

Contribution à la problématique de préservation des documents électroniques dans le cadre du dépôt légal

Bouder Hadjira
hbouder@mail.cerist.dz
Nekri Mounira
mnekri@mail.cerist.dz

Centre de Recherche sur l'Information Scientifique et Technique

Introduction

L'avancement rapide des technologies de l'information a donné naissance à un nouveau mode de publication, à savoir la publication électronique. Cette dernière présente certaines particularités qui ont bouleversé les pratiques établies en matière de collecte, traitement, diffusion et de préservation de ces documents. A noter que la préservation des documents électroniques constitue une problématique pour l'ensemble des institutions productrices d'information numérique, notamment les bibliothèques nationales dans le cadre de l'obligation de dépôt légal. A considérer que ces dernières ont été les premières à faire face à cette problématique. C'est pourquoi nous avons choisi de focaliser notre intervention autour de la préservation des documents électroniques reçus dans le cadre du dépôt légal dont les résultats pourraient être transposés dans les autres secteurs concernés.

En fait, certains pays ont dès la dernière décennie engagé des projets de recherche visant à mettre en place un dispositif juridique et technique susceptible d'assurer la préservation du patrimoine numérique national en constante évolution. Ainsi des projets de révision des législations sur le dépôt légal ont été soumis au débat, dont certains ont été adoptés et sont actuellement en vigueur. Ces derniers sont accompagnés de projets pilotes généralement initiés par les bibliothèques nationales en vue de mettre au point un mécanisme technique fiable et viable garantissant l'accessibilité à ces documents au profit des générations futures.

Au niveau international, l'importance de la masse d'information véhiculée par ces nouveaux supports a également fait de la préservation du patrimoine numérique une préoccupation mondiale, c'est d'ailleurs ce que démontre le rapport concernant un projet de charte sur la préservation du patrimoine numérique, établi par le directeur général de l'UNESCO conformément à la résolution 31 C/34 adopté par la conférence générale à sa 31^{ème} session, par laquelle elle a appelé l'attention sur le développement continu du patrimoine numérique mondial et sur la nécessité d'une campagne internationale pour la sauvegarde de la mémoire numérique menacée. Ce projet de charte devait être soumis pour adoption à la conférence générale à sa 32^{ème} session en 2003.

A cet effet, notre intervention, qui s'inscrit dans le cadre de l'axe « Le livre », s'articulera en premier lieu autour des enjeux liés à la préservation des documents

électroniques : en effet, un document électronique dépend de son environnement (matériel informatique, logiciel, système d'exploitation, format de fichier ainsi que le système d'encodage). Par conséquent, il devient difficile de garantir la possession des moyens adéquats (interfaces logicielles et matériel informatique) pour la lecture au moment voulu d'un document électronique établi des années plus tôt. Ceci supposerait la conservation de ces moyens ou le transfert des données et ce à chaque changement de système informatique.

Par ailleurs, une étude des différentes expériences pionnières en matière de préservation des documents électroniques sera présentée en vue de mettre en exergue les principales techniques de préservation établies en la matière. Ces techniques serviront de base pour les différents scénarii de solutions proposées pour la préservation et l'accès à long terme aux documents électroniques.

I- Enjeux techniques et juridiques liés à la préservation des documents électroniques

De nombreuses difficultés d'ordre technique, économique et juridique doivent être résolues pour assurer la pérennisation de l'information numérique.

En pratique un document est entièrement dépendant de son environnement : matériel informatique, logiciel, système d'exploitation, formats d'encodage et de fichier. Etant soumis à l'obsolescence technologique, il est difficile de garantir que l'on disposera à long terme des interfaces logicielles et matérielles requises pour accéder à la lecture du document électronique établi des années plus tôt.

Les solutions envisageables pour remédier à ces difficultés techniques engendrent à leur tour des problèmes aussi techniques mais surtout d'ordre économique et juridique.

En effet, conserver les matériels et logiciels nécessaires à la restitution des documents, ou procéder à des recopies régulières des données vers le nouveau système informatique, nécessite un large investissement contre une faible rentabilité (coût de la formation du personnel, de sa rémunération, maintenance du matériel et son lieu de stockage...etc.), ce qui constitue une opération économiquement très coûteuse.

