



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: 026-F
Meeting: 148. IFLA-CDNL Alliance for Bibliographic Standards (ICABS)
Simultaneous Interpretation: -

**Collecter et gérer les ressources Internet pour un accès pérenne :
moissonnage et directives pour l'aide à la conservation (Actions de
l'ICABS 3.3 et 3.4)**

Pam Gatenby

Directrice Générale Adjointe
Gestion des Collections
Bibliothèque Nationale d'Australie
pgatenby@nla.gov.au

RESUME :

La Bibliothèque Nationale d'Australie participe à l'objectif 3 du plan d'action de l'ICABS¹, qui consiste à faire progresser la compréhension des questions relatives à l'archivage pérenne des ressources numériques. Elle s'attache plus particulièrement aux actions réalisées afin de faire progresser la coopération dans le domaine de l'archivage du web (action 3.3) et la conservation des ressources numériques (action 3.4), mettant à profit sa propre expérience dans ces domaines. Ces actions spécifiques consistent à orienter le groupe de travail sur le web invisible de l'International Internet Preservation Consortium (IIPC)², à organiser une conférence internationale sur l'archivage du web, à travailler sur l'automatisation du dépôt et de l'archivage des publications gouvernementales en ligne, à mettre en valeur la documentation sur les méthodes d'archivage du web, et à effectuer une veille d'aide à la décision sur la conservation des ressources numériques. L'objectif est de faciliter la mise à disposition des informations relatives à ces domaines d'intérêt, et de rendre ces informations facilement accessibles via le portail thématique PADI (<http://www.nla.gov.au/padi/>).

¹ Alliance IFLA-CDNL pour les normes bibliographiques (n.d.t.)

² Consortium international de conservation de l'Internet (n.d.t.)

1. Introduction :

La Bibliothèque Nationale d'Australie est heureuse de participer à l'ICABS dans la mesure où elle s'est engagée dans l'étude du développement des normes et la rationalisation des coûts de l'accès aux collections patrimoniales du pays.

Notre principale contribution au programme de travail de l'ICABS concerne l'objectif trois : faire progresser la compréhension des questions relatives à l'archivage pérenne des ressources numériques.

Nous avons choisi de participer plus particulièrement à cet objectif parce que nous avons déjà une certaine expérience de l'archivage numérique, et que nous étions conscients du fait que de nombreuses bibliothèques nationales avaient besoin d'assistance dans ce domaine, ne sachant trop comment s'y prendre. Aussi, devant l'impossibilité de résoudre seul tous les problèmes, et afin de continuer à progresser nous aussi, avons-nous reconnu qu'il était indispensable de travailler en partenariat actif avec d'autres.

L'ICABS nous a incités à réfléchir mûrement à un partage plus efficient de nos connaissances et de notre expérience, et aux moyens qui nous permettraient de contribuer plus utilement à notre objectif qui est d'accéder de façon pérenne aux ressources numériques d'importance nationale.

2. Le contexte de l'archivage numérique et des actions de conservation à la Bibliothèque Nationale d'Australie :

L'archivage numérique

La Bibliothèque Nationale d'Australie collecte des sites Internet australiens significatifs pour les organiser au sein de PANDORA, base des archives Internet australiennes depuis 1996 (<http://www.nla.gov.au/pandora/>). Cette collecte est devenue une activité routinière de développement des collections, réalisée par une équipe de cinq personnes.

Les titres sont collectés avec la permission de leurs éditeurs puisque la loi de dépôt légal du Commonwealth australien ne s'applique pas encore aux publications électroniques.

Ces archives comprennent aujourd'hui environ 6000 titres et plus de 10000 réassemblages (ou extraits) de sites Internet, sa taille est d'environ un demi tera-byte. Le développement de PANDORA a été entrepris en collaboration avec les bibliothèques australiennes d'Etat et trois autres institutions nationales déjà impliquées dans la collecte de sites. La sélection des sites Internet est fondée sur des critères précis qui mettent l'accent sur la recherche et la valeur informative aussi bien que sur la portée culturelle, permettant également une sélection par sujet. La qualité du contenu et des fonctionnalités de chaque site Internet est contrôlée.

