

	<p style="text-align: right;">Date : 23/07/2007</p> <p>Archivage d'Internet au niveau international : l'interopérabilité de l'avenir ?</p> <p>Grethe Jacobsen, Dr. Phil. Responsable du dépôt légal et des cartes et plans, Imprimés et photographies, Royal Library – Bibliothèque nationale du Danemark P.O. Box 2149, DK-1016 COPENHAGUE K, Danemark www.kb.dk</p> <p>Traduction juillet 2007 Rosa María Gómez de Regil Ingénierie documentaire CNDP France</p>
Meeting:	73 National Libraries
Simultaneous Interpretation:	No
<p style="text-align: center;">WORLD LIBRARY AND INFORMATION CONGRESS: 73RD IFLA GENERAL CONFERENCE AND COUNCIL 19-23 August 2007, Durban, South Africa http://www.ifla.org/iv/ifla73/index.htm</p>	

Résumé

Plusieurs bibliothèques nationales sont en train de collecter, ou planifient de le faire, une partie d'Internet. Cependant, il faut que les archives du web soient interopérables pour avoir une vision complète d'Internet. Cela permettrait aux utilisateurs de faire des recherches sans difficultés. Un questionnaire à ce sujet a été adressé à 95 bibliothèques nationales. Les réponses vont dans le sens de cette hypothèse et montrent aussi que l'archivage du web se généralise. La création de partenariats est un élément clé pour aller de l'avant. On suggère de distinguer les partenaires de conservation (archives, musées) et les partenaires techniques (entreprises privées, universités, autres institutions de recherche). Le travail avec les entreprises privées pourrait forcer les bibliothèques nationales à laisser place aux pensées non orthodoxes et aux expérimentations. Le défi majeur actuellement est d'adapter le dépôt légal, le copyright et d'autres législations au monde d'Internet, afin de le préserver et de le mettre à disposition des générations contemporaines et futures.

Contexte et questionnaire

Plusieurs bibliothèques nationales ont commencé à archiver certaines parties d'Internet, considéré actuellement comme une partie du patrimoine culturel national. Cependant, et à cause de la nature d'Internet, une archive nationale qui prend uniquement certaines sections d'Internet ne sera jamais capable de montrer une vision exhaustive du Web tel qu'il était utilisé par les internautes à un moment donné. L'objectif à long terme est de rendre

interopérables toutes les archives du web, ainsi, tout utilisateur d'une bibliothèque nationale (ou autre) pourrait faire des recherches dans toutes les archives du web sans difficulté.

La mise en place de cet objectif ne concerne pas uniquement les enjeux techniques et économiques. Elle concerne aussi la problématique du droit incluant le copyright, la publication illégale sur Internet des matériels et la protection des données personnelles, surtout quand il s'agit de donner accès à des ressources qui ont été collectés automatiquement.

Ces actions auront besoin d'une forte coopération entre les bibliothèques nationales et aussi de la création de partenariats afin de partager les coûts de développement du savoir-faire technique dans les domaines de la collecte, l'accès, le partage et la préservation des matériels en ligne. Les bibliothèques nationales ont besoin d'agir ensemble dans des actions de lobbying envers les législateurs afin de demander les permissions pour collecter et donner accès aux archives et d'établir des accords avec les détenteurs des droits d'auteur (copyright).

Pour faire le premier pas vers cet objectif, la bibliothèque nationale du Danemark, *The Royal Library*, a envoyé un questionnaire en mars 2007 aux listes de diffusion de la CDNL¹ et de la CENL² des bibliothèques nationales. Soit 95 bibliothèques.

Nous avons obtenu 23 réponses à la date limite d'envoi du questionnaire (le 27 mars) et 14 réponses en plus après la date, mais à temps pour les incorporer à cet article (taux de réponse de 39%). Toutes les réponses n'abordent pas ou n'étaient pas capables d'aborder toutes les questions, alors les réponses à certaines questions sont moins significatives. Le taux de réponse a été enrichi par les nombreux commentaires permis par le questionnaire qui démontrent que les répondants sont très impliqués dans cette affaire et sont motivés par l'archivage du web.

Dans la section suivante, les questions et commentaires seront discutés. Ils seront accompagnés d'annotations sur les pratiques danoises, pas vraiment pour promouvoir notre expérience mais pour clarifier ce que cela signifie à un niveau pratique.

