



**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:	009-F
Meeting:	89.Cataloguing
Simultaneous	-
Interpretation:	

Le Projet PARADIGMA : en quête de solutions par les métadonnées et de services aux utilisateurs

Carol Van Nuys, Ketil Albertsen, Linda Pedersen et Asborg Stenstad

The Paradigma Project, Bibliothèque nationale de Norvège

carol.vannuys@nb.no, ketil.albertsen@nb.no, linda.pedersen@nb.no, asborg.stenstad@nb.no

Résumé :

Le Projet Paradigma de la Bibliothèque nationale de Norvège cherche les moyens d'assurer un dépôt légal satisfaisant de tous types de documents numériques – y compris les millions de documents trouvés sur le domaine Internet norvégien. La Norvège espère être en mesure de préserver son patrimoine culturel numérique pour la postérité, en donnant aux chercheurs la possibilité d'accéder à un archivage d'Internet, au moyen notamment de métadonnées et de recherche plein texte. Le présent article donne une rapide description du projet lui-même, avant d'aborder les problèmes rencontrés dans sa quête de normes de métadonnées pour la découverte, la préservation à long terme, etc. On présente l'usage qui est fait dans le projet des niveaux d'entités FRBR œuvre, expression, manifestation et item dans la conception de l'archivage, ainsi que des idées de services à venir : un service de vérification et d'authentification et un service d'attribution d'identifiants – l'un et l'autre disponibles via Internet.

1 Introduction

1.1 L'archivage Web dans d'autres pays

Des documents numériques disparaissent quotidiennement, et une étude¹ montre que 20 % seulement des documents trouvés sur le net y sont encore – sans changement – au bout d'un an. Partant, de nouvelles générations de lecteurs n'auront plus la possibilité d'étudier dans l'avenir les documents numériques d'aujourd'hui. La conservation de notre patrimoine culturel numérique est une question de plus en plus cruciale et un défi, et la Bibliothèque nationale de Norvège² n'est que l'une des nombreuses institutions qui travaillent systématiquement pour trouver des réponses aux problèmes juridiques, techniques et bibliographiques dont elle s'accompagne.

Un aspect du travail de conservation numérique réside dans la collecte et l'archivage des documents des domaines Internet nationaux. Différents pays ont opté pour des approches différentes : le Danemark [1] et l'Australie [2] ont opté pour l'approche sélective, tandis que la Suède [3], l'Islande et la Finlande ont recueilli l'intégralité de leurs espaces Web nationaux. La Norvège fait partie des quelques bibliothèques européennes qui recueillent et archivent les documents numériques de leur domaine Internet national sur la base d'une législation existante sur le dépôt légal³.

1.2 La Loi sur le dépôt légal

L'objectif de la Loi sur le dépôt légal [5] est de :

« [...] faire en sorte que les documents qui contiennent de l'information mise à la disposition générale soient déposés dans des fonds nationaux, afin que ces témoignages de la vie culturelle et sociale norvégienne soient conservés et mis à disposition en tant que source de première main aux fins de recherche et de documentation. » (§ 1)

Tenue pour innovante lors de sa promulgation en 1989, l'actuelle Loi sur le dépôt légal couvre tous les documents norvégiens *mis à la disposition générale* sur *quelque support que ce soit* : papier, microforme, photographie, assemblage de divers supports, enregistrement sonore, film, vidéo, documents numériques et programmes radio- et télédiffusés. Les documents publiés à l'étranger pour des éditeurs norvégiens et ceux qui sont spécialement adaptés pour un public norvégien sont également couverts.

Bien sûr, le World Wide Web n'existait pas encore sur Internet en 1989. Les documents numériques – essentiellement sous forme de bases de données – étaient peu nombreux, en comparaison des millions de publications que l'on trouve aujourd'hui sur Internet, mais ils présentaient déjà des difficultés techniques de traitement. À ce jour, le Dépôt pour la Préservation à Long Terme de la Bibliothèque nationale a une capacité de stockage de 100 TBytes de données ; soit un nombre plus que confortable de documents numériques.

2 Le Projet Paradigma

La Bibliothèque nationale de Norvège a lancé le Projet Paradigma⁴ en août 2001. Le but du projet est d'assurer une couverture suffisante au dépôt légal des documents numériques norvégiens, ce qui comprend l'élaboration de techniques, de méthodes et d'opérations de routine pour choisir, engranger,

¹ Manerheim, Johan. *The WWW and our digital heritage* [en ligne]. – URL : <http://ifla.org/IV/ifla66/papers/158-157e.htm> (réf. du 15 avril 2004).

² Pour de plus amples informations sur la Bibliothèque nationale de Norvège, voir l'URL : http://www.kb.nl/gabriel/libraries/pages_generated/no_en.html (réf. du 15 avril 2004).

³ Halgrímsson, Torsteinn (28 février 2003). Web archiving in Europe [discussion]. – *NWA* [en ligne]. – Adresse e-mail : nwa@nb.no.

⁴ Pour de plus amples informations sur le Projet Paradigma, voir l'URL : http://www.nb.no/paradigma/eng_index.html (réf. du 15 avril 2004).

décrire et identifier *tous* types de documents numériques – y compris ceux qui sont mis à la disposition générale sur Internet. Le projet vise en outre à offrir aux utilisateurs l'accès à l'archivage d'Internet en conformité avec la législation actuelle.

Les activités directement liées au Projet s'appuient sur les travaux antérieurs de la Bibliothèque dans divers domaines connexes, et occupent quatre ETP. Une trentaine de personnes sont en outre impliquées dans des activités plus ou moins liées au Projet. Il est prévu que le Projet prenne fin au 31 décembre 2004.

Les sections qui suivent décrivent brièvement les travaux en cours visant à permettre la sélection, la collecte et l'accessibilité des documents obtenus par dépôt légal sur Internet, ainsi que la nature et la taille du domaine Internet norvégien.

2.1 Stratégies de collecte et de sélection

2.1.1 Collecte

S'appuyant sur la Loi sur le dépôt légal ainsi que sur les recommandations du Projet Paradigma, la Bibliothèque nationale a décidé de commencer sa collecte par *tous* les documents numériques *mis à la disposition générale* dans l'espace Web norvégien (« .no »). Au fil du temps, les documents trouvés sur des domaines dont le nom se termine par « .com », « .org » et « .net » seront également collectés.

Plusieurs raisons justifient cette approche : tout d'abord, nous ne pouvons savoir à l'avance quels seront les documents qu'on jugera importants pour la recherche et la documentation dans l'avenir ; ensuite, les coûts de stockage numérique diminuent constamment ; troisièmement, la collecte non filtrée permet d'économiser au moment de la collecte une dispendieuse sélection manuelle ; enfin, un utilisateur d'archives Internet peut trouver des documents par une recherche en texte libre, ce qui lui permet de passer en revue tous les documents, y compris ceux qui n'appellent pas de catalogage manuel. Les critères de choix pour quelque utilisation que ce soit, comme l'approfondissement de la description bibliographique, peuvent aussi être remis en question et modifiés au cours du temps. Ce qui ne serait plus possible si les documents étaient écartés au moment de la collecte.