Par ailleurs, l'adoption de nouvelles lois sur la propriété intellectuelle concernant les documents électroniques peut constituer à son tour un obstacle pour la préservation de ces derniers, notamment lorsqu'il s'agit de recopier leur contenu ou encore d'accéder à ces données à long terme.

II- Etude des expériences pionnières en matière de préservation des Documents électroniques:

L'apparition de nouvelles techniques de production et de diffusion des publications et l'accroissement du nombre de publication sur supports autres que le support traditionnel, à savoir le support papier, a amené certains pays développés à réviser leurs législations relatives au dépôt légal en vue de les élargir à ces nouveaux types de publications et assurer ainsi la préservation de la masse d'information qu'elles véhiculent au profit des générations futur.

Ces projets de révision des législations sont souvent accompagnés de projet visant la mise en place de dispositif technique de préservation des documents électroniques reçus dans ce cadre.

Afin de s'enquérir de l'état du dépôt légal des documents électroniques dans ces pays, nous tenterons de passer en revue certaines expériences pionnières en la matière, et ce tant au niveau national qu'au niveau international.

II-1- Les Expériences Nationales:

II-1-1- Le Dépôt Légal En Australie:

Dispositif juridique :

Le dépôt légal en Australie est régi par la loi de 1968 sur le droit d'auteur qui ne vise que les documents imprimés. En 1995, un projet visant à étendre cette loi à tous les formats électroniques, existants ou futur, a été soumis à la commission de réexamen de la loi sur le droit d'auteur.

Le comité de révision a publié son rapport en février 1999 et a recommandé que l'on élargisse la définition de « matériel de bibliothèque » pour inclure l'électronique. Désormais, la nouvelle loi sur le droit d'auteur le copyright amendment Act 2000 (Digital Agenda), ne comporte aucune modification des dispositions sur le dépôt légal.

Face à cet état de fait, la bibliothèque nationale d'Australie a instauré une pratique consistant à recueillir un certain nombre de publications électroniques sur CD-ROM et sur disquette, dans le cadre d'un dépôt volontaire.

Dispositif technique :

En Australie, le processus de préservation des publications diffusées en réseau a été défini à travers le projet Pandora (juin 1996). Fondé sur la norme ISO de l'OAIS, ce système collecte les monographies, les publications en séries, les journaux récupérés par le biais d'une cueillette, et les sites Internet d'organisations ou personnels (choix de la bibliothèque). Il utilise les normes de la bibliothèque du Congrès et de celles du Comité bibliographique australien (ABN) pour indexer les champs. Quant aux URLs, la bibliothèque a ajouté la zone 856 du format MARC. Une fois les documents collectés et traités, ils sont convertis en deux formats PDF et SGML et pour assurer leur accessibilité à long terme des métadonnées sont ajoutées.

II-1-2- LE DEPOT LEGAL EN FRANCE :

Dispositif juridique

Pays d'origine du dépôt légal, la France s'est dès 1992 dotée d'une loi actualisée sur le dépôt légal suivi d'un décret en 1993 en complément de cette dernière. Il ressort de ce décret que le législateur français établit une distinction entre trois catégories de documents électroniques dont le dépôt est assujéti à des conditions bien déterminées, il s'agit des : bases de données, progiciels et systèmes experts, œuvres multimédias.

A noter, que Cette loi ne mentionne aucun type de support matériel, ce qui la rend plus souple et évolutive.

Par ailleurs, un projet de loi sur le droit d'auteur dans la société de l'information a été adopté en conseil des ministres le 12 novembre 2003. Ce projet contient des dispositions visant à soumettre au dépôt légal les publications en ligne.

Dispositif technique :

Pour la pérennisation de documents électroniques, la BNF s'est engagée sur deux projets. Le premier est le programme PIN (pérennisation de l'information numérique). Le second est le projet NEDLIB (Networked European Deposit Library).