Quelques mots sur l'expérience danoise en général : nous avons moissonné le domaine du plus haut niveau .dk depuis juillet 2005 jusqu'à ce que la nouvelle loi sur le dépôt légal prenne effet. Nous visons la conservation de la partie danoise d'Internet en tant que partie du patrimoine culturel pour les générations futures ; cependant, nous ne serons pas capables, tous seuls, de dupliquer en entier l'Internet de nous jours à moins que nous puissions donner accès aux autres parties du réseau dans d'autres archives de la toile.

Questions et réponses

La première question avait l'intitulé

1. Interopérabilité en général

et la question posée était :

Est-ce que vous êtes d'accord avec l'énoncé « L'objectif à long terme devrait être l'accès sans difficulté aux matériels d'Internet, passés et présents, pour tous les citoyens au travers des frontières nationales »

33 ont été d'accord avec cet énoncé alors que 4 ont dit « non ». Sept ont commenté la question. Il se trouve que seulement un répondant était contre l'énoncé alors que le reste pensait que l'objectif était louable mais peu réaliste ou au moins difficilement atteignable actuellement à cause des problèmes légaux et des principes directeurs variés autour des collections. Le commentaire donne plutôt lieu à un débat :

¹ Conference of Directors of National Libraries

² Conference of European National Librarians

« Je ne suis pas d'accord pour dire que l'accès a besoin ou doit se faire « sans difficulté ». La grande force d'Internet est son hétérogénéité. Je pense qu'une tentative pour imposer une solution homogène va étouffer l'innovation et sera finalement infructueuse. »

Je comprends cela comme une faible argumentation envers l'objectif fixé (accès sans difficulté), mais envers une forme de coopération qui pourrait ralentir les stratégies et les solutions innovantes. Le but principal de tout effort de coopération devrait être, par conséquent, d'assurer le développement en continu. Ce développement devrait laisser une place à la pensée non orthodoxe et à l'expérimentation.

Ceci représente un challenge pour les bibliothèques nationales, spécialement pour leurs directeurs. Ils devront affronter, d'un côté, les demandes des autorités publiques et les politiciens pour trouver des solutions maintenant et de produire quelques résultats pour montrer que l'argent disposé pour mener les actions de l'archivage du web est bien utilisé et, par ailleurs, pour ceux qui sont engagés dans la tâche de l'archivage du web, on demande qu'ils soient capables d'aller dans plusieurs directions en même temps, de parier sur des méthodes et avant tout de laisser la place aux expériences et erreurs.

La seconde question avait comme intitulé

2. L'archivage du web – actualité ou planification

Les questions sont les suivantes :

À ce jour, êtes-vous en train d'archiver le web ?

19 personnes ont répondu « oui » à la question, ils ont engagé des démarches d'archivage du web, alors que 18 personnes ont dit « non ». 11 de ces dernières sont en train de planifier l'archivage du web, et 7 personnes non. Cela signifie que l'archivage du web n'est plus une tâche réservée à un petit nombre d'acteurs comme c'était le cas il n'y a pas si longtemps.

Entre ceux qui sont en train de moissonner ou qui pensent le faire, 15 moissonnent (ou le planifient) le domaine national en entier, 27 collectent (ou pensent collecter) une sélection de sites parmi ceux du domaine national, 13 moissonnent aussi (ou pensent moissonner) en dehors du domaine national, pendant que 8 moissonnent (ou pensent moissonner) les sites par rapport à la langue.

Au Danemark, la loi du Dépôt Légal nous permet de moissonner les matériels publiés dans le domaine .dk et les matériels publiés dans d'autres domaines qui s'adressent à un public danois. Nous moissonnons le domaine national en entier ainsi que des sites web en dehors de ce domaine. Nous avons trouvé autour de 30 000 sites en dehors du .dk qui s'adressent à un public danois, ce sont principalement des sites avec du texte en danois, mais aussi de sites qui appartiennent à des entreprises ou institutions danoises, ou à des particuliers (ex. musiciens) qui sont domiciliés au Danemark.

Ensuite, nous avons demandé des informations à propos des activités liées à l'archivage du web, surtout si les répondants avaient créé des sites avec ce type d'information et si c'était le cas, nous avons demandé les URL. L'intitulé et les questions étaient les suivantes :

3. Information sur vos activités :

3.a Avez-vous un site web avec des informations à propos de vos activités sur l'archivage de la toile ?

