

L'archivage des données du Web (1ère partie)

[Retour au sommaire de la lettre](#)

A mesure que croît la richesse du Web, la question de l'archivage des données numériques se pose de plus en plus. Faut-il conserver une trace de tous les articles, témoignages, vidéos et autres documents audiovisuels disponibles sur le Net ? Ces contenus ne présenteront-ils pas en effet un intérêt dans quelques décennies, notamment pour les historiens et les sociologues ? Pour éviter que ces données ne tombent progressivement aux oubliettes du Net, plusieurs solutions d'archivage ont d'ores et déjà vu le jour. Ces systèmes sont pour l'instant surtout développés par de grandes institutions dédiées à la conservation du patrimoine et ils pourraient aussi être proposés à l'avenir par des moteurs de recherche souhaitant enrichir leurs contenus. Nous vous proposons ici un panorama de ces outils d'archivage du Web. Le mois prochain, nous publierons plusieurs entretiens avec des responsables de grandes bibliothèques nationales à ce sujet.

En France, les livres sont archivés depuis 1537 (le dépôt légal a été instauré par François 1^{er}), les estampes depuis 1648, les partitions musicales depuis 1793, les photographies et les phonogrammes depuis 1925, les vidéogrammes depuis 1975 et l'édition électronique sur support depuis 1992. Depuis 1999, la BnF prépare également l'extension du champ du dépôt légal aux sites Internet mais le dépôt légal obligatoire des sites Web n'est pas encore en vigueur.

Le fait de légiférer pour le dépôt légal des pages Web ne suffira pas à garantir la sauvegarde des ressources électroniques. Le caractère immatériel et mouvant du Web nécessite en effet la création de nouveaux outils d'archivage automatiques et performants. Les techniques utilisées par ces outils d'archivage ne sont pas vraiment nouvelles et elles sont très proches de celles qui sont utilisées par les moteurs de recherche. L'archivage des pages Web nécessite aussi le déploiement par les bibliothèques de nouveaux moyens financiers, notamment pour l'acquisition et la maintenance des serveurs de stockage.

Nous allons nous concentrer ici sur l'archivage des données du Web et nous n'aborderons pas la question de la numérisation des ouvrages des bibliothèques. Nous traiterons ultérieurement ce sujet qui suscite actuellement de nombreux débats en France suite à l'annonce par Google, en décembre 2004, d'un accord visant à numériser et à mettre en ligne dans plusieurs années environ 15 millions d'ouvrages provenant de cinq grandes bibliothèques américaines (<http://actu.abondance.com/2004-51/google-print.html>).

Quelles sont les solutions pour retrouver un document Web qui a disparu ?

Qui n'a pas un jour perdu la trace d'une information importante publiée à un moment donné sur un site Internet avant d'être effacée par son auteur ? Sans avoir à contacter l'auteur du document, il existe plusieurs solutions pour retrouver un document qui a disparu.

La première méthode, moyennement fiable, consiste à **utiliser la fonction "Cache" de Google**. Si vous vous souvenez d'une expression exacte de ce document, vous pouvez saisir une requête dans Google avec cette requête entre guillemets. Ou bien, vous pouvez tenter de retrouver cette page en effectuant une recherche avec des mots clefs susceptibles d'y apparaître. Si vous retrouvez la page dans la liste de résultats, vous pourrez alors la consulter en cliquant sur le lien "En cache" proposé juste en dessous du résultat en question.

Chaque page indexée par Google est sauvegardée par le moteur dans une mémoire cache pendant toute sa durée de vie. Cependant, cette solution est peu sûre car elle ne fonctionne que si le spider (robot indexeur) de Google n'a pas encore découvert que cette page n'est plus active. La vocation première de Google n'est en effet pas l'archivage mais, au contraire, la fourniture de résultats aussi actuels que possible et Google élimine systématiquement de son serveur les pages correspondant à des liens disparus. De plus, le lien "En cache" n'apparaît pas toujours, le propriétaire du site ayant la possibilité de demander à ce qu'un contenu ne soit pas sauvegardé en cache par Google. Point important cependant : il semblerait que Google "stocke" actuellement le contenu de ses "caches" pour utilisation ultérieure. Des rumeurs nous ont indiqué que ce stockage s'effectuait depuis plus de trois ans au sein du moteur de recherche. Il ne nous a malheureusement pas été possible d'obtenir une confirmation officielle de ce fait par Google. Difficile, donc, de dire si ce stockage est effectif...

