



World Library and Information Congress: 70th IFLA General Conference and Council

22-27 August 2004
Buenos Aires, Argentina

Programme: <http://www.ifla.org/IV/ifla70/prog04.htm>

Code Number: 113-F
Meeting: 141. Government Information and Official Publications
Simultaneous Interpretation: Yes

La stratégie nationale de conservation préventive aux Etats-Unis : défis et perspectives

Laura E. Campbell
Bibliothèque du Congrès
Washington, DC USA

Résumé

Avec l'inflation du volume d'information numérisée, l'expression créatrice à l'oeuvre dans les sciences pures, la technologie, les arts, les sciences humaines et les sciences sociales s'incarne de plus en plus dans ce support fragile, éphémère et dynamique. Si un tel support venait à disparaître, cela aurait des conséquences considérables et irréversibles sur la culture des générations futures. C'est pourquoi le Congrès américain a chargé la Bibliothèque du Congrès de conduire un effort national pour la mise en place d'une infrastructure permettant d'identifier, d'acquérir, de gérer et de conserver les oeuvres importantes existant sous forme numérisée : le Programme National pour l'Infrastructure de l'Information et la Conservation préventive (sigle américain : NDIIPP), loi votée en décembre 2000 (PL - 106 554). Cette communication décrit notre travail à ce jour et avance des idées sur la façon dont le gouvernement peut encourager des partenariats pour atteindre cet important objectif national.

Avec l'inflation du volume d'information numérisée, l'expression créatrice à l'oeuvre dans les sciences pures, la technologie, les arts, les sciences humaines et les sciences sociales s'incarne de plus en plus dans ce support fragile, éphémère et dynamique. Si un tel support venait à disparaître, cela aurait des conséquences considérables et irréversibles sur la culture des générations futures. C'est pourquoi le Congrès américain a chargé la Bibliothèque du Congrès de conduire un effort national pour la mise en place d'une infrastructure permettant d'identifier, d'acquérir, de gérer et de conserver les oeuvres importantes existant sous forme numérisée : le Programme National pour l'Infrastructure de l'Information et la Conservation

préventive (sigle américain : NDIIPP). La loi votée en décembre 2000 (PL - 106 554) affecte environ 100 millions de dollars à ce programme, dont l'engagement sera autorisé par tranches successives : 5 millions aussitôt pour les études préparatoires, 20 millions devant être débloqués après l'approbation du projet par le Congrès et les 75 millions restants subordonnés à la collecte de subventions d'un montant équivalent. Le projet, *Conserver le patrimoine numérique*, a été soumis au Congrès à l'automne 2002 et, en décembre de la même année, a été approuvé officiellement, ce qui nous a permis de passer à l'étape suivante qui s'appuie sur environ 20 millions de dollars de fonds publics et environ 15 millions collectés auprès de donateurs privés ou ne dépendant pas de l'Etat fédéral.

Nous avons depuis lors mis en oeuvre un programme à plusieurs dimensions, regroupant à travers le monde des partenaires et des communautés intéressées dans les différentes branches d'activité. Nous avons lancé un concours pour encourager la construction d'un réseau de partenariats afin d'identifier et capturer les contenus numériques en péril. Notre travail pour peaufiner une architecture technique se poursuit avec l'intention d'identifier les protocoles nécessaires à la conservation préventive des données pendant la saisie et le transfert. Le projet *Archive Ingest and Handling*, connu désormais sous ce nom, repose sur quatre partenaires qui testent sept technologies de conservation préventive. Le corpus des données pour le test est un archivage de 12 gigabits de documents publics rassemblés la George Mason University pendant la période qui a suivi le 11 septembre 2001 et qu'elle a donné à la Bibliothèque du Congrès avec autorisation de s'en servir pour ledit test. En collaboration avec NSF et d'autres agences fédérales, la Bibliothèque envisage aussi de financer un programme de recherche fondamentale ; un appel à propositions est attendu l'an prochain.

