



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>

octobre 28, 2005

Code Number:

151-F

Meeting:

119 Acquisitions and Collection Development

Différents exemples de collections pour les utilisateurs

Notre patrimoine électronique comme matériau pour les utilisateurs : collecte et accès aux publications du net à la Bibliothèque nationale de Norvège

Kjersti Rustad

The National Library of Norway
Mo i Rana, Norway

Traduction : Christine Girard (Université Bordeaux 4, France)

Résumé :

La Bibliothèque nationale de Norvège est un centre multi media du savoir qui offre aux utilisateurs de la documentation sur tous les supports. Cette communication traite de la façon dont la Bibliothèque nationale collecte les publications d'internet, conformément à l'Acte du dépôt légal, et comment ces publications pourraient être mises à disposition de l'utilisateur final.

Une archive web de millions d'URL collectés sur plusieurs années devient un vrai défi quand il s'agit de la rendre accessible. Comment l'utilisateur final peut-il accéder aux publications du net dans cette Archive Internet pour les besoins de la recherche et de la documentation ? Cette communication examine les possibilités d'accès à Archive Internet, mais aussi les limites qu'impose la législation existante.

1. Introduction

Un chercheur travaillant sur l'histoire récente de la presse norvégienne a récemment interrogé la Bibliothèque Nationale pour obtenir des informations. Il menait une étude comparative sur

la présentation des nouvelles dans la presse imprimée, à la radio et dans la presse sur Internet et souhaitait accéder aux documents source.

Cet épisode peut illustrer une demande typique de l'utilisateur final. Se pose alors la question : la Bibliothèque Nationale peut-elle fournir ce type d'information ?

Dans ses principes, la Bibliothèque Nationale de Norvège est la mémoire de la nation, un centre multimedia du savoir. La Bibliothèque Nationale doit conserver et mettre à disposition l'héritage culturel norvégien, et enregistrer les documents relatifs à la vie culturelle et sociale norvégienne pour les utilisateurs actuels et futurs.

Donc la réponse à la question est oui. Il est prévu que la Bibliothèque Nationale fournisse à l'utilisateur final le document source quel que soit le medium.

Notre principal outil pour remplir cette mission est l'Acte sur le Dépôt légal.

Quand l'actuelle loi sur le dépôt légal est entrée en vigueur en 1990, la Norvège a été parmi les premiers pays à inclure les documents numériques dans sa législation. La loi est basée sur le principe que toute information disponible, quel que soit sa forme et son support, doit être conservée et rendue accessible pour les besoins de la recherche et de la documentation.

Bien qu'Internet tel qu'on le connaît aujourd'hui n'existait pas en 1990, les documents accessibles sur Internet sont soumis au Dépôt légal conformément à cet Acte.

C'est ce qui fonde l'action menée par la Bibliothèque Nationale pour collecter ce qui est publié sur Internet.

2. Collecte des publications du net

Le moissonnage des publications sur internet a débuté sur des échantillons au milieu des années 1990 et le premier moissonnage portant sur la totalité du web norvégien a été mis en œuvre en décembre 2002. C'est cependant cette année, en 2005, que la Bibliothèque Nationale a débuté le moissonnage régulier du domaine Internet norvégien.

Notre principe de sélection consiste à moissonner tous les documents largement accessibles de l'espace web norvégien, c'est-à-dire le domaine.no. Puisque le Norwegian Act s'applique à tous les documents généralement accessibles, qu'ils soient produits par un éditeur norvégien ou destinés au public norvégien, les documents d'autres domaines comme .org ou .com peuvent aussi relever du dépôt légal. Donc nous prévoyons à plus long terme d'inclure les autres domaines. Nous prenons aussi en compte les sites web avec robots.txt. En effet notre moissonneur ignore le protocole robots.txt.. Ceci a fait l'objet d'une longue discussion.

Comme vous le savez, robots.txt est un protocole standard qui est utilisé pour s'assurer que le moteur de recherche ne récupère pas un site web ou des documents que le webmestre considère inadaptés. La règle communément admise sur Internet est de respecter robots.txt. Mais, dans la mesure où il s'agit de la Bibliothèque Nationale et de notre mission de préservation exhaustive, nous devons ignorer robots.txt pour nous assurer que d'importants documents sont bien conservés. Notre expérience nous a montré que si nous suivions robots.txt, nous risquons de perdre des sites web entiers. Ce qui irait à l'encontre de l'Acte du dépôt légal.