3.b Avez-vous des principes directeurs sur la collecte des matériels provenant d'Internet ?

3.c Avez-vous des principes directeurs pour la découverte des sites incluant les sites importants ?

15 personnes ont affirmé avoir un site contenant des informations sur leurs activités. Un examen plus détaillé montre que la plupart d'entre eux ont une version anglaise du site ;

l'archivage du web est clairement vu comme un effort international, l'information ne s'adresse pas uniquement à un public local, mais aussi aux collègues et autres parties intéressées à l'international. 4 répondants ont des sites web dans leur propre langue et un a un site en allemand et en français.

Le partage de principes directeurs est une autre question. Tous les collecteurs actifs ont des principes directeurs ou sont en train de les créer.³ Cependant, uniquement 6 répondants mettent leurs principes directeurs en ligne. Cela peut être dû plus au manque de temps pour faire les traductions qu'à la volonté de ne pas mettre à disposition du public les principes directeurs.

Certainement, le premier cas est le Danemark. Nous maintenons un site bilingue www.netarkivet.dk (danois) et <http://netarkivet.dk/index-en.php> (anglais) et nous avons décidé de traduire et publier nos principes directeurs en anglais et en danois. Jusqu'à présent un seul principe directeur a été publié.

La quatrième question traitait des bases légales pour collecter les données si elles existent (ou d'autres réglementations gouvernementales) ou de l'accord volontaire.

4. Conseils pour la collecte de données

Vous collectez (cochez une ou plusieurs options, s'il vous plaît) :

- Selon la législation (spécifiez, ex. dépôt légal):
- Au travers d'un accord volontaire (spécifiez):
avec d'autres institutions
- Au travers d'un accord volontaire (spécifiez):
avec les communautés
- Au travers d'un accord volontaire (spécifiez):
avec des organismes ou personnes
du secteur privé
- Autre (spécifiez):

11 répondants ont collecté les données selon le dépôt légal, un selon d'autres mandats légaux, un selon la loi du copyright et un autre a fait une exception à la loi sur les données personnelles. Un répondant est en train de collecter le domaine national comme partie d'un projet. 4 répondants ont collecté selon les accords avec les éditeurs, autres communautés et l'Internet Archive et 1 a ne se prononce pas sur aucune de ces approches.

Il semble qu'il y a du chemin à parcourir pour les pays qui veulent collecter les données par le biais du dépôt légal ou un autre type de législation, mais plus de la moitié de ceux qui ont indiqué être engagés dans une démarche d'archivage du web le font avec par le biais d'une loi sur le dépôt légal, alors que trois autres ont différents types de mandats. Il faut noter qu'une bibliothèque active dans l'archivage du web, la bibliothèque nationale des Pays Bas, n'a pas de dépôt légal et par conséquent a une longue expérience de négociation avec les éditeurs pour déposer le patrimoine national soit en ligne ou dans son format physique.

La loi danoise du dépôt légal, permettant de moissonner les parties danoises d'Internet, a été passée en décembre 2004 et elle s'est renforcée en juillet 2005. Avant nous étions capables de collecter les publications entre 1998 et 2005 selon la précédente loi du dépôt légal (surtout entre 1998 et 2005). En plus, nous avons eu la permission du ministre de la culture pour faire du moissonnage sélectif entre 2001 et 2004 dans le cadre des différents projets de préparation à l'arrivée d'une nouvelle loi sur le dépôt légal.

Un enjeu important de l'archivage du web est son intégrité. Nous avons demandé, par conséquent, si les bibliothèques engagées dans l'archivage du web manipulaient l'archive en abandonnant certains matériels récupérés lors de la collecte et si l'on pouvait avoir des traces.

³ Une personne qui fait de l'archivage du web n'a pas répondu à cette question.

5. Manipulation du contenu de l'archive

5.a Quelles quantités du matériel moissonné dans l'archive gardez-vous ?

- Tout ce qui a été moissonné (spécifiez):
Seulement une partie des matériels moissonnés (spécifiez):

5.b Si vous renoncez à certains matériels, peuvent-ils être tracés ?