Enfin, nous coopérons avec d'autres bibliothèques nationales pour soutenir un consortium international en vue de la conservation préventive de la Toile. Nos partenaires sont la Bibliothèque nationale de France, la British Library et les bibliothèques nationales d'Australie, du Canada, du Danemark, de la Finlande, de l'Islande, de l'Italie, de la Norvège et de la Suède ainsi que l'*Internet Archive*. Les objectifs clés sont les suivants :

- un travail de collaboration, dans le cadre des différentes législations nationales, en vue d'identifier, développer et faciliter la mise en place de solutions pour sélectionner, rassembler, conserver et donner accès au contenu de l'Internet
- une aide pour parvenir à une couverture internationale des collections d'archives de contenu de l'Internet dans le cadre des législations nationales et en accord avec les différentes politiques documentaires nationales
- un plaidoyer international pour les initiatives encourageant l'acquisition, la conservation et l'accès au contenu de l'Internet

Pour atteindre ces objectifs, le Consortium devra :

- fournir un forum de partage de connaissances au sujet de l'archivage des données internet à la fois au sein du consortium et à l'extérieur
- développer et préconiser des normes
- développer des outils et des techniques avec interfaces pour acquérir, archiver et fournir l'accès aux sites web
- faire mieux connaître les problèmes et les initiatives de conservation préventive de l'Internet par le biais de conférences, ateliers, séminaires, publications, etc.

Nos accords de partenariat avec de grandes bibliothèques universitaires et leurs partenaires réuniront en fin de compte quelque 25 organisations appartenant à l'enseignement supérieur américain, aux bibliothèques des Etats, et aussi bien au secteur à but non lucratif qu'au secteur à but lucratif. Nous espérons beaucoup apprendre sur la manière de structurer et administrer des partenariats au fur et à mesure que nous approfondirons notre savoir-faire en matière de collecte de l'information numérique. Les Archives nationales américaines participent aussi à l'un des projets, mais en raison des contraintes législatives sur notre financement et des exigences compensatoires, les Archives sont un partenaire 'silencieux' mais actif. D'autres agences d'Etats ou fédérales sont également impliquées. Nous sommes extrêmement satisfaits de la vaste couverture géographique ainsi que de la diversité des formats représentés : géospatial (plusieurs formats), télévision numérique, texte, sites web, courriels et bases de données en sciences sociales. Parmi les thèmes que nous voulons traiter, il y a la protection de la vie privée et la confidentialité des données, la propriété intellectuelle, les systèmes de transmission de l'héritage, la maintenance des logiciels, et toute une série de questions techniques liées à l'identification et à la collecte de l'information en péril.

Dans l'ensemble, les problèmes techniques sont assez bien connus. Les supports et les signaux d'encodage se dégradent ; l'obsolescence des matériels et des logiciels rend la relecture des données problématique ; les schémas de métadonnées semblent proliférer et ne sont pas toujours compatibles d'un format à l'autre, étant donné les besoins différents, mettons, pour donner un exemple, du cinéma et du texte. Le simple fait de copier des données d'un support sur un autre (opération dite 'rafraîchissement des données') peut potentiellement introduire des erreurs, et si certaines de ces erreurs sont sans conséquences sur le contenu de l'œuvre, il n'en va pas de même pour toutes et de plus, les deux types d'erreurs posent des questions sur l'intégrité et l'authenticité des œuvres enregistrées. Les spécialistes de conservation préventive du numérique ont débattu des mérites relatifs de la migration, de l'émulation, de l'encapsulation, et actuellement, il semble y avoir un consensus autour de l'idée que différentes stratégies de conservation peuvent être mises en place selon la nature des données, la fréquence de leur utilisation et l'importance de préserver l'aspect original¹. Les droits de propriété intellectuelle sont complexes et leur gestion a varié dans le temps et selon les supports : la radiodiffusion et l'édition universitaire par exemple se trouvent sous la même rubrique de la loi sur le copyright alors que les structures industrielles et la pratique diffèrent.

Finalement, en dépit d'extraordinaires réussites techniques en matière de rapidité des processeurs et de capacité de mémoire, la conservation préventive, ce n'est pas simplement la possibilité de compresser des données sur un disque de haute densité et de pouvoir les relire. Le défi, nous le savons tous, est de rendre de l'information stockée accessible - non pas seulement au sens où nous pouvons la visualiser, si ardu que cela soit, mais accessible au sens où les futurs utilisateurs seront capables de trouver ce qu'ils cherchent, de s'en servir d'une façon qui respecte l'intégrité de l'information et les intérêts légitimes en jeu. Cela revient à dire qu'il nous faut anticiper les attentes de l'utilisateur du 22^e siècle. Ou du moins qu'il ne faut pas commettre de dommages dans le présent tout en permettant au progrès technologique de s'occuper des attentes de l'utilisateur dans cent ans.