Pour les documents en ligne, la BNF s'inspire des projets suédois et canadiens pour contrôler son propre programme. Les documents sont collectés par des robots et sans sélection. Elle utilise le système de l'archivage hiérarchique (HSM), par le biais de CD-ROM à haute résistance (Century-Disc).

II-1-3- LE DEPOT LEGAL AU CANADA :

Dispositif juridique

Au Canada, la loi sur le dépôt légal fait partie de la loi sur la bibliothèque nationale de 1995. Cette loi ne vise que les publications électroniques publiées dans un format physique. Les publications électroniques en réseau ne sont pas considérées comme étant couvertes par la loi actuelle.

Dispositif technique

Le processus technique de conservation des documents électroniques au Canada a été défini à travers le projet pilote sur les publications électroniques **PPPE**. L'objectif principal de ce projet consiste en premier lieu à donner la définition exacte d'un document électronique, ensuite l'élaboration des critères de sélection et de conservation pour la collection électronique et à la fin intégrer ces publications dans les politiques de la BNC en ce qui concerne la sélection, la gestion des collections et la conservation.

D'autre part, en partenariat avec le ministère des Relations avec les citoyens et de l'Immigration (MRCI), la BNQ inaugure le dépôt légal des publications gouvernementales québécoises diffusées sur Internet.

II-2- Projet collaboratif de préservation de documents électroniques :

Lancé en janvier 1998, NEDLIB Networked European Deposit Library est le projet le plus représentatif des projets de coopération en matière de préservation des documents électroniques. C'est un projet subventionné par le Programme Télématique de la commission européenne. Il est conçu pour être utilisé par n'importe quelle institution responsable de dépôt des documents électroniques dans le but d'acquérir, de préserver et de permettre l'accès aux documents électroniques.

Il regroupe les documents électroniques sur supports électroniques et en ligne.

L'installation et la préservation sont basées sur le modèle de référence **OAIS**, qui peut être appliqué à n'importe quelle archive et plus spécialement aux organisations ayant la responsabilité de rendre l'information disponible à long terme.

Aucune plate forme n'est définie pour l'installation de ce projet, chaque bibliothèque est libre d'utiliser la plate forme qui répond à ses besoins et bien sûr aux exigences du système.

Il est piloté par la bibliothèque nationale des pays bas, en partenariat avec 8 bibliothèques nationales européennes, les archives nationales des pays bas, deux sociétés informatiques et trois éditeurs.

Son objectif était de développer une architecture commune et d'outils de base pour la construction d'un **Système de dépôt des documents électroniques (Deposit Systems for Electronic Publications ou DSEP)**: Qui est un modèle des spécifications fonctionnelles de dépôt des publications électroniques. Il est basé sur le modèle de référence OAIS (Open Archives Information System). Il est constitué de modules indépendants, mais en interaction permanente. Il reçoit les données sur les documents, génère les métadonnées qui seront utilisées par la suite dans le stockage des documents, leur gestion (établissement du catalogue) et l'accès. Il utilise tous les modules du modèle OAIS à savoir :

- Ingest
- Stockage
- Gestion des métadonnées

- L'accès
- L'administration

Un module spécifique intitulé préservation a été ajouté au DSEP ; ce qui le différencie du modèle OAIS. Ce module gère la conservation à long terme des documents publiés sur support ou sur le réseau, en utilisant deux techniques la migration et l'émulation (voir supra). DSEP permet aussi l'interaction avec le système de gestion de bibliothèque s'il existe, afin d'acquérir les notices descriptives des documents et leur diffusion.

III. Les techniques établies de préservation des documents électroniques

En absence d'un système complet pour la préservation des documents électroniques. Plusieurs techniques ont été développées au cours de ces dernières années en vue de palier, ne serait ce qu'à une partie des difficultés posées par les différents types de supports électroniques. Actuellement il existe plusieurs techniques d'archivage dont quatre principales, à savoir: la migration, l'émulation, l'encapsulation et le rafraîchissement des supports.