Afin de gérer et de maintenir ses responsabilités en termes d'archivage numérique, la Bibliothèque Nationale a développé son propre système d'archivage connu sous le nom de PANDAS³. Le système comprend les fonctions suivantes :

- gestion des métadonnées des titres sélectionnés et rejetés
- initiative du rassemblement des titres sélectionnés
- gestion du contrôle qualité
- préparation de l'affichage public des articles dans la base d'archives
- gestion des restrictions d'accès
- gestion d'une identification permanente
- fourniture de rapports de gestion

³ Système d'archivage numérique de Pandora (n.d.t.)

Plusieurs institutions collectrices dans le monde entier se sont renseignées sur la possibilité d'utiliser PANDAS dans la mesure où aucun système commercial équivalent de collecte et de gestion des ressources Internet n'existe sur le marché. Pour répondre à ces requêtes, la Bibliothèque Nationale a mis PANDAS à disposition, sur demande, pour effectuer des tests. Nous sommes également prêts à mettre le logiciel disponible afin que d'autres institutions puissent l'implanter, étant bien entendu que nous ne fournirons pas de service de maintenance.

La conservation numérique

Les réflexions stratégiques liées à la conservation ont été dès le début au cœur du développement des collections numériques de la Bibliothèque Nationale. Le but de rassembler des ressources Internet et de les archiver dans PANDORA est de les gérer en temps réel, afin de fournir un accès public continu. Ceci réclame à la fois une planification très en amont et une gestion dynamique quotidienne des ressources archivées. Une charte de conservation numérique (<http://www.nla.gov.au/policy/digpres.html>) et un plan d'action brossent les grandes directions stratégiques poursuivies par la Bibliothèque Nationale.

L'accent porte sur :

- La gestion des données (rafraîchissement, retour d'information, etc.)
- L'identification et la caractérisation des ressources archivées
- L'identification et la gestion des risques et menaces connexes aux différents formats de fichiers
- Le développement de politiques de conservation pour tous les formats de fichiers
- L'utilisation des métadonnées pour la conservation numérique
- La prévision des aléas

La collaboration est également essentielle à la méthode de conservation des documents numériques telle qu'entreprise par la Bibliothèque Nationale. Afin de faciliter ces collaborations, la Bibliothèque Nationale gère le portail thématique PADI⁴ (<http://www.nla.gov.au/padi>). PADI propose des liens vers un grand nombre d'informations liées à la conservation des documents numériques et fournit également une information contextuelle sur les mots clés. C'est la clef de voûte de notre contribution à l'ICABS, nous permettant de maximiser le partage d'information.

3. L'implication de la Bibliothèque Nationale d'Australie dans l'ICABS :

La Bibliothèque Nationale d'Australie contribue à deux axes du plan d'action de l'ICABS : le moissonnage Internet (action 3.3) et la conservation des documents numériques (action 3.4).

Le moissonnage Internet (action 3.3)

L'objectif de nos activités dans ce domaine est de cerner et promouvoir la mise en œuvre de méthodes de collecte et d'archivage des publications Internet, et d'identifier les problèmes associés, ceci afin de diffuser largement cette information et de la rendre facilement accessible via PADI, dans le but plus particulier de venir en aide aux institutions qui ont un projet de mise en place ou de restructuration de leur archivage numérique et de leurs politiques de conservation.

Les activités clés sont les suivantes :

⁴ Préservation de l'accès à l'information numérique

- Orienter le groupe de travail sur le web invisible du Consortium International de Conservation de l'Internet
- Organiser une conférence internationale sur l'archivage des ressources Internet
- Automatiser le dépôt et l'archivage des ressources gouvernementales en ligne
- Cerner et promouvoir la mise en œuvre des méthodes d'archivage des publications en ligne collectées par moissonnage.