Plus récemment, l'investigation a été poussée plus loin : comment les différentes composantes s'harmonisent-elles ? Comment le système répond-il aux besoins de l'institution et de ses

¹ Voir Kyong-Ho Lee, Oliver Slattery, Richard Lu, Xiao Tang, Victor McCrary, « The State of the Art and Practice in Digital Preservation », *Journal of Research of the National Institute of Standards and Technology* 107 (1) (January-February 2002), p. 93-106.

usagers présents et futurs ? C'est-à-dire à quoi ressemble une architecture technique de conservation préventive ?

Nous avons réuni un petit groupe de techniciens experts en février 2001 qui a dressé les grandes lignes des principes du projet et donné l'ébauche de l'architecture que nous avons examinée soigneusement dans différents groupes de discussion. Nous avons affiné la proposition initiale au terme de nouvelles réunions en 2002, mais les principes sont restés les mêmes. Ils prônent les valeurs de transparence, collaboration, développement incrémentiel, stabilité, flexibilité, hétérogénéité et innovation, en même temps qu'ils suggèrent une approche souple permettant à de nombreuses parties de travailler sur des aspects différents tout en assurant la cohérence d'ensemble.

Des notes techniques sont disponibles sur notre site web, www.digitalpreservation.gov, et j'espère que, si vous vous occupez d'archives numériques, vous prendrez le temps de les consulter. Mais je voudrais m'arrêter un instant sur plusieurs points étudiés par notre consultant Clay Shirky (*Update to the NDIIPP Architecture, Version 0.2*) qui traitent directement des relations entre organisation et technologie :

1. Nous sommes conscients du fait que beaucoup d'énergie est déployée par nombre d'institutions sur cette question de la conservation préventive du numérique. En conséquence, l'architecture NDIIPP admet la nécessité d'interfaces entre les institutions afin de pouvoir échanger aussi bien des documents particuliers que des collections entières, et d'autre part la possibilité pour les institutions de jouer différents rôles simultanément.
2. Il existe de nombreux systèmes de conservation préventive prêts à être testés et il y a une forte demande pour que ces systèmes soient fédérés ou autrement pour trouver le moyen de permettre à ces systèmes de fonctionner ensemble - d'où notre projet d'archivage et de test
3. Une convergence rapide sur un seul système ou sur une famille de systèmes pouvant fonctionner ensemble est peu probable. Chaque système est précisément conçu pour répondre aux objectifs de l'institution source ou de l'institution qui finance, et ces objectifs ne sont bien sûr pas tous les mêmes. La conservation préventive du numérique ne s'en portera que mieux, car l'hétérogénéité est un garant contre une défaillance à grande échelle du système. Mais la compatibilité fondée sur les mêmes outils et formats, ou même sur le même modèle conceptuel est un but inaccessible. La compatibilité au premier degré fondée sur la notion « chacun utilise les mêmes outils et formats » et la compatibilité au second degré « chacun utilise le même modèle conceptuel » sont deux buts inaccessibles, tant maintenant que dans un avenir prévisible.
4. En conséquence, l'architecture NDIIPP est construite pour soutenir des institutions qui coopèrent tout en fonctionnant sur des technologies différentes.

La tâche est ardue. Mais nous croyons que c'est à la fois possible et nécessaire. Les documents que nous et d'autres produisons favoriseront les échanges entre le NDIIPP et ses partenaires, ainsi qu'à l'intérieur de l'ensemble plus vaste des projets de conservation de numérique qui ne sont pas forcément ou formellement liés à notre initiative. Nous ne doutons pas non plus que les expériences concrètes que nous et d'autres engrangeons serviront à enrichir et affiner le travail.

Au fil du temps, les bibliothèques et les archives ont élaboré des pratiques, des normes et des relations de coopération dont la somme est une infrastructure intangible qui soutient l'acquisition et la protection de la production intellectuelle de nos usagers. De concert avec

des organisations telles que l'IFLA, nous avons, à la Bibliothèque du Congrès avec nos homologues américains ou étrangers, étendu nos missions nationales respectives pour faire en sorte que nous employions des normes internationales compatibles et que nous conservions globalement la mémoire du génie humain. Nos arrière-petits-enfants nous en seront reconnaissants.

août 10, 2004