III.1 La migration

La migration est une technique de préservation des documents électroniques. Elle peut être appliquée au niveau : du support physique, du système d'exploitation et du format de données. Cette technique consiste à retraiter des données du document électronique afin de le rendre compatible avec le nouvel environnement. En effet les documents électroniques doivent être migrés dès que leur environnement devient obsolète. Quelques risques sont à soulever lors de l'application de cette technique. Il s'agit de la perte d'information d'une ou plusieurs parties du document et les problèmes d'ordre juridiques liés à la reproduction du document électronique.

III. 2 L'émulation

Contrairement à la migration, l'émulation ne retraite pas les données mais elle consiste à émuler l'environnement matériel et logiciel nécessaires pour accéder et manipuler le document électronique. La difficulté de cette technique réside au niveau du développement des émulateurs.

III. 3 L'encapsulation

L'encapsulation consiste à rattacher à chaque document électronique les informations nécessaires pour le codage de ses bits par n'importe quel système et par conséquent assurer son exploitation à long terme. Ces informations contiennent les caractéristiques du système d'exploitation, le matériel et le logiciel pour la manipulation du document.

III. 4 Rafrâchissement des supports:

Le rafraîchissement est une opération qui s'effectue périodiquement (certains parlent d'un cycle qui varie de trois à cinq ans) et qui consiste à recopier les données d'un support à un autre ou d'une technologie à une autre. Mais il faut savoir que cette technique de

sauvegarde et de préservation malgré son apparente simplicité possède quelques lacunes qu'il ne faut pas négliger à savoir le temps à consacrer à cette opération sachant que le nombre de documents électroniques croît d'une façon exponentielle. A noter que l'utilisation de l'une de ces méthodes n'exclut pas les autres. Elles peuvent être combinées pour assurer l'accessibilité complète aux documents électroniques à long terme. Mais ce qu'il faut prendre en considération lors du choix d'une ou plusieurs de ces méthodes c'est le ratio coût/efficacité.

IV. Propositions de solutions pour la préservation de documents électroniques:

L'étude des expériences étrangères en matière de préservation des documents électroniques, ainsi que l'ensemble des techniques développées à cet effet, nous a permis de constater que malgré les tentatives de mise en place d'un système collectif de dépôt légal (cas de NEDLIB), il n'existe pas de solution unique pour la préservation de ce type inédit de documents. Et ce dû en premier lieu à la divergence de conception du document électronique à préserver. Chose qui a amené chaque Etat à se doter d'un système de préservation de document électronique qui répond à ses objectifs et à la typologie des documents électroniques soumis au dépôt légal.

Néanmoins, force est de constater que l'ensemble des expériences étudiées opte pour une démarche commune pour la préservation des documents électroniques qu'on peut schématiser comme suit :