Cette question a, toutefois, provoqué de longues discussions avec les éditeurs web et la Bibliothèque Nationale a dû défendre sa pratique et expliquer que notre moissonnage a un fondement légal qui le différencie des autres moteurs de recherche.

Nous effectuons le moissonnage avec Heritrix,¹ un moteur du web développé par Internet Archive² en collaboration avec divers partenaires dont la bibliothèque nationale norvégienne. Notre première expérience de moissonnage du net avec Heritrix a débuté en février de cette année et notre projet initial était de moissonner la totalité du domaine .no quatre fois par an. Notre calcul était que chaque session de moissonnage devait durer environ 3 semaines et collecter 10 millions de documents (URLs).

Nous avons toutefois vite réalisé que ces prévisions étaient complètement dépassées. Au bout de 3 semaines prévues, le moteur avait ramené au moins 3 fois plus de documents et le total augmentait d'heure en heure.

Moissonner le domaine Internet en son entier assure la préservation de photographies de ce qu'est le domaine Internet norvégien à un moment donné. C'est une source importante pour l'utilisateur final. Mais il est également important de conserver l'historique complet d'un document particulier : par exemple, tous les numéros d'un magazine, tous les rapports d'une série, et les mises à jour quotidiennes de la presse. Cette approche sélective complète le moissonnage exhaustif des sites norvégiens. Elle constitue le deuxième niveau et le deuxième objectif de notre entreprise de moissonnage. La récupération quotidienne de tous les journaux norvégiens est prévue à partir de 2005. Nous prévoyons également de débiter le moissonnage régulier de tous les périodiques norvégiens en 2005.

3. Coopération internationale

La Bibliothèque Nationale de Norvège participe à plusieurs projets internationaux d'archivage d'Internet. Elle fait partie du consortium international pour la préservation d'Internet et coopère avec les autres bibliothèques nationales ainsi qu'avec Internet Archive aux Etats-Unis³.

En outre la bibliothèque participe au projet nordique d'archivage du web, forum des bibliothèques nationales des pays nordiques qui coordonne les actions de moissonnage et d'archivage de documents sur Internet.⁴

Bien que l'intitulé du forum semble indiquer que nous ayons un web d'archivage commun, ce n'est pas le cas. Les bibliothèques nationales des pays nordiques font des récupérations séparées et ont des sites web d'archivage distincts. Toutefois le projet NWA a développé un outil d'accès commun intitulé « NWA Toolset ».

4. Préservation

La Bibliothèque nationale conçoit la préservation à un horizon de 1000 années. Ceci veut dire que notre objectif est que l'information contenue dans les documents de tous formats soit accessible dans mille ans. Seuls les utilisateurs du futur pourront dire si c'est un objectif réaliste ou non, mais c'est notre but et le principe selon lequel nous organisons nos actions de

¹ Pour plus d'information sur Heritrix voir : <http://crawler.archive.org>

² pour plus d'information sur Internet Archive voir : <http://www.archive.org>

³ Pour plus d'information sur IIPC voir : <http://www.netpreserve.org/about/index.php>

⁴ Pour plus d'information sur le projet NWA voir : <http://nwa.nb.no>

préservation. Les publications internet moissonnées par Hérित्रix sont archivées dans le Centre de préservation à long terme des publications électroniques de la Bibliothèque nationale.

5. Accès

5.1 Métadonnées

Comment l'utilisateur accède –t-il aux publications internet archivées pour ses besoins de recherche et de documentation ? Il y a au moins deux aspects à prendre en compte à propos de cette question : le métadonnées et la législation.

La bibliothèque nationale décidera dans le courant de l'année jusqu'à quel point les documents archivés doivent être catalogués et inclus dans la bibliographie nationale.