La Section de Dépôt légal collecte semi-manuellement un choix de documents Web depuis 2001, à l'aide du logiciel HTTrack⁵, et ces documents sont catalogués dans le catalogue BIBSYS⁶ de la Bibliothèque nationale. Cette activité se poursuivra jusqu'à ce que les activités de collecte générale du Projet Paradigma et les procédures qui y sont associées soient pleinement établies. Cette même section effectue également une collecte en rapport avec un événement : elle a par exemple collecté les sites Web de partis politiques avant, pendant et après les élections. D'autres sections sont aussi engagées dans les activités du dépôt légal numérique, et les Archives du Son et de l'Image de la Bibliothèque cherchent des solutions pour traiter la question du dépôt légal des programmes de radio et de télévision « natifs numériques », en collaboration avec la Société norvégienne de radio- et télédiffusion.

Le dépôt du *Web profond*, par exemple les journaux sur Internet, les médias à flux continu, les documents produits par webcam, les médias interactifs et documents numériques de tous types stockés dans des bases de données, représente un enjeu particulièrement délicat. Le Projet Paradigma a lancé la collecte quotidienne d'environ 65 journaux sur Internet, et très prochainement plusieurs bases de données de journaux en ligne seront intégralement déchargées, ce qui complétera les « instantanés » quotidiens. Nous discutons les problèmes liés au Web profond dans le cadre du *Consortium international pour la conservation d'Internet*⁷, mais un grand nombre de questions d'ordre administratif, juridique et technique restent pour le moment sans réponse.

⁵ Pour de plus amples informations sur le logiciel HTTrack, voir l'URL : <http://www.httrack.com> (réf. du 15 avril 2004).

⁶ Pour de plus amples informations sur BIBSYS, voir l'URL : <http://www.bibsys.no/english.html> (réf. du 15 avril 2004).

⁷ Pour de plus amples informations sur les activités liées au Web profond, voir l'URL : <http://www.nla.gov.au/ntwkpubs/gw/66/html/p15a01.html> (réf. du 15 avril 2004).

En résumé, la Bibliothèque nationale de Norvège pourra recevoir des objets numériques par le biais de plusieurs canaux : collecte documentaire automatisée sur Internet, mises à jour de bases de données fournies par batch, périodiques et listes de diffusion à abonnement distribuées par e-mail, groupes de discussion NetNews et documents fournis sur support physique tels que les cédéroms.

2.1.2 Sélection

Il y a de nombreux documents valables sur Internet, et nous sommes en train de travailler à la définition de *critères de sélection* pour les documents dont nous sentons qu'ils « méritent » une description bibliographique manuelle plus ou moins détaillée. Ces critères de sélection reposent sur la législation du dépôt légal ainsi que sur la politique documentaire générale de la Bibliothèque telle qu'elle est formulée dans notre plan stratégique. Les critères de sélection pour les documents numériques rejoignent ceux qui sont formulés pour des types de documents plus traditionnels dans le Guide pour la Sélection établi par la Bibliothèque.

Le Projet Paradigma envisage d'implémenter une architecture de système qui permette un processus de *sélection* en trois phases, afin que les bibliothécaires reçoivent une aide technique pour trouver les quelques documents qui devraient être catalogués de manière plus ou moins détaillée. La première phase consiste à trouver et collecter les documents norvégiens et sames sur Internet. La seconde phase donne aux bibliothécaires l'occasion de produire *automatiquement* des listes pondérées sur la base de requêtes spécifiques. Ces listes s'appuient sur l'utilisation de vecteurs comportant des métadonnées automatiquement extraites des documents collectés. Dans la troisième phase, les bibliothécaires choisissent des documents spécifiques dans les listes pondérées en vue d'un enregistrement manuel plus ou moins détaillé, en faisant appel aux critères de sélection mentionnés ci-dessus. Un jour, nous pourrions peut-être aussi maîtriser les ressources intégratrices qui ont été cataloguées manuellement, aidant ainsi les bibliothécaires à découvrir et modifier ces notices bibliographiques, par exemple à des intervalles de temps déterminés, ou quand les modifications apportées au texte dépassent un pourcentage déterminé, etc.

2.2 Le domaine Internet norvégien

La taille exacte du domaine Internet norvégien est encore inconnue à ce jour. La première collecte menée par le Projet Paradigma autour de décembre 2002/janvier 2003 a donné environ 3,1 millions d'URL (c'est-à-dire de fichiers), dont une quantité d'environ 53 % consiste en images (.jpg, .gif, .png). Le NEDLIB-harvester⁸ a commencé avec environ 1 000 URL de démarrage, et la collecte a été limitée au protocole HTTP, au domaine national norvégien (« .no »), et aux URL non paramétrées. La seconde campagne de collecte a eu lieu en août 2003, et a donné environ 4,1 millions d'URL. La troisième campagne de collecte est actuellement en cours, et il n'y a pas de statistiques disponibles pour l'instant.

En supposant une distribution comparable à celle qu'ont donnée les campagnes de collecte menées en Suède et en Finlande, nous nous attendons à trouver entre 45 et 55 % de sites Internet norvégiens n'ayant pas l'extension « .no ». Il va sans dire que le traitement manuel et l'évaluation de chaque objet n'est pas possible ; l'immense majorité doit passer par des procédures automatisées.

2.3 Stratégie d'accès

2.3.1 *Qui va chercher quoi* dans nos archives ?

Lorsque l'on cherche à trouver des solutions par les métadonnées pour décrire la documentation numérique riche et variée telle que celle qui est stockée dans nos archives, il faut se poser la question suivante : qui va utiliser cette documentation, et dans quel but ? Il est difficile d'imaginer ce que seront les

⁸ Pour de plus amples informations sur le NEDLIB-harvester, voir l'URL : <http://www.csc.fi/sovellus/nedlib/ver11/documentation11.doc> (réf. du 15 avril 2004).

questions spécifiques des chercheurs dans 10, 20 ou 50 ans, mais nous pouvons tenter d’imaginer certains *groupes d’utilisateurs* et certains *types* de questions.

Il peut y avoir un groupe constitué d’utilisateurs s’intéressant à l’étude d’Internet et de la documentation numérique en tant que *médium*, c’est-à-dire, pour la raison même que cette documentation a été obtenue d’Internet et qu’elle présente les caractéristiques de ce médium. On peut envisager que certains utilisateurs aient besoin d’étudier la langue employée sur le net et les relations entre différentes formes de langage ; les médiologues pourraient avoir envie d’étudier les relations entre imprimé et numérique, ou entre les tendances de l’évolution technologique et le contenu ; les webgraphistes peuvent s’intéresser à l’utilisation de la publicité, à la mise en page, etc. ; les chercheurs en informatique peuvent étudier divers protocoles de communication, l’utilisation des formats au cours du temps et même les virus ; les sociologues peuvent s’intéresser aux influences réciproques entre société et information disponible sur Internet. Bien sûr, il peut y avoir aussi des recoupements entre ces différents domaines d’intérêt.