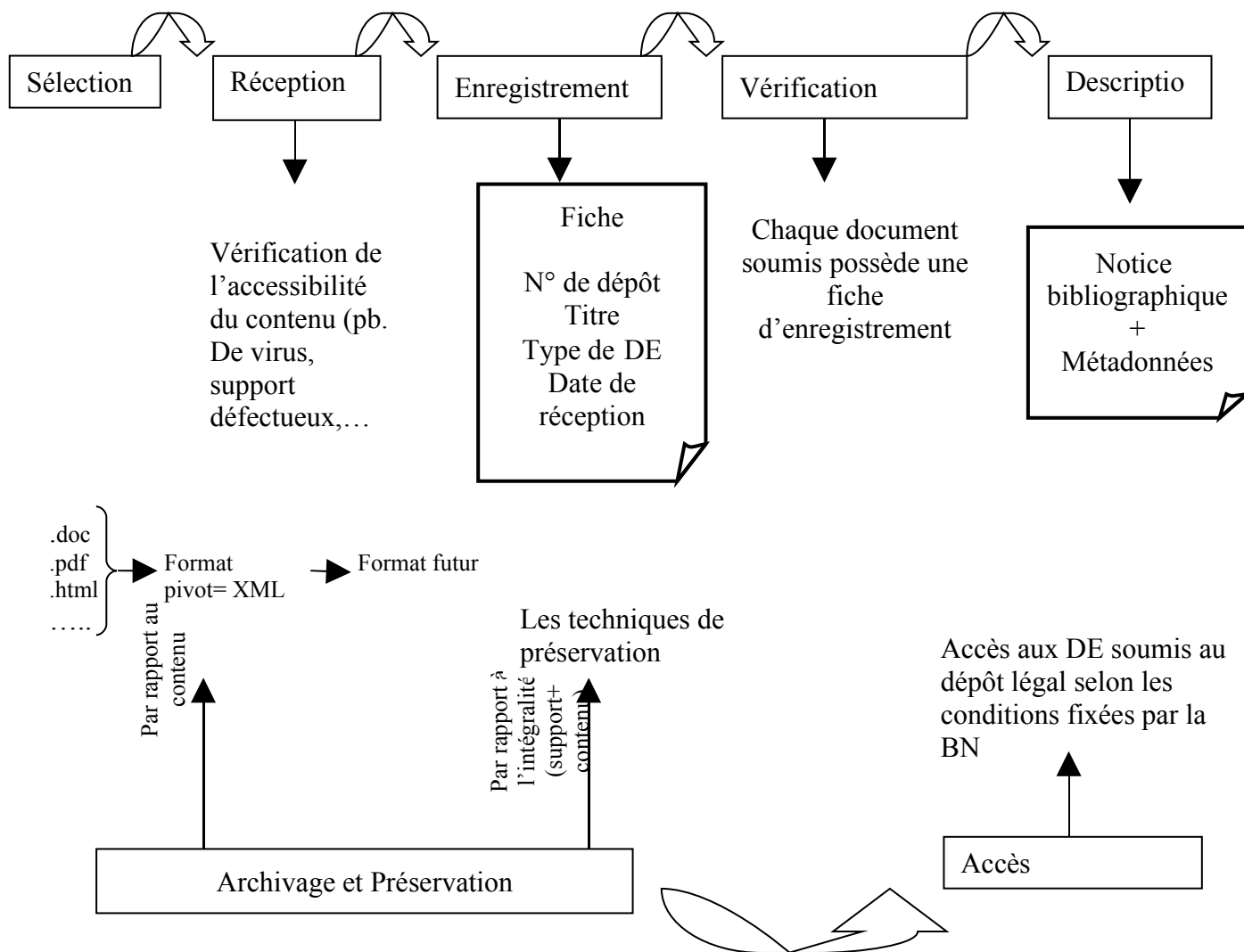


Figure 1 : Modèle proposé

Ce schéma retrace les principales étapes que doit parcourir un document électronique depuis sa réception jusqu'à sa préservation. Et c'est à partir de là que notre solution pour la préservation sera dégagée.

A noter, que la solution que nous préconisons est applicable dans toutes institutions concernées par la problématique de la préservation des documents électroniques. Elle constitue une sorte de feuille de route pour assurer la pérennisation d'un document électronique. Elle se présente comme suit :

IV-1- Sélection et réception du document électronique :

La sélection des documents électroniques, objet de la préservation, est tributaire des critères de sélection établis par l'instance compétente (exemple : pour le dépôt légal c'est la loi sur le dépôt légal qui définit les critères de soumission d'un document électronique). Après sélection, les documents réceptionnés sont soumis à une vérification du contenu pour s'assurer de leur intégralité et de la fiabilité du support. Cette opération est assurée par le

service chargé de l'archivage des documents électroniques (pour le dépôt légal, cette tâche est assurée par le service de dépôt légal de la bibliothèque nationale ou autre organisme dépositaire).

IV-2- Enregistrement et vérification des documents électroniques

Cette étape est réservée aux documents électroniques reçus en dépôt légal. Elle consiste à attribuer une fiche à chaque document. Elle comporte le numéro de dépôt légal, le titre, le type de document et la date de dépôt. Ces fiches viennent remplacer le registre du dépôt légal tenu par les institutions dépositaires. Elles sont également vérifiées de manière systématique.