Notons enfin que Google n'est pas le seul moteur à proposer un cache. Des outils comme Yahoo! (<http://search.yahoo.fr/>), MSN (<http://search.msn.fr/>) ou Dir.com (<http://fr.dir.com/>) ont également opté pour ce type de fonctionnalité.

L'**interrogation d'un outil spécialisé dans l'archivage du Web** offre une deuxième possibilité pour retrouver une page disparue. Le site Wayback Machine (www.archive.org) permet notamment d'effectuer une recherche par URL dans les archives du Web. Si la page a été prise en compte par l'archiveur automatique de ce site, elle pourra être retrouvée et consultée. D'autres outils nationaux, le plus souvent initiés par de grandes bibliothèques, sont également dédiés à l'archivage des pages Web d'une zone géographique. C'est le cas, par exemple, de PANDORA en Australie qui permet de visualiser les archives d'une sélection de sites australiens organisés par thématiques.

Typologie des acteurs intervenant dans l'archivage du Web

Plusieurs catégories d'acteurs proposent des outils d'archivage du Web :

- Les **grandes bibliothèques nationales** : ces institutions ont historiquement pour mission de préserver la mémoire collective nationale pour les générations à venir. Jusqu'ici surtout limitée aux documents écrits et audiovisuels, leur mission d'archivage est en train d'être progressivement étendue aux documents électroniques. Les bibliothèques nationales sont d'autant plus sensibles à cette mission d'archivage du Web qu'elles réalisent que certaines publications sont désormais exclusivement publiées en ligne, et non sur support papier.
- Les **associations et organismes de recherche** : la fondation américaine à but non lucratif Internet Archive s'est spécialisée dans l'archivage systématique des sites internationaux, avec son outil Wayback Machine.
- Les **moteurs de recherche** conservent les pages indexées par leur spider "en cache" pendant toute leur durée de "validité".

Les bibliothèques nationales ont commencé leurs expérimentations en 1994 et de plus en plus de bibliothèques nationales lancent aujourd'hui des initiatives d'archivage du Web.

Années de lancement des principaux outils

	Pays	Auteur (s) - Projet
1994	Canada	Bibliothèque Nationale du Canada - EPPP (Electronic Publication Pilot Project)
1996	Suède	Bibliothèque Royale de Suède - Kulturarw3
1996	US	Internet Archive - WayBack Machine
1996	Australie	National Library Of Australia - PANDORA
1997	Finlande	Helsinki University Library - EVA
2000	République Tchèque	National Library of the Czech Republic - WebArchiv
2001	Autriche	Austrian National Library - AOLA (Austrian On-Line Archive)
2002	US	Bibliothèque du Congrès - MINERVA
2004	UK	British Library / National Archives / National Library of Wales / National Library of Scotland / JISC / Wellcome Trust - UKWAC (UK Web Archiving Consortium)

Processus techniques mis en œuvre pour l'archivage du Web

Crawling et collecte des pages

Les outils actuels ont le choix entre deux approches pour l'archivage :

- la **sélection d'un nombre limité de sites** suivant des critères définis par les administrateurs.
- le **crawling et la capture automatisée et systématique de toutes les ressources du Web** (national ou non). Cette méthode a été choisie par les bibliothèques de Finlande, de Norvège ou de Suède. Pour ses partisans, il est impossible de définir les documents qui seront intéressants pour les générations à venir et donc nécessaire de tout sauvegarder. L'augmentation des capacités de stockage facilite également cette opération. L'exhaustivité absolue sur un pays n'en reste pas moins impossible, dans la mesure où certains sites "géographiques" sont inaccessibles et/ou hébergés à l'étranger...

Chaque site repéré est aspiré à distance par un logiciel spécialisé (également appelé "web harvester" ou "harvester robot") chargé de récupérer un ensemble de pages pour qu'elles soient stockées suivant des critères précis. De la même manière que les spiders ou crawlers des moteurs de recherche, l'archivageur automatique analyse les pages sauvegardées pour identifier de nouveaux liens et les sauvegarder à leur tour s'ils sont jugés pertinents. Parmi les archiveurs connus, on peut citer NEDLIB (pour "Networked Deposit Libraries"), un archiveur gratuit développé à l'occasion d'un projet financé par l'Union européenne et utilisé notamment pour l'archivage du Web finlandais (voir <http://www.lib.helsinki.fi/tietolinja/0203/webarchive.html>) ou encore le logiciel opensource HTTrack, utilisé notamment par la Bibliothèque Australienne.