Un autre groupe d’utilisateurs pourrait n’avoir besoin de ces documents numériques que comme *matériau primaire* – exactement au même titre que les sources documentaires traditionnelles d’aujourd’hui. Dans ce groupe on trouvera certainement des chercheurs de tous les horizons, aussi est-il intéressant de découvrir leurs attentes concernant la documentation numérique en particulier. Les sources pertinentes sont-elles disponibles au format numérique uniquement ? Importe-t-il que le contenu soit dynamique, qu’il y ait des animations, un affichage interactif, du son et de la vidéo intégrés, etc. ? Les chercheurs ont-ils besoin d’avoir accès à la documentation par une recherche en texte libre ou bien en mettant en relation de grandes quantités d’informations puisées à diverses sources ?

2.3.2 État actuel de la législation

Donner à des utilisateurs l’accès aux archives Internet du dépôt légal est une question complexe, et la Bibliothèque nationale doit trouver des solutions satisfaisantes, malgré les nombreuses (et parfois contradictoires) dispositions que renferment la Loi sur le Dépôt légal, la Loi sur le Copyright et la Loi sur les Données personnelles.

Nous sommes en train d’essayer de trouver des réponses à des questions telles que : quels utilisateurs peuvent avoir accès à différents types de documents numériques ? peuvent-ils avoir accès aux collections à partir d’ordinateurs situés hors de la Bibliothèque nationale ?

2.3.3 Outils d’accès

Des besoins tels que ceux que nous avons décrits ci-dessus sont intéressants pour nous dans notre tentative d’élaborer des outils d’accès devant permettre de lancer des recherches dans nos archives Internet. Bien sûr, il nous faut prendre en considération le fait que les bibliothécaires ne catalogueront que très peu des documents qui y seront disponibles.

Sur un plan plus technique, le Projet Paradigma espère donner aux utilisateurs accès aux archives Internet grâce à l’Outil d’Accès du Nordic Web Archive (NWA)⁹ (voir figure 1). Aujourd’hui, la recherche en texte libre avec des opérateurs booléens, la recherche par URL et la présentation de l’histoire d’un document au cours du temps sont des choix standard. Nous espérons que l’outil nous donnera encore plus de possibilités dans l’avenir : combinaisons de recherches booléennes entre plusieurs listes de résultats, recherche en parallèle sur des documents catalogués dans des catalogues bibliographiques extérieurs, recherche sur des métadonnées extraites automatiquement, possibilités avancées de navigation programmée et paramètres de recherche pré-programmée, constitution de « paniers » de listes de réponses, accès à des groupes de documents pondérés selon divers critères (éditeur, etc.), maximum d’une seule

⁹ Pour de plus amples informations sur le Projet Nordic Web Archive, voir l’URL : <http://www.nwa.nb.no/> (réf. du 15 avril 2004).

réponse pour les doublons existants, reconstitution sous forme d'une seule réponse d'un document logique composé de nombreuses pages web distinctes, etc.

Nous comptons adapter l'interface de l'Outil d'Accès NWA à diverses fonctions spécifiques, et la manière dont nous faisons appel au modèle *FRBR* de l'IFLA jouera dans l'avenir un rôle important dans les modalités d'accès à la documentation engrangée.

3 En quête de solutions par les métadonnées

Le Projet Paradigma est à mi-chemin de sa quête de solutions et de formats de métadonnées convenables. Au cours de l'année écoulée, l'une de nos principales activités a consisté à définir des métadonnées de *découverte*, ainsi qu'à chercher des solutions satisfaisantes pour l'extraction automatique de métadonnées techniques. Dans la section qui suit nous tentons de donner un aperçu des raisons pour lesquelles nous comptons décrire les nombreux documents numériques engrangés dans nos archives Internet, et de la manière dont nous comptons nous y prendre.

3.1 Pourquoi cataloguer des ressources Internet ?

Nancy Olsen, dans l'introduction à son ouvrage *Cataloging Internet resources* [3], donne trois raisons fondamentales de cataloguer des ressources Internet :

1. Il y a beaucoup d'information valable sur Internet.
2. Ces ressources doivent être organisées pour être rendues accessibles.
3. L'utilisation des techniques bibliothéconomiques existantes et les procédures de création de notices pour la recherche documentaire sur des catalogues en ligne constituent la méthode la plus efficace pour avoir accès à ces ressources.

Nous sommes d'accord avec Olsen sur ces trois points, mais en même temps nous estimons que *bien moins de 1 %* de la documentation collectée sur le domaine Internet norvégien pourra jamais faire l'objet d'une description bibliographique à un niveau ou un autre. Cela est dû bien sûr à la quantité même des documents archivés. Nous pouvons essayer de nous consoler avec la pensée suivante : bien qu'un pourcentage très supérieur des documents plus traditionnels de la Bibliothèque fasse l'objet d'une description bibliographique, il est vrai que différents documents sont traités de manière différente : les publications éphémères ne donnent lieu qu'à une description sommaire, tandis que les ouvrages et les périodiques sont catalogués plus en détail.

À l'inverse, 100 % des documents Internet seront complètement indexés après collecte, à l'aide du logiciel d'indexation FAST¹⁰. Cela permettra au personnel de la Bibliothèque et aux utilisateurs d'effectuer des recherches dans les documents Internet archivés – tant en texte libre que sur des index. L'infime proportion de documents Internet catalogués manuellement sera disponible en full text dans les archives et par l'intermédiaire de notices bibliographiques dans le catalogue de la Bibliothèque – avec, on l'espère, un lien entre les deux qui soit commode pour les utilisateurs.

Non seulement nous cataloguerons certains documents et indexerons la totalité des documents, mais nous collecterons aussi les métadonnées incorporées ainsi que les documents Internet qu'elles décrivent, et la Bibliothèque Nationale envisage de mettre sur pied dans l'avenir un service qui permettra aux éditeurs de produire et de fournir leurs métadonnées avec leurs documents au moment du dépôt.

3.2 Qu'est-ce que les métadonnées ?

Notre quête de solutions en termes de métadonnées nous a, bien sûr, conduits à rechercher des définitions exactes. Le terme « métadonnées » n'a cessé d'être défini et redéfini dans la littérature. La définition la plus fréquente est sans doute « données sur des données », et les métadonnées recouvrent

¹⁰ Pour de plus amples informations sur FAST Search & Transfer (ASA), voir l'URL : <http://www.fast.no/> (réf. du 15 avril 2004).

toute une série de types d'information¹¹. Nous nous sommes rendu compte que les schémas de métadonnées sont aussi abondants que variés, mais ils ont tous une chose en commun : ils peuvent nous aider à *décrire* et à *trouver* les nombreux documents valables de notre fonds – y compris ceux qui n'appellent pas un catalogage de haut niveau.