La nécessité de contrôle bibliographique, la façon de rendre ces documents accessibles ainsi que la masse des documents conservés sont des éléments à prendre en compte. Cependant le catalogage des documents internet archivés ne s'appliquera sans doute qu'à une faible part des documents.

Si l'on réalise que probablement moins de 1% des documents internet collectés du domaine norvégien feront l'objet d'un enregistrement bibliographique, nous devons avoir d'autres outils d'accès aux archives du web. La solution est la suivante : Tous les documents des archives internet seront entièrement indexés en utilisant le logiciel d'indexation FAST. Cela permettra aux utilisateurs d'effectuer une recherche libre. Les documents seront accessibles par NWA Toolset. Ce logiciel permettra d'accéder à plus de 99% des documents conservés. Pour illustrer la façon dont l'archive internet apparaîtra à l'utilisateur, les exemples qui suivent présentent les affichages NWA Toolset, qui est le principal logiciel d'accès aux archives internet norvégiennes.

L'exemple 1 présente l'interface utilisateur recherche avancée de NWA Toolset. Les modalités de recherche comprennent la recherche plein texte avec les opérateurs booléens et la recherche par certains URLS. Les recherches peuvent être limitées par date, langue, et site.

L'exemple 2 montre l'exemple d'une recherche sur l'intitulé « jazz », limitée au site de la Bibliothèque Nationale de Norvège. La liste des réponses peut être présentée chronologiquement ou générale.

L'exemple 3 nous montre une présentation chronologique de l'historique du document, avec une fenêtre comportant une ligne chronologique. Comme vous le voyez, c'est la version 5 de 5 chargements d'un document spécifique, effectuée le 15 janvier 2005 à 12h20. On peut choisir, année, mois, jour, heure et minutes dans la ligne chronologique. Cela donne à l'utilisateur la possibilité d'un voyage virtuel dans le temps.

L'exemple 4 nous montre l'affichage de l'historique du document à travers le récapitulatif des différents enregistrements.

5.2. Législation

Moissonner le domaine Internet norvégien permet à la Bibliothèque Nationale de préserver notre héritage électronique. Cependant, la législation norvégienne actuelle limite les possibilités d'accès du public à ces archives Internet.

Pour donner à l'utilisateur final l'accès aux archives Internet la Bibliothèque Nationale doit prendre en compte à la fois les règles du dépôt légal et celles qui concernent les données personnelles et le copyright.

L'Acte du dépôt légal prescrit que les documents déposés selon l'Acte doivent être accessibles pour les besoins d'étude et de recherche. Ceci signifie que l'Acte apporte des restrictions claires quand il s'agit de donner accès aux publications Internet. C'est le but de la recherche qui détermine l'autorisation d'accès.

Ceux qui pourront avoir accès aux archives Internet sont par exemple les chercheurs, les savants, les étudiants, les enseignants et d'autres utilisateurs spécialisés, par exemple dans un domaine spécifique comme la généalogie.

Les publications du net, qui sont mises à disposition des usagers, peuvent contenir des données personnelles et des données personnelles sensibles. Le traitement légal des données personnelles est défini par l'Acte sur les données personnelles. L'Acte doit permettre de vérifier que les données personnelles sont produites en accord avec le respect fondamental du droit à la vie privée, incluant la nécessité de protéger l'intégrité de la personne et la vie privée. La possibilité de traiter les données personnelles dans les archives Internet grâce aux outils de recherche rend les informations beaucoup plus accessibles dans les publications sur Internet que dans les documents traditionnels.

Ceci implique que la Bibliothèque Nationale doit avoir une autorisation de l'institution norvégienne d'inspection des données pour conduire ses opérations de moissonnage. En outre, une autorisation similaire est requise pour donner aux utilisateurs l'accès aux publications Internet.

Selon l'Acte du copyright, la Bibliothèque Nationale ne peut rendre accessible au public les documents si cela s'oppose à leurs droits de propriété économique ou intellectuelle. Selon la législation courante, la Bibliothèque Nationale ne pourrait donner accès aux publications Internet que sur place, c'est-à-dire sur les ordinateurs installés dans les locaux de la bibliothèque.