- Oui (expliquez):
Non

25 répondants ont dit qu'ils gardent tout ce qu'ils moissonnent, 1 répondant a dit que la bibliothèque conserve uniquement une partie des matériels moissonnés, à savoir, un nombre limité de fichiers. La limite n'a pas été spécifié dans la réponse, ni dans le site web de la bibliothèque. Il paraît que les bibliothèques font des efforts pour maintenir l'archive intacte.

L'accès est un élément de base dans la raison d'être d'une archive du web – ou de toute archive en général. S'il n'y a pas d'accès à l'archive, la collecte de données n'a pas de sens, étant donné que le but d'une archive du web est de documenter une partie du patrimoine culturel de la nation. En outre, l'accès devrait être proposé pour tous, ainsi que pour tous les matériels patrimoniaux d'un pays. Par conséquent, nous avons posé des questions à ce sujet :

6. Accès aux matériels archivés

Permettez-vous d'accéder en ligne à votre archive web ?

Si « **non** » répondez à la question suivante :

Accès limité	(spécifiez s'il vous plaît, ex. réservé à la recherche, objectifs statistiques)
Pas d'accès	

8 de ceux qui collectent des données ont répondu « oui » à la question « Permettez-vous d'accéder en ligne à votre archive web ? » Cependant, 3 de ceux qui ont dit « oui » ont remarqué, que techniquement, l'accès au grand public n'est pas encore possible. Le reste propose différents types de possibilités d'accès en ligne et hors ligne, tels que l'accès avec la permission des éditeurs (6), l'accès dans des locaux (4) et l'accès réservé aux chercheurs (4). Nous pouvons conclure que l'accès général en ligne aux archives de la toile n'est pas encore très répandu, à cause des aspects légaux (copyright et protection des données) ainsi qu'à des aspects techniques.

Au Danemark, nous permettons uniquement l'accès pour la recherche et avec des objectifs statistiques et uniquement aux chercheurs qui ont un niveau de post doctorat. La raison n'est pas le copyright, comme on pourrait le croire, mais la loi danoise sur le traitement des données personnelles. L'agence danoise pour la protection des données a déterminé que la collecte des matériels publics sur Internet et leur dépôt dans une archive du web, pourrait amener au traitement des données personnelles, qui est couvert par la loi sur le traitement des données personnelles. Nous allons commencer la négociation avec l'agence afin de permettre l'accès au moins à certaines parties des données collectées, par exemple, aux sites des institutions publiques. Notre premier but était de permettre l'accès public depuis les salles de lecture de la *Royal Library* et de la *State and University Library*, afin de rendre accessible cette partie du patrimoine culturel à tous les citoyens, comme c'est le cas pour les livres et d'autres types de publications qui sont disponibles pour tout le monde, soit en location ou pour leur utilisation dans les locaux de la bibliothèque.

L'échange des matériels moissonnés, des URLs collectés ou d'autres informations sur une archive du web pourrait intéresser d'autres archives et par conséquent, il s'agit d'un élément important dans la coopération internationale sur l'archivage du web. La question 7 traite cet aspect :

7. Matériels moissonnés sans problème

Êtes-vous capable de fournir des copies du matériel moissonné à d'autres archives du web ?

Si « **oui** » répondez à la question suivante :

Pouvez-vous fournir (cochez les cases appropriées) :

Copie des matériels moissonnés

Information sur les URL collectés et horodatage
pour le moissonnage

Autre information (spécifiez):

Uniquement 3 répondants ont été capables de fournir des copies des matériels moissonnés, 2 ne savaient pas et 15 ont répondu « non » à la question. 8 personnes sont capables de fournir des informations sur les URL collectées et de donner l'horodatage des matériels moissonnés et une personne n'était pas sûre de ce type de coopération internationale. En plus, 2 ont noté que les métadonnées et les notices de certains matériels moissonnés sont disponibles dans le catalogue en ligne.

Au Danemark, nous sommes incapables de fournir au grand public des copies et des informations sur ce que nous moissonnons ; nous fournissons ces informations aux chercheurs qui ont obtenu la permission d'accéder à l'archive. Nous n'avons aucune notice dans notre OPAC et nous ne planifions pas de cataloguer l'archive du web dans des notices sur l'OPAC. Nous avons des notices pour les publications que nous avons collectées entre 1998 et 2005 mais nous ne pouvons pas donner l'accès à ces publications pour le moment, parce qu'elles font partie de l'archive de la toile et elles sont couvertes par la loi danoise sur le traitement des données personnelles.