Pour sa collecte du Web, la bibliothèque nationale suédoise utilise quant à elle le système "Combine" pour tous les domaines en ".se". Ce système a été développé par l'université de Lund, avec le soutien financier de la Commission européenne (<http://www.lub.lu.se/combine/>). Cette bibliothèque utilise également le crawler open source Heritrix de l'Internet Archive pour la sauvegarde quotidienne de la presse suédoise (voir <http://crawler.archive.org/index.html>).

Tous les sites ne peuvent hélas être "crawlés" et archivés et, lorsqu'il est impossible pour le robot d'aspirer un site, certaines bibliothèques demandent à l'éditeur de fournir son contenu par dépôt volontaire (la BnF compte procéder ainsi avec son nouvel outil).

Stockage des documents

Le stockage est l'un des principaux enjeux à relever pour tout projet d'archivage du Web. Des capacités de stockage importantes sont en effet nécessaires pour la base de données contenant l'archiveur et les différents critères de sélection des pages, pour une autre base de données ou les documents traités sont analysés et, enfin, pour le stockage des documents archivés.

Pour limiter les espaces de stockage nécessaires, l'outil doit éliminer les pages en doublons et les sites miroirs. Afin de limiter les espaces de stockage utilisés, il est possible de compresser les pages archivées (un outil de décompression doit alors être utilisé pour permettre la visualisation de la page).

Pour qu'elles puissent être visualisées par la suite, les pages archivées doivent être "préparées", avec entre autres le remplacement des liens Internet par des liens locaux, la suppression des données "inutiles" telles que les adresses e-mail...

Analyse et indexation des pages archivées

Un ensemble de critères (définis par les administrateurs) permet de définir les pages qui doivent être archivées. Les pages pourront, par exemple, être sauvegardées uniquement si leurs URL correspondent à un nom de domaine géographique pertinent (par exemple, ".fr" s'il s'agit d'un outil spécialisé sur l'archivage du Web français), si elles sont rédigées dans une langue donnée ou si elles ont été publiées par des organismes nationaux (une solution française pourrait, par exemple, décider de retenir les versions anglaises du site Renault.com).

Un outil d'indexation est ensuite utilisé pour créer les métadonnées correspondant aux pages archivées. Le processus d'indexation est quasiment le même que celui des moteurs de recherche, avec une extraction systématique des mots clefs. La bibliothèque de Finlande utilise, par exemple, la technologie FAST depuis 2001 pour indexer les contenus Web qu'elle sauvegarde.

Pour un outil d'archivage du Web, le moteur d'indexation doit également recueillir des métadonnées propres aux archives :

- l'identifiant de chaque page archivée.
- les informations sur la situation géographique de cette page au moment de sa collecte.
- les données concernant la dernière date de sauvegarde de chaque page...

Descriptif de plusieurs outils / projets d'archivage du Web

Wayback Machine

Aux Etats-Unis, l'Internet Archive, une association à but non lucratif, financée par plusieurs entreprises, a été créée en 1996 par l'entrepreneur Brewster Kahle (également fondateur de la société Alexa, qui est désormais une filiale du groupe Amazon). Cette association est à l'origine de l'un des outils d'archivage du Web les plus puissants et exhaustifs, consultable gratuitement à l'adresse <http://www.archive.org/>. Baptisé "Wayback Machine" (la machine à remonter le temps), cet outil effectue des sauvegardes régulières d'un ensemble de pages Web, de fichiers textes, de vidéo et de logiciels. Entrez, par exemple, l'URL d'un site en première page et vous pourrez consulter les différentes versions de ce site sauvegardées par la "Wayback Machine" (196 sauvegardes à ce jour pour le site Abondance.com !).

Le nombre de pages sauvegardées depuis 1996 est évalué à 40 milliards, avec une taille de 600 téraoctets de données, qui croît d'environ 12 téraoctets par mois. L'outil est un grand succès puisqu'il reçoit actuellement plus de 20 000 visites par jour.

Heritrix, le crawler / archiveur opensource développé par l'Internet Archive pour sa collecte du Web, est en outre proposé en téléchargement gratuit sur Sourceforge et utilisé par certaines bibliothèques (http://sourceforge.net/project/showfiles.php?group_id=73833&package_id=73980).

