



IFLA  
2005  
OSLO

## World Library and Information Congress: 71th IFLA General Conference and Council

### "Libraries - A voyage of discovery"

August 14th - 18th 2005, Oslo, Norway

*Conference Programme:*

<http://www.ifla.org/IV/ifla71/Programme.htm>

juillet 20, 2005

**Code Number:**

032-F

**Meeting:**

150 SI - ICABS (IFLA/CDNL Alliance for Bibliographic Standards)

**Entreprendre un projet d'archivage électronique: ce qu'il faut mettre en place pour préserver l'accès aux collections numériques<sup>1</sup>**

### **Pam Gatenby**

Adjointe au Directeur général du Département des collections  
Bibliothèque nationale d'Australie

*Traduction en Français : Gildas ILLIEN, SCU*

*Paris 8 (trad. Le 20/07/2005)*

*[Illien.gildas@wanadoo.fr](mailto:Illien.gildas@wanadoo.fr)*

### **Résumé**

*Les institutions nationales chargées de l'archivage électronique sont confrontées à des enjeux de taille. Alors que l'information à collecter est pléthorique et que les ressources sont insuffisantes, le chemin à parcourir en décourage plus d'un. De nombreuses expériences à travers le monde démontrent néanmoins qu'il est possible de faire la différence en s'engageant dans des projets modestes. Dans tous les cas, il est préférable de commencer par quelque chose, même d'insatisfaisant, plutôt que de rien faire du tout sous prétexte d'attendre que tous les problèmes aient été préalablement résolus.*

*Des politiques institutionnelles adaptées, des procédures, des méthodes de travail, des compétences et des infrastructures techniques sont nécessaires afin d'organiser la gestion et*

---

<sup>1</sup> This presentation is based on the documents listed in the bibliography at the end of the paper. Additional information relevant to the theme of the presentation is contained in those documents. I would like to acknowledge Colin Webb, Director of Preservation at the National Library of Australia, the author of the UNESCO guidelines document, which the presentation draws on in particular.

*la conservation durable des ressources électroniques. Toutefois, il n'est pas indispensable et peut-être même pas souhaitable d'avoir développé dès le départ des programmes aboutis d'archivage et de conservation électroniques avant de s'engager dans un tel projet.*

*Des programmes a minima permettent déjà de sauvegarder des éléments essentiels du patrimoine numérique qui seraient sans cela condamnés à disparaître.*

*Cet article présente les procédures essentielles à la mise en place d'un accès continu aux ressources numériques :*

- *la sélection et la collecte sécurisées des données dont la valeur semble durable (l'archivage électronique) ;*
- *l'organisation des informations collectées afin d'en faciliter la gestion et la consultation ;*
- *la mise en œuvre d'un programme de **conservation numérique** (data management) qui limite les menaces que représente l'évolution des logiciels et des technologies dont la consultation des documents numériques est tributaire.*

## **Introduction**

Les ressources électroniques, le Web et Internet occupent désormais une place centrale dans l'environnement informationnel mondial. Dans l'intérêt des générations futures, la plupart des bibliothèques nationales et des autres institutions chargées de l'archivage électronique endossent une responsabilité de taille en s'engageant à préserver au moins une petite partie de ces ressources. Cette tâche s'inscrit dans la continuité de leurs missions : nous nous devons de sauvegarder la mémoire des tentatives et des réalisations humaines.

Chacun est désormais à même de comprendre que les enjeux auxquels nous sommes confrontés sont considérables. Alors que l'information à collecter est pléthorique et que les ressources sont insuffisantes, le chemin à parcourir en décourage plus d'un. De nombreuses expériences à travers le monde démontrent néanmoins qu'il est possible de faire la différence en s'engageant dans des projets modestes. Dans tous les cas, il est préférable de commencer par quelque chose, même d'insatisfaisant, plutôt que de rien faire du tout sous prétexte d'attendre que tous les problèmes aient été préalablement résolus.