Un autre aspect de la coopération internationale est de s'entraider pour trouver des matériels qui font partie du patrimoine culturel d'un pays, mais difficiles à trouver sans l'aide de ses collègues. Nous avons demandé, par conséquent, si une coopération de ce type est possible.

8. Coopération avec d'autres archives du web

Seriez-vous en mesure de (cochez les cases appropriées) :

Donner des informations sur les URL de votre domaine national
qui intéresseraient d'autres bibliothèques nationales

Collecter les sites web de votre domaine national qui
intéresseraient d'autres bibliothèques nationales

Permettre à d'autres bibliothèques nationales et à leurs utilisateurs
d'accéder à votre archive

14 personnes ont répondu être en mesure de donner des informations sur les URL de leurs domaines nationaux qui pourraient intéresser d'autres bibliothèques nationales, 13 sont aussi en mesure de collecter les sites web du domaine, alors que 10 permettent l'accès à d'autres bibliothèques nationales.

L'archive du web du Danemark ne peut pas fournir des informations sur les URL issus de matériels moissonnés, ni collecter des sites web pour d'autres bibliothèques nationales. Nous pouvons, bien sûr, donner des informations sur les URL et les contenus publics dans le domaine « .dk » si nous possédons ce type d'informations.

En étroite connexion avec ces sujets de coopération, on trouve le choix de partenaires, sujet de la dernière question intitulée :

9. Partenariats pour le moissonnage des matériels Internet

Nous avons posé les questions suivantes :

9.a Moissonnez-vous des données en partenariat avec d'autres institutions ?

Si « oui » répondez à la question suivante :

9.b Travaillez-vous avec (cochez les cases appropriées) :

- | | | |
|--|--------------------------|---|
| D'autres bibliothèques nationales | <input type="checkbox"/> | |
| D'autres bibliothèques | <input type="checkbox"/> | |
| D'autres institutions publiques | <input type="checkbox"/> | Spécifiez le type (archive, musée, université, etc) : |
| Des communautés | <input type="checkbox"/> | Spécifiez s'il vous plaît : |
| Des entreprises / institutions privées | <input type="checkbox"/> | Spécifiez le secteur (TIC, commerce, édition, etc) : |
| Des particuliers | <input type="checkbox"/> | |

13 de ceux qui moissonnent des données ont dit « oui » à cette question et ils ont fourni plus de détails sur les partenariats.

4 travaillent avec d'autres bibliothèques nationales, 9 avec d'autres bibliothèques, 8 avec d'autres institutions publiques, 4 avec des communautés et 4 avec des entreprises ou institutions privées. Comme les chiffres l'indiquent, plusieurs des répondants travaillent avec plus d'un type d'institution, les bibliothèques étant un partenaire privilégié. Cependant, d'autres institutions publiques ou privés ont été mentionnées : Internet Archive, universités, archives nationales, archives du film et du son, musées, éditeurs, sociétés fiduciaires. À partir des commentaires, nous pouvons observer que ce qui guide les bibliothèques dans leur choix de partenariat c'est l'engagement des partenaires vis-à-vis de l'archivage du web ou leur savoir-faire technique.

Le fait d'avoir des partenaires implique aussi des complications par rapport à l'appartenance de l'archive et à la question :

9.c Comment avez-vous résolu les enjeux concernant l'appartenance de l'archive et de ses contenus ?

4 personnes ont dit qu'ils avaient partagé la propriété, alors que 7 ont affirmé que leur institution était propriétaire des matériels et 1 a dit qu'ils n'avaient pas résolu la question complètement ; 3 ont différents types d'accords avec les partenaires.

Parmi les coûts de construction et maintenance d'une archive web, il y a les coûts de développement, facteur important pour les bibliothèques dans le cadre du partenariat.

À la question

9.d Comment avez-vous résolu les problèmes techniques

- | | |
|-----------------------------------|--|
| Développement commun | <input type="checkbox"/> |
| Responsabilité de mon institution | <input type="checkbox"/> |
| Responsabilité de mon partenaire | <input type="checkbox"/> |
| Autre arrangement | <input type="checkbox"/> (spécifiez s'il vous plaît) |

6 ont répondu qu'ils ont engagés dans du développement commun avec les partenaires, alors que 4 institutions sont en charge du développement, 2 laissent cela aux partenaires et 1 repose sur le logiciel développé par l'IIPC (International Internet Preservation Consortium).