3.3 Qu'est-ce qu'un document Internet ?

3.3.1 Définition d'un document Internet d'un point de vue technique

Quand un document Internet est sélectionné pour la collecte et, partant, l'archivage, le sens de l'expression « un document » peut être ambigu : quels sont les composants qui devraient être collectés et archivés en tant qu'éléments à part entière du document ? Quels sont les composants qui devraient faire l'objet d'une appréciation individuelle ? Nous supposons que tout composant qui affecte « l'apparence » (y compris le *son* et autres éléments *non graphiques*) d'une page Web doit absolument être conservé dès lors que la page Web est sélectionnée, c'est-à-dire : l'arrière-plan, le contenu des frames, les icônes des boutons, etc.

Les documents vers lesquels pointent des liens ne sauraient être confondus avec le document de départ, mais ils entretiennent un rapport avec lui. À un plus haut niveau sémantique, nous voudrions souvent traiter tout un ensemble de documents reliés les uns aux autres comme un seul immense document. Si nous les traitons comme des documents complètement indépendants, nous courons le risque, par exemple, de collecter quelques chapitres d'un rapport et de perdre les autres (cela pourrait arriver parce qu'ils contiennent de larges citations, des résumés, etc., en d'autres langues que le norvégien).

Aussi, en réponse à notre question « qu'est-ce qui constitue un document Internet ? », pouvons-nous répondre qu'un document Internet se compose de nombreux fichiers associés, par exemple du texte, de l'image, du son, des animations, etc., et que ces derniers sont la plupart du temps mis en relation par des liens et quelquefois contenus dans des frames.

3.2.2 Définition d'un document Internet d'un point de vue bibliographique

Bien sûr, nous ne pouvons jamais compter sur un ordinateur pour nous dire où un document Internet commence et où il s'arrête – même si nous intégrons à son programme des instructions qui vont dans ce sens. Heureusement, les bibliothécaires sont à même de décider quelles sont les nombreuses parties d'un document Internet qui constituent un tout logique. Ainsi, d'un point de vue bibliographique, nous pouvons définir un document Internet comme une unité informationnelle qui peut être décrite bibliographiquement. Cette définition, volontairement, ne spécifie *pas* un ensemble d'éléments documentaires uniques ou déterminés, mais laisse le bibliothécaire identifier l'objet décrit : une seule notice peut suffire à décrire un site Web, et une ressource donnée présente sur ce site peut également recevoir sa propre description. Le bibliothécaire est libre d'inclure ou d'omettre les sons en arrière-plan, les feuilles de style, etc., et peut collecter dans un seul document plusieurs pages Web étroitement liées, par exemple les chapitres d'un rapport. Nos futures procédures automatisées proposeront au bibliothécaire des définitions du document, sur la base d'une analyse du contenu, des classes de liens, etc. : par défaut, les images incorporées, les clips sonores ou vidéo et les feuilles de style directement reliés sont inclus dans le document. Les liens de certaines classes, identifiant une page Web référencée, par exemple en tant que table des matières ou en tant que section, sont également inclus.

Ainsi, une unité informationnelle qui peut être décrite bibliographiquement constitue le point de départ pour une description par métadonnées – que le document numérique soit déposé sur un support physique tel que cédérom, DVD, ou qu'il soit collecté sous forme de fichiers distincts sur Internet. Ce qui

¹¹ L'une des nombreuses études sur lesquelles nous nous sommes penchés est : *DESIRE : a review of metadata : a survey of current resource description formats*. (1997). Voir l'URL : http://www.koln.ac.uk/metadata/desire/overview/rev_toc.htm (réf. du 15 avril 2004).

veut dire que tous les documents numériques – depuis les types de documents *traditionnels* tels que les monographies, les thèses, etc., jusqu'à des types de documents *éphémères* tels que les journaux sur Internet, la cyberpoésie, le cyberthéâtre, etc. et de *nouveaux* types de documents tels que les pages d'accueil, les weblogs, etc. – sont candidats à une description par métadonnées dans le cadre de notre archivage d'Internet.

3.4 Enquête sur les métadonnées et les travaux apparentés

3.4.1 De quels types de métadonnées avons-nous besoin ?

Il nous a paru intéressant de nous enquérir des formats de métadonnées usités aujourd'hui par la Bibliothèque Nationale pour la description de divers types de documents numériques. Cette information peut être utile, dans la mesure où nous espérons pouvoir un jour importer et exporter des données dans nos archives. Notre enquête a montré que plusieurs formats sont en usage : BIBSYS-MARC (le format MARC du système BIBSYS) pour les textes numériques, l'Ensemble d'Éléments de Métadonnées Dublin Core¹² pour les programmes radio, MAVIS¹³ (un système et un format australiens) pour les documents télédiffusés, les sons et les images, ainsi que d'autres formats usités dans des systèmes développés en local.

Ces formats de métadonnées sont bien adaptés à leur usage, mais n'apportent pas de solution satisfaisante à tous nos besoins en métadonnées. Les archives Internet nécessitent de nombreux types de métadonnées : des métadonnées *administratives* concernant par exemple la création et la modification des enregistrements de métadonnées, des métadonnées de *gestion des droits et accès* afin d'enregistrer les informations relatives au copyright et de définir les groupes d'utilisateurs qui ont le droit d'avoir accès aux archives et les documents qu'ils peuvent lire, des métadonnées *structurelles* afin de montrer les relations logiques entre les objets, entre les métadonnées ou entre les objets et les métadonnées, des métadonnées de *préservation à long terme* pour spécifier, par exemple, les types de fichiers, le logiciel requis et l'historique de la conversion/migration du document, et enfin, des métadonnées *techniques* pour spécifier la taille, les écritures, les détails de communication, etc., des documents. *Last but not least*, nous avons besoin de métadonnées *descriptives* et *analytiques* à des fins de recherche documentaire.

3.4.2 Quel modèle de description devrions-nous choisir ?

Les opinions sont partagées quant au niveau de description à appliquer à un document numérique. Lorsque nous tentions de définir le contour de nos métadonnées descriptives et analytiques, nous avons examiné deux modèles qui étaient en balance. L'un de ces modèles propose d'utiliser trois niveaux de description :

1. Cataloguer en vue d'une intégration à la Bibliographie nationale, au catalogue de la Bibliothèque nationale, à BIBSYS ou à d'autres bases de données spécifiques.
2. Cataloguer à un niveau plus simple dans un format commun.
3. Procéder à une extraction automatique des métadonnées du document lui-même ainsi que des protocoles de communication, etc.

L'autre propose deux niveaux seulement – « cataloguer ou ne pas cataloguer » :

1. Cataloguer en vue d'une intégration à la Bibliographie nationale, au catalogue de la Bibliothèque nationale, à BIBSYS ou à d'autres bases de données spécifiques.
2. Procéder à une extraction automatique des métadonnées du document lui-même ainsi que des protocoles de communication, etc.

¹² Pour de plus amples informations sur la Dublin Core Metadata Initiative, voir l'URL : <http://www.dublincore.org/> (réf. du 15 avril 2004).

¹³ Pour de plus amples informations sur le système MAVIS de Wizard, voir l'URL : <http://www.wizardis.com.au/ie4/products/mavis/introducingmavis.html> (réf. du 15 avril 2004).