Au cours de cette intervention, j'aborderai à rebours les termes qui composent le titre de ma communication. Je dresserai tout d'abord un panorama des procédures et des systèmes qu'il faut impérativement mettre en place pour organiser la gestion des ressources numériques et en préserver la consultation à long terme. J'apporterai ensuite quelques avis et conseils sur la manière dont les institutions qui ne se sont pas encore engagées dans cette démarche pourraient le faire. Bien qu'il puisse être appliqué plus largement à d'autres types de ressources numériques et à d'autres institutions, je tiens enfin à préciser que mon propos concerne avant tout les ressources issues du Web et le rôle spécifique des bibliothèques nationales dans ce domaine.

Ma communication se veut générale, je n'entrerai pas dans les détails. Elle doit en effet servir de cadre introductif aux communications suivantes qui se concentreront sur des aspects plus spécifiques des procédures ou des actions auxquelles j'aurai fait allusion dans mon panorama.

Ce qu'il faut toujours garder en tête :

Avant d'entrer dans le vif du sujet, je voudrais poser quelques principes qu'il convient de garder toujours à l'esprit au commencement d'un tel projet :

- Si la plupart des programmes sont lancés sur le mode du projet, il faut envisager dans une étape ultérieure leur consolidation dans une organisation pérenne du travail;
- Les institutions doivent clarifier les aspects juridiques liés à la collecte, à la reproduction, à la conservation et à la consultation des ressources ;
- La conservation des documents électroniques ne saurait être exhaustive : il ne faut garder que les données dont la valeur semble durable – précisons toutefois que si l'on peut décider après coup de ne pas conserver une ressource archivée, le choix de ne pas archiver une ressource est *a priori* définitif ;
- Pour les ressources qu'on a choisi de conserver, l'objectif est de garantir leur consultation permanente : on ne peut parler d'archivage électronique si les ressources ne sont plus accessibles ;
- Endosser la responsabilité de cet archivage implique un engagement continu et efficace – on n'obtient que ce qu'on entreprend effectivement : une bienveillante négligence ne suffit pas ;
- Si l'on attend que des solutions d'ensemble fiables apparaissent avant de s'engager, des données seront perdues. Mieux vaut tâtonner dans des démarches limitées et pas toujours fiables que ne rien faire du tout. Les petits pas sont préférables à l'immobilisme.
- La gestion des ressources électroniques implique l'acceptation et la gestion du risque ainsi que des décisions qui impliquent parfois que l'on mette en péril certains contenus ou la possibilité d'y accéder ;
- La coopération est souvent un moyen de mutualiser les charges comme les expertises et de gagner en efficacité, car elle permet d'établir des programmes de conservation partagée. Elle ne dispense pas pour autant d'assumer des investissements financiers ni de faire des choix ;
- L'utilisation de systèmes standard est nettement préférable. Elle limite en effet le nombre d'options à gérer, facilite la coopération et le partage des tâches grâce à l'interopérabilité dans le temps et dans l'espace, et permet de se prémunir contre une gestion hasardeuse des procédures d'archivage.

### **Que faut-il mettre en place pour réussir l'archivage électronique ?**

En résumé, on a d'abord besoin de politiques institutionnelles adaptées, de procédures, de méthodes de travail, de compétences et d'infrastructures techniques. Toutefois, une condition préalable demeure l'acceptation par l'institution chargée du projet de sa pleine responsabilité dans la mise à disposition pérenne des ressources numériques.

Les programmes d'archivage électronique ont plus de chances d'aboutir s'ils s'inscrivent dans les missions et dans la politique générale d'un établissement qui est reconnu et rend régulièrement des comptes à la collectivité en matière de collecte, de gestion, de diffusion et de conservation de l'information. C'est en effet ce contexte qui va permettre de préciser les critères de sélection, les nouvelles infrastructures à installer, les normes à employer et le cadre juridique adapté à la conduite du projet. Le succès peut aussi dépendre de la place accordée au programme d'archivage électronique dans l'organisation de l'institution à qui cette mission aura été confiée. Compte tenu du caractère à la fois coûteux et complexe des opérations à engager, l'existence de soutiens privés peut également être un atout.