Au Danemark, deux des bibliothèques engagées dans l'archivage du web, la Royal Library (bibliothèque nationale) et la State and University Library, ont établi une institution virtuelle, « netarchive.dk » qui est gouvernée par un comité de pilotage de 6 membres (3 de chaque bibliothèque) afin de représenter l'expertise sur les technologies web, les TIC, le dépôt légal et la construction des collections. Le comité se réunit de 2 à 3 fois par an pour discuter et décider sur les aspects économiques, techniques, légaux et de gouvernance. Netarkivet.dk a un manager qui fait les rapports au comité de pilotage et supervise le travail quotidien. Le développement est en partie partagé entre les deux bibliothèques et par les partenaires de l'IIPC.

Nous avons également demandé à ceux qui ne travaillent pas avec d'autres institutions sur le moissonnage de répondre à cette question :

Si vous deviez vous engager dans des partenariats pour moissonner des matériels sur Internet, avec quel type d'institution voudriez-vous travailler et quelle est votre solution préférée par rapport aux enjeux qui traitent de l'appartenance de l'archive et de ses contenus et sur les aspects techniques ?

Actuellement, plusieurs de ceux qui font de l'archivage du web, ont aussi répondu à cette question. Les commentaires soulignent le besoin de coopération avec des institutions qui ont une expertise technique et sur les collections afin de proposer un engagement sur la conservation des données. Dans quelques cas, les bibliothèques font la distinction entre les partenariats techniques et les partenariats de conservation. Les partenaires techniques mentionnés sont les universités, les institutions de recherche et les entreprises privées ; les partenaires de conservation sont les archives, musées et d'autres répertoires de données reconnus. La bibliothèque nationale d'Australie a faite remarque que « plus le travail est partagé et mieux il peut être fait ».

Au Danemark, « coopération » et « partenariat » sont les mots clés de l'archivage du web. Nous participons aux activités de l'IIPC et aux initiatives et projets européens. Nous travaillons aussi de très près avec d'autres pays nordiques qui font aussi de l'archivage du web.

Conclusion

Tous les répondants soutiennent l'idée de l'archivage du web, néanmoins certains trouvent peu réaliste le fait d'archiver le réseau Internet en entier. Le défi pour les bibliothèques nationales sera de rendre cet objectif viable.

Les partenariats et la coopération proches sont des éléments clé pour avancer. Même si ce n'est pas nouveau pour les bibliothèques, la nouveauté est que les partenaires sont issus d'une grande variété de milieux, publics et privés. Il semble utile de distinguer les partenaires de conservation et les partenaires techniques. Pour trouver des partenaires de conservation il sera nécessaire de regarder dans les bibliothèques, les archives, les musées et d'autres institutions avec des fonctions de dépôt de données. Probablement, ces institutions seront toutes publiques. Les partenaires techniques peuvent être les mêmes que ceux de conservation. Les entreprises privées, les universités et des institutions de recherche sont le type dominant d'institutions citées. Le travail avec des entreprises privées pourrait avoir comme avantage que les bibliothèques seront forcées de garder à l'esprit et de planifier l'archivage du web afin qu'il soit accompagné par la recherche en continu et le développement qui doit laisser place à la pensée non orthodoxe et à l'expérimentation. Le défi pour les bibliothèques nationales sera de trouver les meilleurs partenaires pour les différentes fonctions.

Le challenge majeur pour les bibliothèques nationales en ce moment est cependant, d'adapter les lois sur le dépôt légal, le copyright et autres au contexte d'Internet, afin de conserver cette partie du patrimoine national et de le rendre disponible pour les générations actuelles et futures.

À la fin du questionnaire, nous avons demandé aux participants de commenter la problématique évoquée et le questionnaire lui-même. Un des commentaires fournit la conclusion parfaite de cet article et un encouragement à l'IFLA et aux bibliothèques nationales : « nous souhaitons que ce questionnaire nous aide à tous pour que l'IFLA soit capable de poursuivre les enjeux et les problèmes légaux liés à l'archivage du web. »