Plusieurs arguments plaident en faveur de cette seconde solution : 1) la recherche de documents numériques (en texte libre, etc.) ne dépend pas de l'enregistrement à l'inventaire comme c'est le cas pour les documents analogiques qui ne sont pas portés à l'inventaire. 2) La Bibliothèque ne doit pas forcément inscrire les documents à l'inventaire afin de garder trace de sa logistique, par exemple quelles bibliothèques universitaires ont reçu des exemplaires. 3. Nous pouvons toujours revenir sur notre décision de ne pas cataloguer tel type de document.

La section qui suit donne une brève description de chacun des trois niveaux.

► **Cataloguer en vue d'une intégration à la Bibliographie nationale, etc.**

Pour le moment, nous ne pouvons nous prononcer sur les types de documents qui devraient recevoir le plus haut niveau de catalogage, mais nous pouvons affirmer avec certitude qu'un petit nombre de documents numériques valables continueront d'être catalogués en format MARC afin d'être intégrés à la Bibliographie nationale. (Indiquons au passage que le format MARC utilisé en Norvège s'appelle NORMARC, que certains systèmes ont adopté des versions locales, par exemple BIBSYS MARC, et qu'à un niveau national on parle d'adopter MARC21¹⁴. Le code de catalogage norvégien s'appuie sur la seconde édition des Règles de catalogage anglo-américaines (AACR2), dont les chapitres 9 et 12 sont désormais disponibles en norvégien.)

Nous pouvons également affirmer avec certitude que le catalogage des documents audiovisuels pour la conservation à long terme nécessite un haut niveau de détail – notamment lorsqu'il s'agit de garder la trace d'informations en relation avec la restauration d'originaux, de copies, etc. La Bibliothèque continuera sans aucun doute d'utiliser MAVIS pour cela.

► **Cataloguer à un niveau plus simple dans un format commun.**

Comme il a été dit plus haut, la Bibliothèque nationale envisage dans l'avenir de proposer un service qui permettra aux éditeurs de produire et de fournir des métadonnées avec leur document lors du dépôt. Aujourd'hui, le Projet Paradigma s'efforce de définir le ou les formats de métadonnées qui constitueront la base d'un futur outil simple d'utilisation qui fournira ce service. Au bout du compte, les bibliothécaires pourront traiter les enregistrements de métadonnées fournis par les éditeurs, les prenant pour base de notices bibliographiques de plus haut niveau.

Nous avons analysé et comparé quelques formats de métadonnées au cours de notre recherche de solutions adaptées : MARC et Dublin Core, puisque l'un et l'autre sont en usage dans les bibliothèques et organismes assimilés ; MODS¹⁵ et METS¹⁶, puisque ces formats ont été élaborés par des bibliothèques pour des bibliothèques, et ONIX, format élaboré par les industries du livre et de l'édition. Nous prenons note également de ce que la communauté ISBN a proposé qu'à l'avenir des déclarants puissent fournir aux agences d'enregistrement ISBN des métadonnées compatibles avec ONIX lors de l'attribution de chaque ISBN.

Nous avons comparé les formats énumérés ci-dessus en nous demandant : qui est responsable de la gestion de ce format ? S'agit-il d'une norme internationale ? Quel type de média décrit-il ? Comporte-t-il des définitions sémantiques et/ou syntaxiques ? Comment décrit-il les relations qui existent entre les documents ? Le format dépend-il de règles ou de codes spécifiques ? Est-il compatible ou en relation avec d'autres formats ? Est-il largement usité, et par quelles communautés ?

¹⁴ Pour de plus amples informations sur MARC21, voir l'URL : <http://www.loc.gov/marc/bibliographic/ecbdhome.html> (réf. du 15 avril 2004).

¹⁵ Pour de plus amples informations sur MODS, voir l'URL : <http://www.loc.gov/standards/mods/> (réf. du 15 avril 2004).

¹⁶ Pour de plus amples informations sur METS, voir l'URL : <http://www.loc.gov/standards/mets/> (réf. du 15 avril 2004).

Nous espérons que cette enquête pourra déboucher sur un plus vaste débat sur les métadonnées à la Bibliothèque, en relation avec sa révision en cours. Nous comptons également regarder de plus près la manière dont nous pourrions satisfaire les exigences de nos utilisateurs en termes de fonctionnalités en faisant appel aux *Notices minimales* proposées par le Groupe de travail de l'IFLA sur l'utilisation des schémas de métadonnées [6] et au *Rapport final sur les Fonctionnalités requises des notices bibliographiques* (FRBR) [5]. Nous avons en outre un projet de collaboration sur des solutions par les métadonnées avec un projet bibliographique en cours au sein de la Bibliothèque nationale elle-même, ainsi qu'avec la *Bibliothèque numérique norvégienne*, un autre projet de portée nationale. Enfin, nous espérons que ces travaux déboucheront sur la recommandation de formats de métadonnées pour la description à différents niveaux.

Entre-temps, nous avons cherché à spécifier nos besoins en métadonnées techniques pour notre logiciel d'archivage. Nous avons identifié plusieurs facteurs qui peuvent influencer notre choix d'un format de métadonnées pour un catalogage de plus bas niveau. Voici quelques facteurs techniquement souhaitables :

- Interopérabilité sémantique avec MARC : il importe que les attributs du format de métadonnées soient sémantiquement harmonisés avec le format MARC, prédominant dans la communauté bibliothéconomique. Si possible, le format devrait être un sous-ensemble fonctionnel de MARC. Cela faciliterait l'échange de données.
- Simple mais riche : il importe de trouver un format de métadonnées qui soit simple à utiliser, mais suffisamment riche pour nous permettre de représenter un niveau de détail adéquat.
- Facile à convertir vers d'autres formats : les passerelles de conversion entre le format choisi et MARC devraient être disponibles ou relativement faciles à définir. Ici, nous voyons que des passerelles existent déjà entre MARC21 et MODS – ainsi qu'entre MARC21 et ONIX –, ainsi qu'une passerelle entre le Dublin Core non qualifié et MODS.
- Compatibilité avec XML : XML est plus ou moins un standard, et un format compatible avec XML nous permettra de manipuler le format avec des logiciels disponibles. Une structure cadre plus large sera aussi définie en XML, permettant ainsi aux archives de recevoir des métadonnées de diverses sources, de gérer la modification des métadonnées, de définir les métadonnées d'origine, de garder trace de l'historique des versions, etc. (par exemple, METS).
- Extensibilité : un format de métadonnées devrait nous permettre de définir de nouveaux éléments chaque fois que c'est nécessaire.
- Éléments fondamentaux : il importe de définir des éléments de métadonnées fondamentaux, c'est-à-dire un dénominateur commun qui pourra faciliter la recherche documentaire entre différents types de documents.

Si nous comparons ces facteurs avec les formats de métadonnées décrits dans notre enquête, nous voyons que les formats qui sont compatibles avec MARC et XML sont préférables. Il n'y a toutefois pas de recette toute faite. Il faut définir de nouveaux éléments pour les métadonnées techniques, structurelles et de gestion des droits et accès, et peut-être les consolider dans le cadre de METS. Cela reste bien sûr vrai aussi en ce qui concerne les métadonnées pour la préservation à long terme. Mais là, le Silo Numérique à Long Terme de la Bibliothèque exige que nous utilisions des métadonnées compatibles avec OAIS¹⁷.