## Les actions prioritaires

S'il existe différents modèles à peu près équivalents pour le traitement de la documentation électronique, certaines procédures fondamentales doivent être impérativement respectées pour répondre à l'objectif d'une consultation pérenne. On peut les regrouper autour des trois axes suivants :

- la sélection et la collecte sécurisées des données dont la valeur semble durable (l'archivage électronique) ;
- l'organisation des informations collectées afin d'en faciliter la gestion et la consultation ;
- la mise en œuvre d'un programme de conservation numérique (*data management*) qui limite les menaces que représente l'évolution des logiciels et des technologies dont la consultation des documents numériques est tributaire.

### La sélection et la collecte des données

La politique documentaire et les procédures qui précisent les critères de sélection, le degré d'exhaustivité et les modalités de la collecte doivent avoir été définis au préalable.

D'un point de vue conceptuel, la sélection du patrimoine numérique est identique à celle des autres types de documents. Des approches similaires s'appliquent – par exemple : les choix de sélection fondés sur des plans de développement des collections, l'appréciation de ce qui donne une valeur durable à un document, ou encore la nécessaire connaissance du contexte de production et de la provenance du document. Les documents électroniques soulèvent toutefois des défis nouveaux, à commencer par leur masse considérable, leur qualité très inégale (du fait des conditions largement répandues de leur production), leur nature changeante et éphémère, enfin, la difficulté à en définir les frontières et à en identifier les auteurs. Si la sélection des documents repose sur des jugements critiques et doit laisser une place au doute, elle doit néanmoins répondre aux critères de la politique documentaire qui reflète les objectifs de l'organisation chargée de l'archivage et qui fonde sa légitimité publique. Ces critères doivent refléter les fondements de la politique documentaire : pourquoi l'on conserve certaines informations et pas d'autres.

La protection des données numériques sélectionnées doit se faire dans un lieu de stockage sécurisé (archives numériques) plutôt que dans leur environnement d'exploitation. En effet, c'est là qu'il reste possible de garantir leur consultation sur le long terme. Certaines institutions créent leurs propres archives tandis que d'autres, qui travaillent en réseau selon un mode collaboratif ou distributif, choisissent de recourir aux services d'un autre organisme. Le type de stockage des archives doit être sécurisé et fiable ; il doit supporter des procédés qui protègent les flux de données des altérations, destructions et pertes involontaires. Il n'est pas nécessaire d'acheter des équipements coûteux pour garantir cette sécurité pourvu que les machines et le système soient bien gérés. Toutefois, plus les données à stocker sont hétérogènes et complexes, plus le système de stockage doit être sophistiqué.

Le stockage des données dans un espace sécurisé s'accompagne de mesures de gestion essentielles :

### *Préalables juridiques à l'archivage*

La question juridique du point de vue de la propriété intellectuelle d'une part et du dépôt légal d'autre part doit avoir été réglée avant la collecte des données. En particulier, il est indispensable d'obtenir la permission de reproduire et de donner accès aux ressources archivées à des fins de conservation.

### *Modalités de l'archivage*

La plupart des stratégies de transfert découlent de l'une ou l'autre des approches suivantes : le dépôt des ressources par leurs auteurs ou la collecte des ressources par la bibliothèque (pour les ressources du Web, la collecte consiste en une copie de la ressource originale). Quelle que soit l'approche retenue, l'organisation du travail et les outils utilisés devront permettre de prendre des décisions relatives à la fréquence de collecte des ressources dynamiques, ou, encore plus compliqué, à la restriction de l'accès. Le consortium IIPC développe actuellement un outil de moissonnage du Web (du nom de Heritrix) qui devrait être mieux adapté que d'autres logiciels du domaine public aux besoins spécifiques d'archivage des institutions culturelles. Le consortium a également développé des outils et des procédures qui facilitent la collecte et l'archivage des sites Web qui contiennent des bases de données. Enfin, il réalise une boîte à outils destinée aux professionnels des petites institutions qui ne disposent pas d'infrastructures technologiques très développées pour collecter et gérer les ressources du Web.

