

Archivage et diffusion électroniques des thèses et mémoires à l'université Cheikh-Anta-Diop de Dakar :

→ **UN PAS VERS LE DÉSENCLAVEMENT SCIENTIFIQUE
DE L'AFRIQUE**

DIÉYI DIOUF

*Bibliothèque centrale
de l'université Cheikh-Anta-Diop,
Dakar, Sénégal
dieyidiouf@ucad.sn*

*Titulaire du diplôme supérieur en
sciences de l'information et de la
communication de l'université Cheikh-
Anta-Diop de Dakar, Diéyi Diouf est
conservateur à la bibliothèque centrale
de l'UCAD.*

Le ministère français des Affaires étrangères et européennes (MAEE) s'est engagé, à travers le projet Système d'information scientifique et technique (SIST)¹ du Fonds de solidarité prioritaire (FSP), à désenclaver et promouvoir la recherche africaine par un meilleur accès à l'information scientifique et technique. Dans le cadre de ce projet, un atelier de formation à la numérisation, à la création, à la gestion et à la diffusion d'une collection de documents électroniques s'est tenu en octobre 2007 à Antananarivo (Madagascar). Un volet de cet atelier a été consacré à la présentation de projets de création de bibliothèque électronique, en cours ou achevés, intégrant des actions de numérisation de fonds documentaires, de gestion et de diffusion de collection de documents électroniques. Ces retours d'expériences ont constitué une base de réflexion sur la conduite d'un projet de numérisation, la mise en accès de fonds documentaires numériques, et leur valorisation via la plateforme SIST. Cet article rapporte l'expérience conduite à la bibliothèque centrale de l'université Cheikh-Anta-Diop de Dakar (Bucad) au Sénégal.

Le projet de numérisation de l'université de Dakar

L'université Cheikh-Anta-Diop de Dakar (Ucad) est la plus grande université d'Afrique de l'Ouest. Elle regroupe des facultés autonomes, des écoles nationales supérieures, des instituts d'enseignement et des instituts de recherche. Elle compte aujourd'hui environ 55 800 étudiants et plus de 1 200 enseignants-chercheurs. Depuis 2003, l'université s'est engagée dans une profonde mutation, structurée à partir de grands projets où les techniques informatiques occupent une place de premier ordre.

Des méthodes et des outils plus performants ont favorisé le développement de la recherche. Cette dynamique s'est traduite par une forte augmentation des produits de la recherche et du nombre d'ouvrages universitaires (250 à 300 environ par an) déposés à la Bucad.

Face à cet accroissement de production et au-delà des questions de stockage et de conservation, les responsables de la bibliothèque se sont interrogés sur leur mission d'information scientifique et plus particulièrement sur une diffusion

1. www.sist-sciencesdev.net

accrue des savoirs auprès de la communauté universitaire nationale et internationale.

L'acquisition des ouvrages et des revues devenant de plus en plus difficile en raison de leur coût, il s'agissait de créer les conditions d'un partage et d'un accès gratuit au plus grand nombre des documents produits par un pôle universitaire africain d'envergure. C'est dans ce contexte que les autorités de la bibliothèque universitaire ont pris l'initiative de mettre en place un nouveau service d'archivage et de diffusion électroniques chargé de garantir la gestion et la visibilité des travaux de l'Ucad.

La création du service de numérisation

L'objectif majeur recherché à travers la mise en place du « Service de numérisation des publications académiques » était, en assurant le traitement électronique des documents, de conserver de manière pérenne et de diffuser en texte intégral les thèses et travaux de recherche déposés à la bibliothèque centrale de l'université. Le projet de numérisation visait, à travers la mise en ligne sur internet des documents universitaires, à la fois un accès plus large et plus rapide aux contenus scientifiques, une meilleure valorisation des résultats de la recherche et la promotion d'une littérature grise le plus souvent peu connue et peu exploitée par les enseignants-chercheurs et les étudiants tant sur le plan national qu'international.