► Extraction automatique de métadonnées

Malheureusement, les bibliothécaires ne catalogueront jamais la masse inconcevable de 99 % des documents Internet que nous aurons archivés. C'est pourquoi nous faisons porter nos recherches sur l'utilisation de systèmes d'analyse et d'extraction automatiques des métadonnées de documents Internet dans le cadre de nos travaux sur les métadonnées et la conception de systèmes. Les métadonnées extraites

¹⁷ Pour de plus amples informations sur le modèle de référence OAIS, voir l'URL : <http://ssdoo.gsfc.nasa.gov/NOST/WWWCLASSIC/DOCUMENTS/PDF/ccsds-650.0-b-1.PDF> (réf. du 15 avril 2004).

seront stockées avec les objets numériques et les autres métadonnées descriptives, et seront rendues disponibles pour la recherche structurée dans les archives Internet.

La technique n'est pas encore suffisamment au point et ne permet pas de déterminer automatiquement le type d'un document, mais elle peut aider à réduire le nombre de documents qui doivent être examinés par un agent humain au cours de la phase 2 du processus de sélection. Comme exemples de propriétés qui peuvent être déterminées automatiquement pour un document, on peut citer : 1) langue, vocabulaire et grammaire, 2) taille et structure du document, 3) source/éditeur/serveur Web, 4) usage des « cookies », 5) âge et durée de vie escomptée du document, 6) présence de son, d'images, d'animations, de vidéo et autres types d'informations avancées, 7) interaction avec l'utilisateur sous la forme de « formulaires », boutons, etc., 8) nombre, type et source de liens, 9) valeurs des URL, par exemple utilisation de mots ou caractères spéciaux dans l'URL, 10) utilisation de scripts côté client, 11) détails sur la communication technique.

Les techniques d'analyse lexicale et grammaticale font des progrès, et nous pensons que ce type d'analyse peut constituer un élément important des futures procédures automatisées. En fin de compte, les propriétés recueillies automatiquement seront rendues disponibles à la recherche structurée dans les archives Internet. Ces propriétés prises isolément seront d'un intérêt limité, mais en combinaison avec d'autres critères, elles pourront bien s'avérer très utiles.

4. Le rôle des FRBR dans les archives Internet

Le Projet Paradigma compte présenter les documents numériques archivés et leurs métadonnées sous forme organisée et structurée, afin d'y faciliter la navigation. Nous avons trouvé que le modèle FRBR élaboré par l'IFLA constituait un outil essentiel à cet égard, et nous utiliserons ce modèle comme pierre de touche de la conception de notre dépôt d'archives Internet.

Nous croyons que le fait d'ajouter au modèle FRBR des mécanismes de modélisation des agrégats sera bénéfique à notre travail sur des médias dynamiques tels que des documents Internet, des multimédias et autres ressources continues. Les mécanismes d'agrégation peuvent être implémentés comme de simples extensions au modèle, sans imposer de changements lourds aux concepts FRBR déjà définis. Un article consacré aux mécanismes d'agrégation que nous proposons sera publié dans le courant de l'année dans le numéro spécial de *Cataloging & Classification Quarterly* sur les FRBR.

Afin d'adapter le modèle FRBR aux documents Internet dynamiques, il est nécessaire de procéder à une légère réinterprétation des concepts de *manifestation* et d'*item*, point développé dans la section qui suit.

4.1 Adaptation des FRBR au contexte des documents Internet dynamiques

4.1.1 Les documents dynamiques

Les documents Internet sont souvent *dynamiques*, par exemple un journal sur Internet, faisant l'objet de multiples mises à jour au cours d'une même journée. Un utilisateur peut se référer à ce type de document dynamique comme un forum ou un canal d'information : « Le *Daily News* rapporte que... » Nous pouvons peut-être dire qu'un document dynamique correspond *grosso modo* à une URL. Les concepts de « livraisons » et d'« éditions » successives doivent également être repensés dans le contexte d'Internet : d'un point de vue formel, il se peut qu'une mise à jour d'une page Web soit analogue à une nouvelle édition d'un livre. Toutefois, les lecteurs voient par exemple la page d'accueil d'un journal sur Internet comme une seule entité qui ne cesse de changer – et non comme un ensemble d'éditions distinctes.

En utilisant le modèle FRBR avec des extensions pour les composants d'agrégats, nous avons défini le concept de *document dynamique* comme étant « l'intégralité du cycle de vie d'une page Web modifiée en permanence ou autre document Internet similaire ».

S'il nous fallait cataloguer un document Web à mise à jour de ce type en conformité avec les AACR2, nous appliquerions normalement les règles valables pour les ressources intégratrices, c'est-à-dire