### L'organisation des données

Une fois transférées dans un espace de stockage ou d'archivage, les données numérisées doivent être contrôlées et organisées afin qu'on puisse aisément les localiser, les consulter, les gérer et les conserver. Ceci implique les opérations suivantes :

#### *Assigner des identifiants*

Chaque fichier numérique affecté à un espace d'archivage doit être doté d'un nom de fichier unique afin de ne pas être confondu avec un autre ; il doit également disposer d'un identifiant durable qui permettra de retrouver la ressource sans qu'il y ait rupture du lien URL si le site change d'emplacement ou d'adresse.

#### *Décrire les ressources digitales en utilisant des schémas de métadonnées standard*

Les métadonnées (qui sont des données structurées qui décrivent les documents) permettent de retrouver les informations au moyen d'un système de recherche. Du point de vue de la conservation, elles servent à gérer et à représenter les documents à l'écran. Il est préférable de retenir un schéma standard de métadonnées car les logiciels les reconnaissent automatiquement et les métadonnées pourront être partagées avec d'autres bibliothèques. L'utilisation de schémas standard garantit par ailleurs la cohérence et la qualité des recherches de même qu'elle encourage la normalisation des procédures de conservation. Plusieurs projets de normalisation sont en cours pour les métadonnées, parmi lesquels le projet PREMIS, initié par OCLC et le RLG. PREMIS recommande l'utilisation d'éléments de métadonnées et des procédures d'implémentation destinées à la gestion, à la conservation durable et à l'utilisation des ressources électroniques. La Bibliothèque du Congrès est également très active en matière de développement des normes de métadonnées et il sera fait état de ses avancées dans ce domaine au cours du programme de la journée.

Afin de faciliter la recherche et la consultation des ressources, un niveau basique de métadonnées (qui comporte par exemple le titre, le nom de l'auteur, celui de l'éditeur et des descripteurs de sujet) est requis. Les métadonnées doivent par ailleurs être intégrées à un système de recherche : le catalogue de la bibliothèque ou une base de données publique. Pour ce qui concerne la gestion et la conservation, ce sont des métadonnées relatives à la nature technique des ressources (telles que le type de ressource, le format de fichier, la structure des données, les procédures utilisées pour sa création et le matériel logiciel nécessaires à son utilisation) qui sont nécessaires. On aura également besoin d'enregistrer les informations relatives aux changements et aux nouvelles procédures qui auront affecté le document au cours de sa vie, ainsi que de toutes les données utiles si l'on veut pouvoir continuer à le consulter.

### *Faciliter la consultation*

Pour la plupart des institutions chargées de l'archivage électronique, l'objectif *in fine* de cette politique de conservation numérique est la consultation courante et future. Pour permettre la consultation effective des ressources archivées, il est nécessaire de prévoir les modalités d'accès à ces documents. Ceci peut notamment impliquer la mise en place d'une interface de recherche qui viendra se greffer sur la base d'archivage, celle d'un mécanisme assurant la visualisation des documents retrouvés, ou encore la gestion des droits, des restrictions et de tous les autres aspects liés à la consultation. L'accès aux ressources sera plus ou moins étendu selon les droits qui auront pu être négociés avec les auteurs et les éditeurs dans le respect de la législation en vigueur en matière de propriété intellectuelle. La consultation à proprement parler dépendra de la qualité des métadonnées descriptives employées dans les différents types d'interfaces de recherche utilisées.