Pour atteindre ces objectifs, la nouvelle unité devait se doter d'une base de travail informatique abordable techniquement et financièrement. Le choix s'est porté sur la plateforme Cyberdocs. Cyberdocs est un outil informatique de production, d'archivage et de diffusion de documents numériques structurés. Cette plateforme a été développée à partir du programme Cyberthèses, lancé en 1998 par les presses de l'université de Montréal et l'université Lumière (Lyon-2), avec le soutien du fonds francophone des inforoutes. Ce programme s'est transformé au fil des ans en un programme international de coopération entre les universités de Lyon, Montréal, Genève, Santiago du Chili, Dakar, Antananarivo. Le réseau Cyberthèses regroupe aujourd'hui de nombreuses institutions, francophones ou non, à travers le monde².

L'application de la plateforme Cyberdocs

La plateforme Cyberdocs est entièrement libre et gratuite. Elle se compose d'un ensemble d'outils qui permettent la conversion et la mise en ligne des thèses et des mémoires tout en offrant la possibilité de traiter d'autres types de documents.

Le processus de publication électronique de la plateforme Cyberdocs vise à transformer un document tel

La plateforme Cyberdocs

L'ensemble de la chaîne se structure sur la base de logiciels libres* et de pilotes qui lui sont associés, c'est-à-dire :

- une suite bureautique « OpenOffice.org » comprenant un traitement de texte, un tableur, un module de dessin vectoriel, une base de données, etc. ;
- un serveur web Apache, logiciel libre de serveur HTTP produit par Apache Software Foundation ;
- un environnement de développement intégré (PHP), interface permettant de développer, compiler et exécuter un programme dans un langage de script PHP (Hypertext Preprocessor) et qui est utilisé pour produire des pages web dynamiques via un serveur HTTP ;
- une machine virtuelle Java, programme qui exécute des applications et des programmes développés en langage Java ;
- un moteur de servlet, application permettant d'exécuter de petits programmes échangeant des informations avec le serveur et affichant ensuite la page sur l'écran de l'internaute sur un serveur web ;
- un système de mise en ligne et de recherche documentaire SDX.

D.D.

* Protégés par la licence GPL (General Public License).

qu'il est fourni par les auteurs, c'est-à-dire un fichier au format traitement de texte, en un document structuré au format XML, document à la fois de référence et qui sert à l'archivage et à la publication.

Cyberdocs comporte un module de conversion, un module de publication et un module de gestion. L'ensemble du processus de traitement d'un document comprend :

- une phase de préparation des documents qui consiste à utiliser des feuilles de styles pour identifier les différentes parties et la structure du document (niveau des titres, polices de caractères, tableaux, illustrations...);
- une phase de préparation, où des métadonnées sont associées au document ;
- une phase de conversion au cours de laquelle le module de conversion de Cyberdocs convertit le document ainsi préparé en un document XML; cette phase permet également de transformer ce document XML en un document HTML, XHTML ou PDF pour une diffusion en réseau ou sur cédérom ;
- la phase de publication, réalisée par le module de publication – application SDX –, assure la disponibilité du document, sa consultation ainsi que des recherches à l'intérieur du document.

Mise à part la restructuration du document, l'ensemble du processus de traitement est automatisé et ne nécessite pas d'intervention spécifique.

Cette plateforme Cyberdocs a été installée à l'Ucad grâce à une mission d'expert. La plateforme, qui peut être adaptée aux besoins des différentes institutions (adaptation aux types de documents, etc.), a été personnalisée par

2. www.cybertheses.org

l'équipe interne de la BU avec l'appui du service informatique et a été dénommée Cyberdocs-Ucad.



Logo Ucad

Mise en place du service et organisation du circuit de numérisation

Les intervenants et leurs rôles respectifs

Lors de la création du service, quatre personnes ont été détachées du personnel de la bibliothèque pour constituer l'équipe de numérisation: le bibliothécaire porteur du projet de 2002 à 2003 qui, formé à l'archivage électronique a lui-même formé la nouvelle équipe composée d'un autre bibliothécaire chargé de sensibiliser les étudiants et chercheurs aux règles de présentation des travaux scientifiques, et de deux secrétaires affectées à la restructuration et la préparation des documents (stylage). Par la suite, un chef de service connaissant l'édition scientifique a été nommé pour diriger l'équipe.