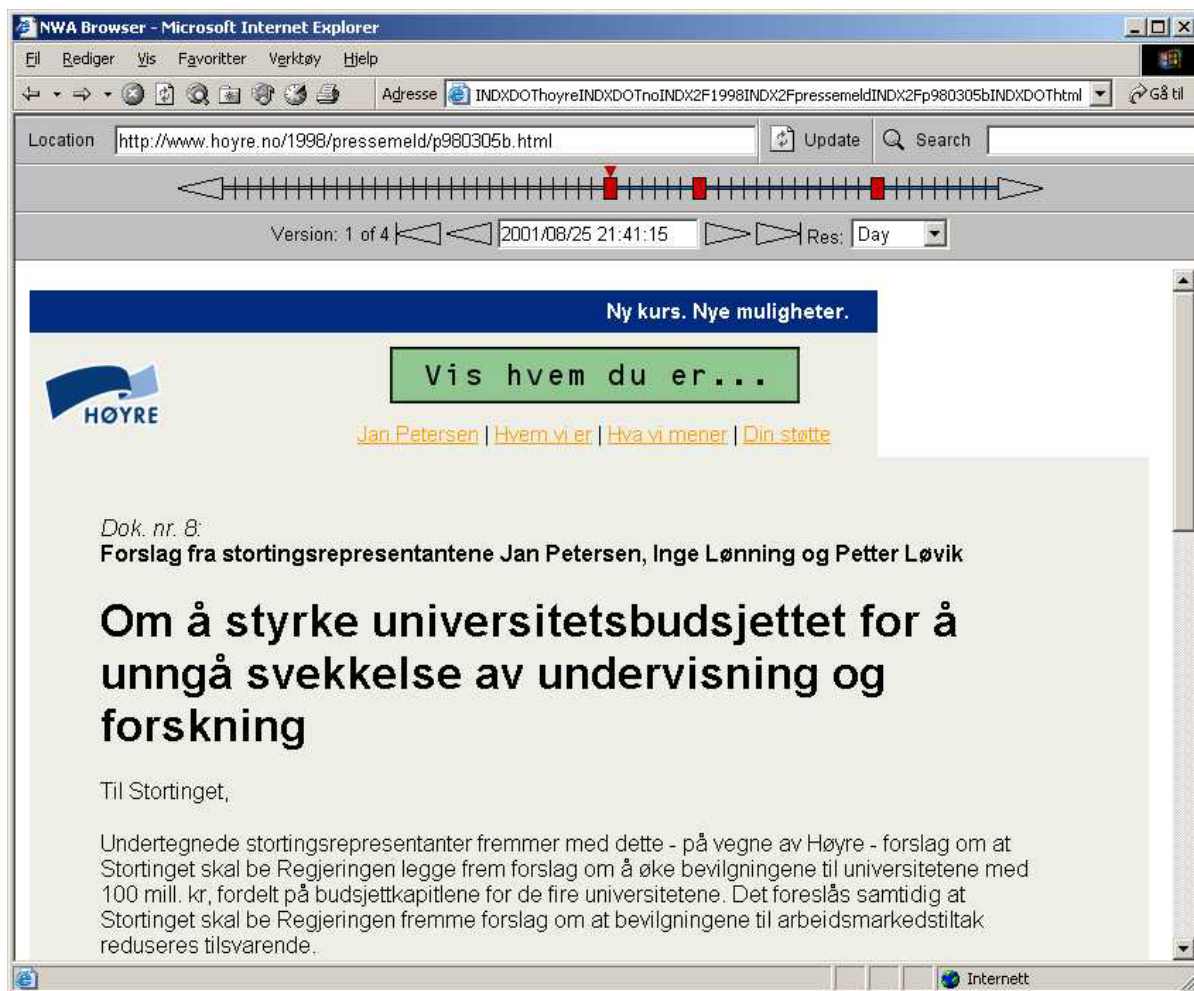
les ressources bibliographiques auxquelles on ajoute ou modifie quelque chose au moyen de mises à jour qui ne conservent pas leur autonomie mais se fondent dans l'ensemble. Mais des documents tels que les journaux sur Internet ressemblent plutôt à un canal radio, un flux d'informations éphémères en perpétuel changement. Ils ne « se fondent pas dans l'ensemble ». La capture à un moment donné du contenu d'un document en perpétuelle transformation s'apparente à l'enregistrement d'un *échantillon* d'une radiodiffusion éphémère. Nous nommons *document spécifique* chacun de ces échantillons ou « instantanés ».

Lorsque l'on accède à un document dynamique sur le Web, l'*item* (c'est-à-dire l'exemplification) retrouvé par un utilisateur peut être différent de tous les autres *items* du même document : cela peut dépendre de la combinaison d'un certain nombre de facteurs : l'identité de l'utilisateur, l'outil d'accès employé (browser Web), les informations relatives aux accès précédents au même document (informations conservées dans les cookies), les paramètres expressément spécifiés par l'utilisateur (dans un formulaire, par exemple), et enfin, mais non des moindres, l'état de la base de données au moment où l'accès a lieu. Il arrive souvent que l'*item* soit généré à la volée quand un utilisateur demande une exemplification. En d'autres termes, un appel HTTP se comporte comme un service d'« impression à la demande » : l'exemplaire fourni reflète le contenu de la base de données documentaires, quel qu'il puisse être à l'instant même de l'impression. La base de données peut être considérée comme une représentation matérielle (semi-)permanente du document dynamique, à partir de laquelle des *items* spécifiques sont dérivés. Les *items* eux-mêmes n'ont pas de représentation permanente – ils sont éphémères tant qu'ils ne sont pas recueillis et conservés, par exemple dans un archivage Internet.

4.1.2 Les documents spécifiques

Nous avons défini un *item* constituant un exemplaire d'un document dynamique comme étant un *document spécifique*, qui diffère d'un *item* traditionnel sur un point fondamental : il s'agit d'un membre d'un groupe d'*items* constituant des exemplaires du même document dynamique. Un document stocké dans les archives, ou affiché pour le compte de l'utilisateur, est de toute évidence un document spécifique, mais avec une nuance importante : une recherche plein texte donnera au maximum *une seule* entrée pour le document spécifique dans la liste de réponses. Si l'utilisateur demande l'affichage d'une réponse, le document dynamique est présenté comme une seule unité, et l'utilisateur peut choisir un *item* spécifique sur un *axe chronologique*, c'est-à-dire une barre de menu représentant la durée de vie du document. Chaque version conservée, c'est-à-dire chaque document spécifique, est indiquée sur cet axe chronologique avec un marqueur. L'utilisateur peut accéder à n'importe quel document spécifique en cliquant sur le marqueur correspondant à une date/un moment donné, ce qui lui permet de retrouver l'*item* (cf. figure 1).

Figure 1. Présentation d'un document dynamique dans l'interface utilisateur Outil d'Accès NWA



4.2 Document défini par l'éditeur ou par l'utilisateur et définitions de métadonnées

La présentation des documents archivés aux fins de recherche et documentation ne constitue que l'un des services qui seront offerts par la Bibliothèque nationale. À côté de cela, et en nous appuyant sur les idées exposées ci-dessus, nous avons proposé de réformer le service, qu'offre déjà la Bibliothèque, d'attribution d'identifiants. À l'heure actuelle, ce service Web attribue des URN:NBN [7] aux universités et autres établissements de la branche norvégienne de l'espace de nommage URN:NBN. Nous entrevoyons cependant des possibilités d'attribution d'ISBN isolés à l'aide de ce même service.

4.2.1 Scénario prospectif pour de futures fonctionnalités

Voici un scénario présentant de futures fonctionnalités : l'ensemble d'identifiants de base attribués par ce service nécessite que l'utilisateur/demandeur fournisse un ensemble minimal de *métadonnées* et une définition exacte du document identifié.

Œuvre, expression, manifestation (y compris ce que nous définissons comme document dynamique) et *item* (y compris ce que nous définissons comme document spécifique) peuvent recevoir des identifiants. Les *items* (documents spécifiques) doivent être spécifiés par une liste complète de leurs composants (par exemple un fichier HTML, des fichiers images, des fichiers sons, etc.) ; les *manifestations* (documents dynamiques) peuvent aussi être spécifiés par des règles, telles que « La page d'accueil du journal sur

Internet à la présente URL et toutes les pages directement liées à la page d'accueil qui figure sur le même site Web ».

Pour les identifiants d'*expressions* et d'*œuvres*, l'utilisateur peut facultativement identifier des *expressions*/documents dynamiques et des *items*/documents spécifiques, qui sont des instantiations de cette *œuvre/expression*.

Les définitions établies par l'éditeur ou l'utilisateur sont considérées comme terminales plutôt qu'automatiquement proposées. L'identité de l'éditeur ou de l'utilisateur qui attribue l'identifiant est archivée ; une définition de document spécifiée par une maison d'édition reconnue ou une université peut être considérée comme ayant plus de poids que si elle émanait d'un utilisateur lambda.

