET LIBRARY (NDL). 2002. Report : Web Resources as Cultural Heritage. In *International Symposuim on Web archiving, Tokyo 30 janvier 2002*. [En ligne]. http://www.ndl.go.jp/en/publication/ndl_newsletter/125/255.html (Page consultée le 01 mai 2003).

NOTESS, Greg R. 2002. The Wayback Machine : The Web's Archive. *Online*, vol. 26, no 2, mars/avril, 2002. [En ligne]. http://www.infotoday.com/online/mar02/OnTheNet.htm (Page consultée le 01 mai 2003).

OAKLEY, Kate. 2002. Qu'est ce que l'e-gouvernance ? In *Atelier sur l'e-gouvernance, Strasbourg 10-11 juin 2002*. [En ligne]. http://www.coe.int/t/f/projets_integres/democratie/Activit%E9s/docs_e-governance/IP(2002)9f.asp#TopOfPage (Page consultée le 01 mai 2003).

PARÉ, Richard. 2002. Les bibliothèques à l'ère de la démocratie électronique et du gouvernement en direct : Le cas du Canada. In *68th IFLA Concil and general Conference, Glasgow, August 18-24, 2002*. [En ligne]. http://www.ifla.org/IV/ifla68/papers/116-163f.pdf. (Page consultée le 01 mai 2003).

PERRIN, Charlotte. 2001. Archivage du Web : l'étranger montre le bon exemple. *Archimag* 145 : 32-36.

PUBLIC RECORD OFFICE OF UK. 2001. *Managing Web resources. Management of electronic records on websites and Intranet : An ERM toolkit*. December 2001. [En ligne]. http://www.pro.gov.uk/recordsmanagement/eros/website_toolkit.pdf. (Page consultée le 01 mai 2003).

REDFERN, Catherine. 2002. Web-archiving : an introduction to the issues. In *DPC Forum, Managing and archiving online documents and records, Londres. 25 mars 2002*. [En ligne]. http://www.dpconline.org/graphics/events/webforum.html (Page consultée le 01 mai 2003).

SECRÉTARIAT DU CONSEIL DU TRÉSOR DU CANADA. 2002. Politique de communication du gouvernement du Canada. [En ligne]. http://www.tbs-sct.gc.ca/pubs_pol/sipubs/comm/comml_f.html (Page consultée le 01 mai 2003).

SOUTH CAROLINA DEPARTMENT OF ARCHIVES AND HISTORY- ARCHIVES AND RECORDS MANAGEMENT DIVISION. 2002. Managing public records on web sites. *Public records information leaflet*, no 26, May 2002. [En ligne]. http://www.state.sc.us/scdah/26.pdf (Page consultée le 01 mai 2003).

TAUBMAN CENTER FOR PUBLIC POLICY (TCPP). 2002. Global E-Government, 2002 Second Annual Global E-Government Study. [En ligne]. http://www.infometre.cefrio.qc.ca/fiches/fiche523.asp (Page consultée le 01 mai 2003).

TRIPPE, William et Mark WALKER. 2003. Content Management and Web Publishing. In *The Columbia Guide to Digital Publishing*. New York : Colombia University Press, p. 418-454.

W3C. 1999. Spécification du modèle et la syntaxe du cadre de description des ressources (Resource description framework ou RDF). [En ligne]. http://www.la-grange.net/w3c/REC-rdf-syntax/#glossary (Page consultée le 01 mai 2003).