Aujourd'hui l'équipe compte six acteurs à temps plein :

- le conservateur, chef de service, supervise l'ensemble des activités. Il bénéficie d'un statut d'administrateur du serveur qui lui permet de contrôler l'ensemble des tâches effectuées par tous les autres acteurs ;
- le second conservateur et le bibliothécaire ont également un statut d'administrateur; ils assurent la correction des erreurs, la conversion, l'indexation et la mise en ligne des documents traités ;
- une secrétaire à temps plein et deux secrétaires à temps partiel chargées du traitement des documents, c'est-à-dire qu'elles opèrent une série de formatages, de codages selon une feuille de style prédéfinie.

Cette équipe a, pour l'essentiel, été formée au stylage par le bibliothécaire responsable du projet. Une formation complémentaire s'est effectuée à travers l'échange et les acquis d'expérience, le site collaboratif Cybertheses, les forums de discussion et internet en fonction des besoins et des questionnements.

L'équipe est assistée par le service informatique de la bibliothèque pour les problèmes techniques liés au serveur.

Le fonds documentaire numérisé

Seuls les documents électroniques sont traités par la chaîne Cyberdocs. Les documents proviennent des instituts et des facultés rattachés à l'université Cheikh-Anta-Diop de Dakar. Ce sont essentiellement des thèses et mémoires disponibles sous format électronique et soutenus depuis 2002 quelle que soit la discipline, et dont l'auteur autorise la mise en ligne. La visibilité de ce fonds permet aujourd'hui

d'en apprécier la richesse et de montrer la qualité et la diversité des enseignements dispensés dans les facultés.

Le circuit de numérisation et la chaîne de production de documents électroniques

Lors du dépôt, les deux exemplaires imprimés accompagnant le cédérom comportant la version électronique de la thèse ou du mémoire sont transmis au service des acquisitions et du traitement. Une fois la vérification faite, le fichier électronique est alors conservé au service de numérisation qui en effectue l'archivage électronique. Les documents sont mis en ligne après l'accord écrit des auteurs. En l'absence d'autorisation, l'accès des documents est limité à l'intranet.

Le registre d'entrée de la bibliothèque universitaire fait état de statistiques annuelles comprises entre 250 et 300 thèses et mémoires publiés et disponibles en accès libre toutes disciplines confondues.

Toutes les thèses en format Word sont et seront traitées. Le traitement des thèses a commencé par les thèses soutenues en 2007. Les thèses présentées sous d'autres formats comme LaTeX ne peuvent pas encore être traitées par la chaîne Cyberdocs ou, plus précisément, l'équipe ne sait pas encore prendre en compte les styles spécifiques à LaTeX.

En 2008, la direction a pris l'initiative de procéder à une sélection plus exigeante sur la pertinence du sujet traité par rapport à l'actualité des recherches ou des domaines ciblés par l'université.

Le processus de production

Le processus de production englobe le « stylage », la conversion, l'indexation et la mise en ligne des documents.

Le « stylage » – qui consiste à identifier les parties de même nature ou de même niveau hiérarchique d'un document et à leur appliquer une mise en forme déterminée à l'aide d'un « style » – est effectué par les secrétaires, mais peut également l'être par toute autre personne du service. Les différents styles sont regroupés dans un modèle appelé « feuille de styles ». L'application de la feuille de style permet de modifier simultanément tous les éléments de même style, d'obtenir une structure homogène et d'utiliser des fonctions automatiques (création de table de matière, liste de tableaux...). La feuille de style appliquée par le service de numérisation est celle de l'université de Lyon-2³. Les fichiers stylés sont ensuite adressés aux « indexeurs » (conservateur et bibliothécaire).

La conversion est la deuxième étape dans le processus de la chaîne de production numérique. Après le « stylage », le document est converti à travers le module de la plateforme Cyberdocs. Le conservateur ou le bibliothécaire, après s'être identifié dans la rubrique « utilisateurs », transfère le fichier contenant le document, puis génère un

3. Cf. www.cybertheses.org

fichier de métadonnées, c'est-à-dire transcrit un certain nombre d'informations descriptives du document et de son contenu et d'informations techniques afin de faciliter son identification et son repérage. La phase de conversion proprement dite fournit un document XML et des documents dérivés HTML et PDF qui peuvent dès lors être indexés et mis en ligne.