Par exemple, les archives numériques de la Bibliothèque nationale d'Australie (<http://www.pandora.nla.gov.au>) peuvent être consultées depuis l'OPAC de la Bibliothèque ou bien à partir de la page d'accueil de son site d'archives (recherche par liste de titres ou à partir d'un index). D'autres archives, comme l'Internet Archive offrent la possibilité de la recherche *time line* qui permet de comparer différentes versions d'un même site web, capturées à des moments différents.

Le système de gestion d'archives électroniques utilisé par la Bibliothèque nationale d'Australie s'appelle PANDAS (PANDORA Digital Archiving System). Il a été développé par la Bibliothèque elle-même afin de répondre à ses besoins spécifiques et est gratuitement utilisable par toutes les autres institutions. Un système d'évaluation est en outre proposé pour aider les autres bibliothèques à voir si PANDAS répond à leurs besoins. (Pour plus de renseignements, consulter le site <http://www.pandora.nla.gov.au> )

### *Mettre en place un programme de conservation numérique*

La conservation numérique recouvre un ensemble de procédures qui visent à garantir la consultation durable des ressources électroniques. Cela nécessite de maintenir intacts les flux de données intacts et les moyens de les lire, de les interpréter et de (re)présenter à l'utilisateur le document tel qu'il avait été créé initialement. Pour arriver à ce résultat, il est vraiment essentiel de bien connaître les caractéristiques des ressources archivées mais aussi les menaces qui pèsent sur leur consultation future. Les supports de stockage numérique sont fragiles, ils sont exposés à la destruction ou à la détérioration. Les logiciels comme les matériels deviennent rapidement obsolètes, c'est pourquoi le cadre temporel à l'intérieur duquel les décisions de conservation devront être prises apparaît particulièrement court en

comparaison avec ce qu'on observe pour la conservation des documents imprimés. Les actions de sauvegarde doivent donc être engagées le plus tôt possible mais devront être poursuivies aussi longtemps qu'on aura besoin de consulter les documents.

Aspects essentiels de la conservation numérique :

#### *Clarifier les objectifs*

Il faut prendre des décisions sur ce qu'il convient de préserver et pour combien de temps. L'étendue et la nature des responsabilités et des décisions prises par l'institution chargée de l'archivage auront été précisées dans un document fixant sa politique documentaire dans ce domaine. Il faut également tenir compte du niveau de fonctionnalité et d'authenticité qui sera attendu par les futurs utilisateurs ainsi que de la nature des informations à enregistrer afin de comprendre le contenu de ces ressources à plus long terme.

#### *Connaître sa collection*

Il est important d'identifier ce qui constitue sa collection numérique et d'enregistrer les caractéristiques techniques de ses documents au moyen d'un schéma standard de métadonnées. La connaissance de la collection permettra de fixer des priorités, notamment de repérer les risques les plus élevés, tels que le vieillissement ou la vulnérabilité des supports, des logiciels et des formats de fichiers devenus obsolètes.

#### *Protéger les flux de données*

Ceci est essentiel pour protéger l'accès aux collections et permettre leur protection durable selon de bonnes pratiques d'informatique documentaire. En particulier, la protection des flux de données peut nécessiter la constitution d'espaces de stockage physiques adaptés aux supports sur lesquels sont conservées les documents, la copie des données dans un système de stockage numérique fiable, la mise en place de sauvegardes régulières, des procédures de récupération des données en cas de sinistre et des transferts réguliers des données vers des supports de nouvelle génération.

#### *Gérer les risques*

Les programmes d'archivage électronique doivent s'efforcer de réagir aux menaces qui peuvent mettre en péril l'accès durable aux ressources. La gestion des risques constitue une approche judicieuse pour savoir quels dangers requièrent une attention particulière et la mise en place de plans d'action destinés à les écarter. Cette approche contribuera également à la planification et à la priorisation des tâches sur le long terme.