Entretien avec Michele Kimpton - Director Web Archive – Internet Archive (www.archive.org)

Quels sont les principaux pré-requis techniques pour archiver le Web ?

Une tonne de capacité de stockage dans votre archive est nécessaire (nous avons plus de 40 milliards de pages) donc il vous faut des pétaoctets d'espace de stockage. Il faut également beaucoup de bande passante pour télécharger de gros volumes de données en très peu de temps.



Quelles sont les difficultés principales rencontrées lors de l'archivage du Web ?

Vous pouvez tomber dans des "pièges", comme des pages inutiles le plus souvent générées dynamiquement lors du processus de web crawling. De même il arrive de capturer le même contenu plusieurs fois car il se trouve hébergé par plusieurs sites différents. Lors du crawling, il faut faire attention à être "poli" et à ne pas se mettre à dos les administrateurs de sites Web.

Que pensez-vous des initiatives d'archivage du Web lancées par les moteurs de recherche ?

C'est excellent, plus les initiatives se multiplient, plus cela signifie qu'un même contenu sera conservé dans de multiples endroits différents. Cela est important car il n'existe pas de copies papier des pages Web.

Les recherches s'effectuent par URL. Wayback Machine s'est également doté d'un module de recherche plein texte (Recall) et il est possible d'effectuer une recherche par mots sur 11 milliards de pages (voir <http://www.dsi-info.ca/actualites/breves.html>).

Enfin, depuis 2003, le site propose également un "bookmarklet", à savoir un petit bouton "Wayback" s'intégrant dans la barre d'outils du navigateur et permettant d'accéder en un clic à l'historique complet du site que vous êtes en train de consulter.

BnF

En France, la BnF a confirmé son intention de "numériser le web francophone" et elle attend actuellement un décret du parlement pour lancer officiellement son projet

(<http://www.neteconomie.com/perl/navig.pl/neteconomie/infos/article/20050217150044>).

En attendant ce décret, la Bibliothèque nationale a déjà commencé à déployer un premier outil d'archivage du Web. Cet outil surveille un échantillon représentatif d'environ 400 sites francophones, des sites supplémentaires étant susceptibles d'être ajoutés suivant une logique événementielle. Un site pourra être uniquement archivé à l'occasion d'un événement politique majeur, comme une élection politique, par exemple. De plus, la BnF est en pourparlers avec l'INA pour l'archivage des contenus audiovisuels.

A noter que la BnF a déjà conduit plusieurs expériences concernant les sites électoraux lors de la présidentielle, des législatives, des régionales et des européennes, entre 2002 et 2004

(http://www.bnf.fr/pages/infopro/depotleg/dli_intro.htm).

Pour des raisons de droits d'auteur, l'outil sera uniquement consultable sur place à la BnF.

PANDORA

Initié en 1996 par la Bibliothèque nationale d'Australie, PANDORA (acronyme de "Preserving and Accessing Networked DOcumentary Resources of Australia") est un outil d'archivage portant sur une sélection de sites Web australiens. Neuf bibliothèques australiennes participent désormais à ce projet mais la Bibliothèque nationale reste en charge de l'administration du système et des serveurs.

**Entretien avec Margaret E Phillips - Director Digital Archiving –
National Library of Australia (<http://pandora.nla.gov.au/>)**

Quelles sont les principales difficultés que vous rencontrez lors de l'archivage des sites Web ?

Jusqu'à récemment, une difficulté majeure concernait l'archivage des sites dépendant de bases de données pour la présentation de l'information. Comme toutes les bibliothèques nationales, notre collecte de sites Web a dépendu de la capacité des web crawlers à accéder à l'information requise via les liens html. La NLA a développé un outil appelé Xinq permettant de créer une interface utilisateur générique pour accéder aux contenus des éditeurs. Cet outil a été développé dans le cadre de la contribution de la NLA au "Deep Web Archiving Working Group" du Consortium International pour la Préservation de l'Internet (IIPC). Voir <http://www.netpreserve.org/about/index.php>. Le projet a été dirigé par Monica Berko, Director Applications Development at the NLA en collaboration avec la Bibliothèque Nationale de France (BnF). Xinq est en phase de devenir disponible via SourceForge sur <http://sourceforge.net/projects/xinq/>. Une autre difficulté concerne les droits de propriété des sites. Il faut obtenir l'autorisation de l'éditeur avant de copier le site dans l'archive. Presque tous les éditeurs acceptent mais cela prend du temps. Une autre difficulté concerne le volume de contenu à archiver, disproportionné par rapport à nos ressources limitées.