IV-3 - Description bibliographique

L'étude des expériences étrangères en matière de préservation de documents électroniques nous a permis de constater que le système de préservation des documents électroniques de chacun de ces Etats n'est qu'une extension du système documentaire destiné à gérer leurs fonds documentaires. Ainsi, les institutions dépositaires qui disposent déjà d'un système documentaire peuvent l'utiliser comme point de départ pour la mise en place d'un système d'archivage des documents électroniques. Les institutions ne disposant pas d'un tel système documentaire, peuvent se doter d'un dispositif d'archivage de document électronique adapté à leurs besoins. En tout état de cause, le choix de l'une ou de l'autre solution se fait après concertation entre l'organisme dépositaire et le fournisseur du dispositif d'archivage.

IV- 4- Archivage et préservation des documents électroniques

Afin d'assurer la sauvegarde du contenu, le document électronique, quel que soit son format d'origine, doit être converti, dès sa réception, en format XML (qui est actuellement considéré comme le format pivot). Et ce tout en laissant la possibilité de le convertir vers les éventuels formats futur.

Quant aux techniques d'archivage et de préservation, l'institution bénéficiaire reste libre de choisir entre les techniques d'archivage et de préservation. L'institution bénéficiaire reste toujours libre de choisir entre les techniques d'archivage déjà citées.

Néanmoins, la plupart des techniques étudiées présentent certains inconvénients liés à leur application, notamment du point de vue coût. C'est ce que atteste l'étude des expériences avancées en matière de préservation des documents électroniques, nonobstant le contexte national en la matière.

A cet effet, nous recommandons l'utilisation du format XML comme format de préservation des documents électroniques, tenant compte du fait qu'il soit un format universel et normalisé, indépendant de toute technologie propriétaire et peu coûteux.

IV-5- Le Format XML comme format de préservation :

L'utilisation du XML comme format de préservation suppose la conversion du contenu du document électronique concerné vers ce format (XML). Pour ce faire, il existe deux types de conversion à savoir :

a. Les types de conversion : Il s'agit de

- Conversion sans utilisation de DTD
- Conversion avec utilisation d'une DTD

b. Les outils de conversion

Qu'on opte pour le premier ou le deuxième type, la conversion du contenu d'un document électronique vers le format XML nécessite l'utilisation d'un certain nombre d'outils à savoir,

Les éditeurs XML – XSL.

c. Mise en œuvre de la conversion vers XML

L'opération de conversion vers le format XML doit être assurée par un personnel qualifié. Dans la mesure où il se chargera de garantir une bonne structuration des documents archivés en vue d'assurer leur accessibilité à long terme. A cet effet, plusieurs scénarii sont proposés pour garantir la qualité des personnes chargées de la conversion en impliquant les différents acteurs : auteurs, éditeurs, organismes dépositaires ou autres institutions concernées.

c.1 Conversion centralisée (conversion par le personnel de l'organisme dépositaire ou autre institution) :

c.2 conversion décentralisée

c.2.1 Conversion avec l'aide de l'éditeur du document électronique :

c.2.2 Conversion avec l'aide de l'auteur du document électronique :

Chacun de ces scénarii présente des avantages et des inconvénients en termes de coût et de temps :

Pour le premier scénario, le facteur coût peut s'avérer élevé pour la plupart des institutions, comparé au facteur temps. Par ailleurs, les deux autres scénarii sont moins coûteux (la formation des éditeurs et autres peut être assurée par une seule personne) mais risque de s'étaler dans le temps puisque il n'est pas toujours possible d'arrêter au préalable la liste des bénéficiaires (former qui et pourquoi ?)

Ainsi, la combinaison des trois scénarii semble être plus opportune pour les organismes de préservation des documents électroniques, qui pourraient dans un premier temps opter pour le premier scénario tout en lançant des sessions de formation au profit des éventuels partenaires qui sont les auteurs et les éditeurs.

IV.6 Recherche et accès au document électronique préservé

La recherche et l'accès aux documents électroniques préservés se feront moyennant le logiciel documentaire utilisé par l'institution concernée. Quant aux conditions de mise à disposition du public des documents électroniques préservés, l'institution concernée fixe les modalités d'accès à ces documents.