(1) Orienter les travaux de groupe de travail sur le Web invisible du IIPC

Le Consortium international de conservation de l'Internet (IIPC) a été établi en 2003 afin de faciliter la collaboration entre bibliothèques nationales sur le développement des normes et des outils permettant l'archivage et la conservation des ressources Internet. Onze bibliothèques nationales ainsi que l'Archive Internet en sont membres et le Consortium est coordonné par la Bibliothèque nationale de France. L'objectif de ce consortium est avant tout concret. Les bibliothèques participantes paient une adhésion et sont obligées de contribuer activement au programme de travail de ce groupe.

La Bibliothèque Nationale d'Australie oriente le groupe de travail sur l'archivage du web invisible de l'IIPC, qui étudie l'identification, l'acquisition, le stockage et la mise à disposition des publications et des sites Internet articulés autour de bases de données. Cette catégorie de ressources Internet nous pose problème, d'autant plus qu'un nombre croissant de ressources importantes sont publiées sous cette forme. Pour des raisons techniques, nous sommes pour le moment dans l'incapacité de les collecter.

Les objectifs du groupe du web invisible sont les suivants :

- Identifier quelles sont informations sont requises par les éditeurs pour accéder aux bases de données déposées
- Développer un modèle de données et un schéma d'archivage des contenus d'une base de données en XML
- Développer un outil d'accès, de recherche et de navigation en ligne dans les bases de données structurées qui sont archivées.

Le logiciel qui est développé sera en accès libre et disponible à l'utilisation par quiconque s'y intéresse.

Les autres groupes de travail de l'IIPC se penchent sur des questions telles que les architectures logicielles requises pour les bases d'archives, l'identification des besoins des chercheurs dans l'usage des archives Internet, et le développement d'outils d'accès et d'un logiciel de recherche sur Internet⁵.

(2) La conférence internationale sur l'archivage des ressources Internet

Une conférence internationale sur l'archivage des ressources Internet et les problèmes qui se posent aux institutions collectrices se tiendra à la Bibliothèque Nationale à Canberra du 9 au 11 novembre cette année. Les événements suivants se dérouleront conjointement à cette conférence :

- Une rencontre du comité directeur de l'IIPC
- La fête nationale le 8, lors de laquelle seront présentés les projets relatifs aux ressources numériques de façon générale

⁵ crawler = type de programme utilisé par les moteurs de recherche pour trouver et indexer les pages de sites (<http://www.alaide.com/dico.php?q=crawler&page=dico&x=24&y=10>)

- Une journée d'information le 12, au cours de laquelle seront présentés les outils et méthodes d'archivage du monde entier

L'objectif principal de cette conférence est d'identifier les questions clés auxquelles sont confrontées les institutions culturelles dans la collecte de ressources Internet et de cerner la façon dont ces questions sont traitées sur le plan international. Une importance toute particulière sera accordée à l'étude systématique des méthodes adoptées dans les différentes institutions aux travaux de réflexion menés pour les appliquer. La conférence ne se veut pas technique, mais se concentre plutôt sur le cœur du métier et les besoins stratégiques de ces institutions de collecte.

Les informations recueillies lors de la conférence touchant aux préoccupations de l'ICABS sur l'archivage de l'Internet viendront celles déjà disponibles sur PADI.

(3) L'automatisation de la collecte et de l'archivage des ressources gouvernementales en ligne

Début 2003, la Bibliothèque Nationale a lancé le Projet pilote de base de données du Commonwealth dont les deux objectifs principaux sont :

- Favoriser la découverte des publications en ligne du gouvernement du Commonwealth australien et faciliter leur accès au travers de la bases de données bibliographiques nationale, disponible via le service suivant : <http://www.nla.gov.au/kinetica>
- Mettre en valeur le signalement et l'archivage de ces publications dans PANDORA ou dans une autre base d'archives publiques pour assurer un accès continu vers elles via le NBD

Un objectif sous-jacent est de revoir les normes des métadonnées obligatoires et de renforcer et diffuser leurs grandes lignes afin d'encourager les organisations gouvernementales à créer des métadonnées de qualité.