Le processus se poursuit par la phase de publication : le document XML est utilisé pour la diffusion par l'outil d'indexation et de publication. Cette opération est réalisée par le module de publication sous le contrôle du bibliothécaire ou du conservateur qui, une fois identifié, peut charger (indexer), c'est-à-dire ajouter un document sur le serveur qui contient la base de données en vue de le rendre consultable sur internet via l'application SDX. Cette étape se fait directement grâce à un formulaire accessible en ligne qui permet le chargement du nouveau document sur le serveur et son indexation dans la base de données, le rendant ainsi accessible.

Le service de numérisation, qui en est à ses débuts, ne maîtrise pas encore parfaitement l'ensemble de la chaîne de publication et en particulier la lecture des rapports d'erreurs lors de l'échec de la conversion. Une certaine expérience et de la pratique sont nécessaires pour déceler l'origine, les points d'erreur ou le logiciel en cause. Par ailleurs, les enseignants-chercheurs et les étudiants sont encore assez peu familiarisés à l'utilisation des feuilles de style et les documents déposés sont le plus souvent insuffisamment structurés. Ce qui implique un travail de vérification et de reprise de style plus long. Le temps consacré au traitement des documents, qui, pour des services expérimentés (Lyon-2, Madagascar...), se situe entre 20 et 30 minutes, s'en trouve allongé.

Les documents numérisés sont accessibles librement en texte intégral sur le site de la bibliothèque universitaire⁴. La recherche sur Cyberdocs se fait par le biais d'une interface composée d'un formulaire avec différentes options de recherche : simple, avancée, par liste ou à l'intérieur d'un document. La consultation est libre et gratuite sans aucune restriction pour l'accès aux notices. L'affichage se fait en texte intégral par le biais d'un sommaire dynamique placé à gauche du contenu.

Grâce à un protocole d'échange des métadonnées, l'université Lyon-2 assure le moissonnage des données des différents serveurs du réseau, les regroupe sur le portail francophone des thèses, un site agrégateur sur lequel toutes les thèses du réseau sont accessibles en texte intégral⁵.

Les actions de communication sur le projet

Pour mieux faire connaître les objectifs du programme, les acteurs de Cyberdocs-Ucad ont procédé à une vaste sensibilisation de la communauté universitaire à travers des cycles de réunion, des affiches, la mise à disposition de

prospectus, des envois de messages aux usagers régulièrement inscrits à la BU... Un deuxième effort a consisté en l'identification, la localisation et la collecte des travaux soutenus ou en cours dans les facultés.

Enfin, une étroite collaboration avec les responsables des bibliothèques du service commun de la documentation (SCD) a permis de faciliter la collecte et le dépôt des travaux soutenus. Toutes ces formes de communication, ainsi que d'autres qui seront mises en place, renforceront les actions à mener lors de la généralisation du processus.

Bilan global du projet de numérisation

Les difficultés rencontrées

Le programme Cyberdocs-Ucad a rencontré quelques difficultés au démarrage en raison de problèmes techniques. C'est seulement le 23 mars 2004 que la première thèse a été stylée, convertie et mise en ligne. Depuis cette date, le service de numérisation est entré dans une phase de pleine activité. Toutefois, toutes les difficultés ne sont pas encore levées, en particulier dans l'interprétation des messages d'erreur lors de la conversion. Des progrès restent à faire pour maîtriser la plateforme et son environnement.

Par ailleurs, le service de numérisation se trouve confronté à un manque de formation dans le traitement des documents en format LaTeX. Ces contraintes sont alourdies par le déficit en matériel de numérisation et par une insuffisance de personnel.

Le service a enregistré un retard dans le traitement des documents en raison principalement de l'absence ou, le plus souvent, d'une insuffisance de structuration des travaux de thèses ou de mémoires déposés à la bibliothèque. Le travail a également été ralenti par le manque fréquent de connexion et les délestages électriques intempestifs provoquant l'arrêt du serveur et la disparition des données. Ces freins ont été d'autant plus pénalisants qu'ils s'ajoutaient à l'absence de deux membres du personnel qui suivaient déjà des formations professionnelles au moment de la création du service.