4.2.2 Champs de métadonnées

Des champs de métadonnées obligatoires et non obligatoires pourraient être rendus disponibles pour la description du document à chacun des niveaux FRBR, et chaque niveau serait identifié par un URN:NBN. Les valeurs des métadonnées seront stockées avec l'identifiant, et les utilisateurs de notre service de résolution fondé sur Internet sera capable de trouver le document, grâce à ce numéro.

Après avoir porté les informations dans les champs de métadonnées d'un futur outil d'attribution de métadonnées et d'identifiant, un éditeur aurait la possibilité de cliquer sur un bouton, par exemple <HTML Dublin Core>, pour visualiser ces métadonnées en HTML dans une nouvelle fenêtre. L'utilisateur pourrait ensuite copier les métadonnées et les coller dans l'élément <HEAD> du document Web en cours de description, avant de poursuivre le processus d'attribution d'identifiant. Après avoir enregistré le document numérique, dès lors pourvu de métadonnées incorporées, l'utilisateur pourrait aisément stocker la copie du document enrichi de métadonnées dans les archives de la Bibliothèque en cliquant sur le bouton d'actualisation du browser.

4.3 Peut-on envisager des services de vérification et d'authentification de documents ?

Nous avons entendu raconter des histoires d'organismes officiels qui modifiaient leurs textes sur Internet et n'avaient ensuite qu'il en avait existé des versions antérieures. Nous avons également entendu parler de sociétés commerciales qui affichent tel prix pour leurs produits, et qui facturent ensuite ces derniers à un prix bien supérieur.

C'est en pensant à de tels récits, et à bien d'autres, que le Projet Paradigma s'est proposé de mettre sur pied un service de vérification et d'authentification qui permette aux utilisateurs de demander le téléchargement d'un document Internet donné, c'est-à-dire un instantané d'une page Web contenant une offre commerciale particulière, une notification de responsabilité légale, un écrit diffamatoire, etc.

Si des doutes venaient à s'élever ensuite au sujet de ce qu'était le contenu de ces documents à tel ou tel moment, la Bibliothèque pourrait confirmer (ou infirmer) toutes prétentions sur ce point. Même lorsqu'aucun aspect juridique n'est impliqué, un *item* de document spécifique conservé peut toujours servir de miroir précis d'un document dynamique à un instant T, par exemple à des fins de citation ou de renvoi bibliographique. Ceci est d'autant plus important que la plupart des documents Internet n'ont pas de numéros de page, de numéro de version, etc.

Dans notre dépôt d'archives Internet, un document spécifique est défini sous la forme sous laquelle il a été reçu du serveur Web. Il y a un train numérique bien défini pour chaque composant du document (texte, images, etc.). Le rendu graphique du document ne fait *pas* partie de sa définition – ce processus est réservé à l'outil d'accès. Le document spécifique est identifié comme étant le contenu du document dynamique donné par certains composants et certaines métadonnées :

- la source de chaque composant (par exemple, une URL)
- tous les paramètres spécifiés par le client lorsqu'il retrouve les composants
- l'heure à laquelle chacun des composants a été retrouvé
- l'ensemble des composants inclus dans le document.

5. Conclusion

Le Projet Paradigma de la Bibliothèque nationale de Norvège ne ménage pas ses peines pour parvenir à l'élaboration d'une technique, d'une méthode et de procédures de routine satisfaisantes pour le dépôt légal de tous types de documents numériques – y compris les millions de documents trouvés sur le domaine Internet norvégien – dans le temps qui nous est imparti. Nous espérons être en mesure de donner à nos utilisateurs accès à la documentation archivée par le biais de notices bibliographiques, de différents types de métadonnées et d'outils de recherche en plein texte dès 2005.

Notre dépôt d'archives Internet structuré selon les FRBR sera sans doute l'un des premiers du genre, et nous espérons également concrétiser nos idées concernant l'utilisation des niveaux d'entité FRBR *œuvre*, *expression*, *manifestation* et *item* dans le cadre d'un futur service d'attribution d'identifiants. Peut-être nos idées concernant un service de vérification/authentification sur Internet prendront-elles corps un jour ou l'autre ? L'avenir le dira, mais en attendant, la Bibliothèque nationale va continuer d'explorer de nouveaux sentiers afin de préserver le patrimoine culturel numérique de la Norvège et de fournir à ses utilisateurs des outils qui puissent ouvrir les portes de cette stimulante bibliothèque numérique.

Orientation bibliographique

(Toutes les références sont du 15 avril 2004).

[1] Rapport final sur le Projet pilote « Netarkivet.dk » [en ligne]. – URL: <http://www.netarkivet.dk/rap/webark-final-rapport-2003.pdf>.

[2] Guidelines for the selection of online Australian publications intended for preservation by the National Library of Australia [en ligne]. – URL: <http://pandora.nla.gov.au/selectionguidelines.html>.

[3] The Kulturarw3 Project – The Royal Swedish Web Archiw3e – An example of “complete” collection of web pages [en ligne]. – URL: <http://www.ifla.org/IV/ifla66/papers/154-157e.htm>.

[4] Olsen, Nancy (2002). Cataloging Internet Resources : A Manual and Practical Guide [online]. – OCLC. – URL: <http://www.oclc.org/support/documentation/worldcat/cataloging/internetguide/1/1.htm>.

[5] Norvège. [Loi sur le Dépôt légal (1989)] Act relating to the legal deposit of generally available documents : no. 32 of 9 June 1989 : with regulations / [published by the Ministry of Church and Cultural Affairs ; unofficial English translation published by the National Library of Norway. – [Oslo] : National Library of Norway, 1997. – 21 p.

[6] IFLA Cataloguing Section Working Group on the Use of Metadata Schemas (2003). Guidance on the structure, content, and application of metadata records for digital resources and collections : draft for worldwide review 27 October, 2003 [online]. – URL: <http://www.ifla.org/VII/s13/guide/metaguide03.pdf>.

[7] RFC 3188 Using National Bibliography Numbers as Uniform Resource Names [online] / J. Hakala, 2001. – URL: <http://www.ietf.org/rfc/rfc3188.txt>.

[8] Van Nuys, Carol (2003). Identification of network accessible documents : problem areas and suggested solutions [online] / Carol van Nuys, Ketil Albertsen. – P. 13-25. – Dans : Proceedings : in conjunction with the 7th European Conference on Research and Advanced Technologies for Digital Libraries, ECDL 2003 / Julien Masanès, Andreas Rauber, Gregory Cobena (eds). – URL: <http://bibnum.bnf.fr/ecdl/2003/index.html>.

[9] Albertsen, Ketil (2003). The Paradigma web harvesting environment. – P. 49-62. – Dans : Proceedings : in conjunction with the 7th European Conference on Research and Advanced Technologies for Digital Libraries, ECDL 2003 / Julien Masanès, Andreas Rauber, Gregory Cobena (eds). – URL: <http://bibnum.bnf.fr/ecdl/2003/index.html>.

[10] Van Nuys, Carol (2003). The Paradigma project [en ligne]. – Dans : RLG DigiNews. – Vol. 7, n° 2. – URL: http://www.rlg.org/preserv/diginews/v7_n2_feature2.html.