Selon vous, quelle est l'importance de l'archivage des données du Web pour les bibliothèques ?

Selon moi, archiver le Web est vital. Sans les initiatives menées par les bibliothèques, une part significative de notre patrimoine documentaire, reflet essentiel de notre société et culture, risque d'être perdu pour les générations futures mais aussi pour notre génération. Je voudrais insister sur le fait que la COLLECTE n'est pas suffisante. Il faut également mettre en place des programmes dynamiques de PRESERVATION afin de garantir l'accès futur aux sites archivés et ce, malgré l'évolution de l'environnement informatique (hardware et software) qui les rend vite obsolètes. Je pense même que la viabilité des démocraties peut-être mise en danger si nous ne trouvons pas le moyen de collecter et de préserver toute l'information émanant des gouvernements et qui souvent n'est publiée que sur le web. Conserver cette information est aussi essentiel que de construire des ponts, des routes ou des hôpitaux. Les communautés nationales doivent comprendre la nécessité de subventionner ce service essentiel.

PANDORA est un outil sélectif d'archivage du Web, dont les critères d'acquisition correspondent à la politique d'acquisition de la Bibliothèque nationale. Une réflexion est néanmoins apparemment en cours pour faire évoluer l'outil vers une sauvegarde exhaustive du Web australien. Pour des raisons de droits d'auteur, les partenaires de PANDORA doivent demander une autorisation à chaque éditeur de site avant de pouvoir intégrer son site dans leurs archives.

Pour collecter les données, à défaut de pouvoir acquérir un système "packagé" convenant à ses besoins, la Bibliothèque nationale a développé son propre archiveur interne dont la première version est sortie en 2001. Baptisé PANDAS (pour "PANdora Digital Archiving System"), ce système permet :

- d'attribuer des métadonnées aux pages archivées (titre, URL, format...), ainsi qu'un numéro identifiant unique à chaque site archivé et aux différents éléments qui constituent ce site (les différents articles d'une publication, par exemple, sont tous numérotés).
- d'organiser la collecte des titres à archiver (un site peut être archivé une seule fois ou une collecte fréquente peut être organisée ; PANDORA utilise le crawler open source HTTrack).
- de préparer la version finale du site archivé, en désactivant les liens externes, les formulaires et les adresses emails.
- de générer automatiquement un titre pour chaque page.

Chaque site archivé est automatiquement revisité par un logiciel (Linkbot) pour vérifier si toutes les pages ont bien été sauvegardées...

Les contenus archivés sont consultables gratuitement :

- sur le site <http://pandora.nla.gov.au/> ; en utilisant l'annuaire thématique (un moteur de recherche devrait être aussi offert prochainement)

- via la base de données bibliographiques "Kinetica" : des liens vers les sites archivés sont proposés aux utilisateurs en complément des références bibliographiques
- via des moteurs de recherche tels que Yahoo ! et Google qui indexent les titres des pages archivées (mais pas leurs contenus)

Web Archiving Consortium

Au Royaume-Uni, à l'initiative de la British Library, plusieurs bibliothèques ont créé en juin 2004 le UK Web Archiving Consortium (UKWAC - <http://www.webarchive.org.uk>). L'objectif de cette organisation est d'archiver un premier échantillon de 6000 sites d'ici 2006, répondant aux politiques d'acquisition de différentes bibliothèques adhérant à ce programme. Les sites devront donner leur autorisation préalable pour être archivés. Une première version de cet outil devrait être lancée prochainement. Les coûts de ce nouvel outil (hardware, software, développement et maintenance) seront équitablement répartis entre les membres du Consortium.

Pour collecter les données, le Consortium va utiliser le crawler open source HT Track. Les données seront ensuite analysées et archivées grâce à la solution PANDAS, le système développé par la Bibliothèque nationale Australienne pour son outil Pandora (voir plus haut). Le déploiement des infrastructures de stockage et la maintenance du système seront assurés par la société Magus Research Limited.

Les sites archivés pourront être consultés gratuitement en ligne via un annuaire et/ou un moteur de recherche.