#### *Garantir l'accessibilité*

Le défi essentiel des programmes d'archivage électronique reste l'usage futur qui sera fait des ressources archivées lorsque les mutations technologiques auront rendu obsolètes les formats et les environnements d'exploitation initiaux. On continue de développer des solutions de long terme, qui incluent le transfert des données vers de nouveaux formats et la création de programmes permettant de préserver la présentation originale des documents grâce à l'émulation des configurations initiales qui permettent de lancer les logiciels d'origine.

A moyen terme, on pourra emprunter les étapes suivantes :

- conserver aussi longtemps que possible les matériels et les logiciels originaux qui sont nécessaires pour accéder aux ressources de la collection ;
- limiter le nombre de formats en transférant les ressources sur des formats standard
- conserver et organiser les données originales au cas où des solutions futures permettraient un jour d'en restaurer l'accessibilité
- surveiller l'environnement technologique et guetter les signes d'obsolescence des matériels, des formats et des normes
- toujours travailler en partenariat avec d'autres institutions pour développer des solutions communes

### **Les infrastructures techniques**

La mise en oeuvre des actions recommandées dans cette intervention nécessite une solide infrastructure technique. L'architecture actuellement utilisée par la Bibliothèque nationale d'Australie fournit un exemple d'infrastructure qui accueille à la fois les activités de numérisation des collections analogiques de la Bibliothèque, la collecte de ressources nativement numériques et l'archivage des ressources électroniques en vue d'une préserver l'accès durable. Cette architecture est représentée dans le diagramme ci-dessous : *[diagramme non traduit]*

#### **Pour commencer (et pour finir)**

Il n'est pas indispensable, et peut-être même pas souhaitable, d'avoir construit *ex nihilo* des programmes avancés d'archivage électronique avant de s'engager dans ce projet. En effet, une telle entreprise découragerait la plupart des institutions chargées de cette mission. Il est préférable de démarrer modestement et avec peu de ressources, en se limitant par exemple à des données simples, dans le but de fournir la meilleure qualité de service possible compte tenu des contraintes et des connaissances actuelles. On pourra suivre les étapes suivantes :

- définir les ressources électroniques pour lesquelles vous avez une responsabilité de conservation et travailler en liaison avec d'autres établissements qui ont des missions similaires aux vôtres pour envisager un travail collaboratif ;
- repérer d'autres organismes qui ont acquis une expérience dans la gestion du type de ressources qui vous intéresse et les solliciter pour bénéficier de leur avis et de leurs conseils ;
- identifier les risques les plus élevés et qui requièrent des actions de prévention immédiates afin d'éviter que les informations les plus précieuses ne soient perdues ;
- repérer les actions les plus urgentes, et surtout les plus simples, qui vous permettront de contrer rapidement ces menaces. Par exemple : assurer la protection des données en s'assurant de leur stockage et de leur archivage dans de bonnes conditions ; acquérir une meilleure connaissance des caractéristiques de ces ressources et de la manière dont les usagers souhaiteront les exploiter et les consulter dans le futur ;
- résoudre les questions de droits et d'autorisations pour s'engager dans le projet et approcher les auteurs afin de négocier avec eux les droits de diffusion lorsque c'est nécessaire ;
- définir aussi tôt que possible les politiques qui vous guideront par la suite dans vos différents engagements ;



- évaluer chacune des étapes de votre démarche et procéder aux remises en cause nécessaires pour mettre en place un programme complet et fiable dans la durée – tirer de votre expérience les enseignements utiles et construire à partir de cette base ;
- évaluer enfin les ressources nécessaires à la conduite d'un programme plus solide et plus fiable à long terme.

Bien qu'ils soient évidemment plus limités dans leur impact direct, les programmes d'archivage électronique *a minima* permettent d'assurer la sauvegarde d'une partie significative du patrimoine numérique qui serait autrement perdu. Comme je l'ai dit au début de cette intervention, attendre que des solutions fiables et complètes soient disponibles avant d'entreprendre quoi que ce soit entraînerait probablement la perte de ces ressources. Il est préférable de commencer par quelque chose, même incomplet et peu fiable, plutôt que de rien faire du tout.