Le projet pilote implique un grand nombre d'organisations gouvernementales et il sera étendu avant 2005. Ses étapes seront les suivantes :

- Cerner les possibilités d'obtention des métadonnées produites par les agences partenaires (y compris le moissonnage)
- Convertir les métadonnées collectées du Dublin Core en Marc 21 pour le chargement dans le NBD par le biais de MODS
- Automatiser les opérations de moissonnage, de conversion, de chargement et de mise à jour des métadonnées
- Améliorer le logiciel PANDAS pour déclencher le moissonnage automatique et archiver les ressources Internet associées aux métadonnées (la finalisation est prévue pour 2005)

(4) Cerner et promouvoir la mise en œuvre des méthodes d'archivage des publications en ligne collectées par moissonnage

Des informations utiles à ce sujet sont déjà disponibles sur PADI et, comme nous l'évoquons plus haut, la conférence sur l'archivage de l'Internet permettra de connaître d'autres pratiques plus performantes pour le moment et leurs développements récents. En outre, une enquête sur les activités actuelles d'un échantillon d'agences culturelles sera entreprise pour s'assurer que celles-ci sont également représentées dans PADI. Afin de rendre plus facilement accessibles les informations disponibles sur PADI, celui-ci subira une restructuration en 2005. D'autre part, les fonctionnalités seront améliorées,

avec la réorganisation des ressources clés en fonction des thèmes les plus recherchés - y compris l'archivage numérique – et de rendre ces sujets (qui s'appelleront des « pistes »⁶) accessibles depuis la page d'accueil de PADI.

La conservation des documents numériques (Action 3.4)

L'objectif de nos activités dans ce domaine est de faire le point sur les normes existantes, les directives et les codes de conservation des documents numériques, en coopération avec la section préservation et conservation de l'IFLA.

La mise en place de programmes de conservation des documents numériques est un défi redoutable pour de nombreuses bibliothèques. Nos objectifs dans ce domaine sont d'identifier les éléments d'implication des bibliothèques et de déterminer les facteurs qui influencent leur réactivité et les modalités d'une assistance qui permette de rendre plus efficaces les programmes existants.

La fonction de PADI d'identification et de facilitation d'accès aux documents d'assistance est au cœur de cette action.

Comme indiqué plus haut, les structure et interface de PADI sont en cours de refonte, afin de rendre plus accessibles les éléments d'informations sur les sujets clés que sont les normes et les directives.

Les relations anciennes de la Bibliothèque Nationale avec d'autres bibliothèques des régions pacifique et du sud-est asiatique fourniront également d'excellentes occasions/opportunités pour déterminer le type d'assistance idoine dans une région vaste et diverse. La Bibliothèque est centre régional pour l'Océanie et l'Asie du sud-est dans le programme de préservation et de conservation de l'IFLA et elle a des relations de travail avec la campagne de conservation du patrimoine numérique menée par la division de la société d'information de l'UNESCO.

En novembre 2002, L'UNESCO a organisé une consultation régionale sur la conservation numérique en association avec la préparation des directives que la Bibliothèque Nationale d'Australie était chargée de rédiger. Sur la base des informations collectées, nous avons préparé en février 2004 une liste des facteurs de réactivité ainsi qu'un document d'assistance à l'amélioration de cette réactivité, pour un comité consultatif international du programme Mémoire du Monde de l'UNESCO, qui s'est tenu à Manille, aux Philippines. Ces facteurs de réactivité sont par exemple, une collection préexistante de ressources numériques, un accès aux connaissances et aux compétences en termes de technologies de l'information, et un personnel dédié à la gestion des collections numériques.

Ces documents seront revus à la lumière des résultats de l'atelier de travail et préparés en vue d'une diffusion plus large vers juin 2004. Ils seront bien entendu accessibles sur PADI.

⁶ « trails » (n.d.t.)