Pour résumer, voici les règles qui doivent être suivies selon l'actuelle législation avant de donner accès aux publications des archives Internet

- L'utilisation des documents doit se faire dans un but de recherche et de documentation
- Une autorisation générale de l'inspection norvégienne des données relative aux données personnelles doit être signée
- Les usagers doivent effectuer leur recherche depuis les locaux de la Bibliothèque Nationale.

Il faut donc reconnaître que la législation existante donne un accès moindre ou en tous cas plus restreint aux publications électroniques qu'aux publications imprimées déposées légalement.

Les propositions de changement de l'Acte norvégien du copyright incluent la possibilité d'effectuer le prêt entre bibliothèques des documents électroniques en réponse à des

demandes spécifiques. La Bibliothèque Nationale enverrait une copie d'un document électronique à la demande d'un utilisateur d'une autre bibliothèque, cette dernière étant responsable de la destruction de la copie après usage.

Si l'on considère le fait que la Norvège est un pays comportant de grandes distances géographiques où la population est dispersée, limiter l'accès aux archives internet à une utilisation dans les services de la Bibliothèque nationale n'est pas la meilleure solution.

Nous pouvons identifier 4 niveaux d'accès pour l'utilisateur final :

- L'accès sur site pour les usagers via les postes informatiques de la Bibliothèque nationale.
Dans la perspective des modifications du Copyright Act norvégien, cet accès sera autorisé.
- L'accès sur site par les usagers via les postes informatiques des universités, des collèges et des bibliothèques publiques. Ce n'est pas permis par la législation existante mais offrirait à bien des égards une bonne solution.
- Accès par les usagers via des ordinateurs privés
- Mise à disposition des documents électroniques par exemple sur le site web de la Bibliothèque nationale, ceci nécessitant un contrat de licence.

L'accès aux archives internet présenté des points 1 à 3 doit être limité aux besoins de recherche et de documentation. Ceci signifie que l'accès libre à tout public ne peut être autorisé selon la législation actuelle. Toutefois il existe des façons de donner accès aux archives internet qui résolvent le problème pour une part de ces archives.

La Bibliothèque nationale peut signer des accords ou des contrats de licences avec les titulaires du copyright afin de mettre les publications à disposition du public, par exemple via le site de la Bibliothèque.

Les propositions de modification du Copyright Act incluent la possibilité de passer des contrats de licence pour les publications électroniques entre les organismes gérant le copyright et les bibliothèques. La Bibliothèque pourrait ainsi rendre les documents électroniques accessibles aux utilisateurs mais ceci implique un paiement de droits.

Le premier objectif de la Bibliothèque nationale est d'offrir à l'utilisateur final l'accès sur site via le poste de la Bibliothèque Nationale. Il s'agit toutefois d'un accès limité géographiquement. Les modalités d'accès futures aux archives web sont à l'étude.

a. A quels documents l'utilisateur a-t-il accès ?

En règle générale, toutes les publications internet relevant du dépôt légal devraient être mises à disposition des utilisateurs pour les besoins de la recherche et de la documentation, l'accès à un utilisateur spécifique étant limité à la part des archives internet correspondant à sa recherche. Quand il s'agit de documents qui ne sont pas protégés par le copyright act comme les textes légaux, la réglementation administrative, et autres décisions des autorités publiques, il ne devrait y avoir aucune limitation d'accès.

6. Conclusion

D'ici quelques années, les archives internet norvégiennes constitueront des billions d'URL parmi lesquels la Bibliothèque nationale permettra à l'utilisateur de naviguer. Le projet de moissonner, préserver et donner accès aux documents du web est énorme, mais la bibliothèque nationale pense que nous possédons l'infrastructure, les services et l'organisation pour y parvenir.

D'un point de vue technique les publications du net peuvent être mises à disposition du public facilement. L'accès est cependant limité par la législation existante, et il est vital pour la Bibliothèque nationale de s'assurer que tout usage des archives internet est conforme à l'Acte de dépôt légal, à l'Acte du copyright, et à l'Acte des données personnelles.

Cependant, ceci ne doit pas empêcher les archives internet, en tant que témoignages de la vie culturelle et sociale norvégienne, d'être mises à disposition des usagers pour les besoins de la recherche et de la documentation.