Enjeux et perspectives

L'archivage du Web est une tâche colossale. Tout d'abord, la taille du Web et son évolution croissante rendent difficile l'archivage des contenus (on estime à 5 à 10 millions le nombre de pages créées chaque jour). Ensuite, le Web est un support « mouvant », en perpétuelle évolution, ce qui ne facilite pas la tâche. **De plus, l'archivage de certaines données peut poser des problèmes juridiques, notamment en ce qui concerne les données nominatives et le respect du « droit à l'oubli »** (voir http://www.archivesdefrance.culture.gouv.fr/fr/archiviste/Expose_Archives_electroniques.p.ppt).

Pour parvenir à archiver le Web, les outils spécialisés doivent relever plusieurs enjeux techniques :

- Il leur faut trouver des moyens pour archiver le Web invisible : de plus en plus de sites sont dynamiques et constitués à partir de documents provenant de différentes bases de données. Ces contenus sont difficiles à repérer car ils ne sont liés à aucune page.
- Outre le fait de sauvegarder les données Web, les bibliothèques doivent s'assurer de la pérennité des informations collectées. Lorsque les formats ou les applications évoluent, les bibliothèques doivent en effet disposer des anciennes applications permettant de lire ces pages, créer des émulateurs (programme permettant de faire fonctionner un système comme un autre système a priori non compatible) ou encore assurer la conversion des données dans de nouveaux formats.

Enfin, de nombreuses organisations nationales développent des partenariats nationaux ou internationaux en vue de mutualiser les coûts de fonctionnement, de réaliser des économies d'échelle et de garantir une interopérabilité entre les systèmes. C'est le cas du UK Internet Archive Consortium (voir plus haut) ou encore du Consortium Netpreserve.org (<http://www.netpreserve.org>), créé en 2003 et coordonné par la BnF. Ce Consortium réunit les bibliothèques nationales de France, d'Australie, du Canada, du Danemark, de Finlande, d'Islande, d'Italie, de Norvège, de Suède, du Royaume-Uni (British Library), ainsi que la Library of Congress américaine et la fondation US Internet Archive. Il a pour objectif "d'approfondir une collaboration technique centrée sur l'identification, l'élaboration et la mise en œuvre d'instruments et de procédés propres à sélectionner, collecter, préserver et rendre accessibles les contenus de la Toile" et "d'établir progressivement un inventaire des collections des contenus de l'Internet, dans le respect des législations de chaque nation et en fidélité aux politiques de sélection propres à chacune".

Conclusion

Les débats sont nombreux en France sur la nécessité d'adopter une approche européenne commune pour la numérisation des ouvrages de bibliothèques (voir <http://www.01net.com/article/271408.html>). Il semble toutefois qu'un peu moins d'attention soit actuellement apportée à la question de l'archivage du Web, un sujet qui devrait pourtant nous préoccuper alors que nous risquons de voir progressivement disparaître la multitude de documents Web actuellement disponibles !

Soucieux d'atteindre l'exhaustivité, certains moteurs de recherche pourraient tenter d'entrer sur le marché de l'archivage Web, un domaine jusqu'ici surtout exploré par les grandes bibliothèques. Mais les grandes bibliothèques ont une carte à jouer face à l'exhaustivité promise par certains moteurs, en faisant notamment valoir leur capacité de traitement et d'analyse des informations.

Quelques liens...

Articles

Archiving the Web: European experiences - Juha Hakala – Bibliothèque nationale de Finlande
<http://www.lib.helsinki.fi/tietolinja/0203/webarchive.html>

Archiving the Web: The PANDORA Archive at the National Library of Australia - Warwick Cathro, Colin Webb et Julie Whiting (Bibliothèque nationale d'Australie)
<http://www.nla.gov.au/nla/staffpaper/2001/cathro3.html>

Outils spécialisés

- PANDORA : <http://pandora.nla.gov.au/>
- Wayback Machine : <http://www.archive.org/>
- WebArchiv : <http://webarchiv.nkp.cz>
- UK Web Archiving Consortium : <http://www.webarchive.org.uk>

Conférence

Chaque année la conférence "International Web Archiving Workshop and Digital Preservation" réunit les spécialistes de l'archivage des contenus numériques. La 5e édition de cette conférence se tiendra cette année à Vienne du 22 au 23 septembre 2005. Pour plus d'information, voir <http://www.iwaw.net/05/index.html>.

Note : cet article a été écrit par un journaliste free-lance pour le compte d'Abondance