Conclusion

L'étude de la problématique de préservation de documents électroniques nous permet de conclure, que les difficultés liées à la préservation de ce type de documents ne sont pas propres aux organismes chargés du dépôt légal. Et que, désormais, ces difficultés concernent toutes les institutions productrices d'informations numériques.

Cette étude nous permet, également, de constater qu'il n'existe pas encore une solution unique et universelle pour le problème de la préservation du patrimoine informationnel

numérique. Et ce en dépit des multiples tentatives émanant de grandes bibliothèques nationales à travers le monde. En effet, chaque pays développé adopte une solution temporaire propre à ses besoins locaux en matière de préservation. D'où les divergences dans les solutions et les approches adoptées par chacune de ces bibliothèques, à commencer par le problème de la définition d'un document électronique, le cadre juridique du dépôt des documents électroniques et les droits à protéger dans ce cadre, ensuite les techniques de préservation : allant de la sélection des documents électroniques à préserver, jusqu'à l'accès à long terme à ces documents.

Références bibliographiques

- [1] Archivage long terme des documents électroniques en physique
Long term archiving of electronic documents in physics (LTA-EDP-1)
<http://ccsd.cnrs.fr/lta1.html>
- [2] Collaboration entre les bibliothèques nationales en vue de conserver l'information numérique. Nouvelles de la Bibliothèque nationale Mars 1999 Vol. 31, n° 3-4
<http://www.nlc-bnc.ca/9/2/p2-9903-03-f.html>
- [3] Décret n° 93-1429 du 31 Décembre 1993 relatif au dépôt légal
<http://www.bnf.fr/pages/infopro/depoleg/dl-decre.htm>
- [4] Danielle Léger. Dépôt légal et publications diffusées sur le Web : Etat des travaux à la bibliothèque nationale du Québec. 28^e Congrès de l'ASTED.
- [5] Equipe du projet pilote des publications électroniques et comité des collections électroniques.- Projet pilote sur les publications électroniques (PPPE) : Rapport final.- Canada : bibliothèque nationale du Canada, 1996
[http:// collection.nlc-bnc.ca/e-coll-e/f-report.pdf](http://collection.nlc-bnc.ca/e-coll-e/f-report.pdf)
- [6] Jean-Michel Rodes, Geneviève Piejut, Emmanuèle Plas. La mémoire de la société de l'information. Publications de l'UNESCO pour le Sommet mondial sur la société de l'information.
- [7] Gail M.Hodge. Best practices for digital archiving: An information life cycle approach. <http://www.dlib.org/dlib/january00/001lodge.html>
- [8] Les actes : Documents électroniques et conservation patrimoniale : le rôle des bibliothèques nationales.
http://www.idt.fr/idt/pages_fra/actes/actes98/act20.html
- [9] Les périodiques électroniques accessibles à distance et la Bibliothèque Nationale de Norvège. 64^{ème} Conférence Générale de l'IFLA 16 août 1998
<http://www.ifla.org/IV/ifla64/024-141f.htm>
- [10] Loi n° 92-546 du 20 juin 1992 relative au dépôt légal
<http://www.bnf.fr/pages/infopro/depoleg/dl-loi.htm>
- [11] Paul McCormick. Le dépôt légal au Canada.
<http://www.nlc-bnc.ca/obj/r7/f2/consult3-f.pdf>
- [12] Rapport du directeur général concernant un projet de charte sur la préservation du patrimoine numérique. 164 EX/21 PARIS, le 9 avril 2002Original anglais
- [13] The Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project
<http://interpares.org/book/index.cfm>

- [14] Witt Maria, Ihadjadene Majid, Archives, Libraries and Museums Convergence= Archive, Bibliothèque et Musées, Paris, 2001, 445p.

Sitographie

[15] http://www.adbs.fr/uploads/journees/2293_fr.php

[16] <http://www.kb.nl/coop/nedlib>

[17] <http://revues.enssib.fr/titre/6tecnic/5archivage/2formats.htm>

[18] http://www.chin.gc.ca/Francais/Contenu_Numerique/Preservation_numerique/guides.html#eval