Les résultats positifs

La mise en application de la plateforme Cyberdocs-Ucad a permis une large visibilité sur internet des travaux universitaires réalisés à l'Ucad, avec plus de 300 documents (thèses et mémoires) en accès direct et en texte intégral depuis 2004.

Dès la mise en ligne des premiers résultats, les auteurs, soucieux d'un accès rapide aux résultats de recherche et d'une large diffusion de leurs propres travaux, ont adhéré spontanément au programme. Cette forte adhésion de la communauté scientifique universitaire au dépôt d'une version électronique de leurs travaux a constitué une source de motivation du personnel du service de numérisation qui s'est fortement investi dans ce nouvel outil de valorisation des travaux universitaires.

4. www.bu.ucad.sn, rubrique « Thèses numérisées ».

5. <http://cybertheses.francophonie.org>



Exemple de recherche sur Cyberthèses

Par ailleurs, l'opération de mise en ligne des thèses et mémoires de l'Ucad a suscité une forte volonté d'intégration des bibliothèques du service commun de la documentation et des autres universités de la sous-région (faculté de médecine de l'université de Cocody, université polytechnique de Bobo-Dioulasso, le Cames/Conseil africain et malgache pour l'enseignement supérieur) au programme Cyberthèses-Cyberdocs.

Pour renforcer la valeur et les résultats positifs de l'entreprise, le service de numérisation a bénéficié de l'adoption et de l'application d'un arrêté rectoral concernant le dépôt systématique d'une version électronique de tout travail académique avant sa soutenance à l'Ucad.

On peut noter que des travaux publiés sur Cyberdocs ont déjà été cités par d'autres chercheurs grâce à leur accès en ligne en texte intégral sur le web confortant ainsi l'intérêt d'une diffusion rapide et intégrale de résultats universitaires dans l'avancée de la recherche.

Perspectives d'évolution du projet

À moyen terme, le service de numérisation envisage d'intégrer d'autres types de documents tels que revues (annales africaines, annales de la faculté des lettres, la *Revue des institutions politiques et administratives du Sénégal/Ripas*, etc.) rapports, livres rares, images, cartes... dans la chaîne de traitement Cyberdocs-Ucad, de sorte à améliorer de façon notable l'accès à toute la documentation produite à l'Ucad.

Dans cet objectif, la numérisation rétrospective du fonds ancien (ouvrages existant en un seul exemplaire et épuisés chez les éditeurs, ainsi que thèses anciennes) reste une priorité, de même que le traitement des travaux universitaires écrits dans d'autres langues que le français, afin de jouer la carte de l'équilibre entre les disciplines enseignées à l'Ucad.

Le service de numérisation doit pouvoir s'appuyer sur la coopération et le partenariat afin d'atteindre les perspectives d'évolution qu'il s'est fixées à moyen terme. Toutefois, les délais d'exécution et de réalisation restent indéterminés car les moyens financiers et matériels ne sont pas entièrement disponibles.

Il est souhaitable que le volet « numérisation » se développe au sein des autres structures d'enseignement et de recherche pour un meilleur partage du savoir mais surtout un accès libre et gratuit à l'information pour tous les utilisateurs.

Désenclavement de la recherche en Afrique

L'activité de Cyberdocs-Ucad concourt à une plus grande visibilité des résultats de la recherche et, de ce fait, contribue au processus de désenclavement de la recherche en Afrique. La mise en place d'un archivage et d'une diffusion électronique pertinente des travaux de recherche participe à un meilleur partage du savoir, indispensable à l'essor de la recherche africaine.

Contrairement aux collections matérielles stockées dans les différentes bibliothèques, les ressources électroniques de Cyberdocs, via internet, sont directement accessibles aux chercheurs quels que soient le lieu ou le moment de leur recherche. La recherche africaine ne peut que tirer le plus grand profit du développement d'un environnement numérique auquel le réseau des bibliothèques de l'Ucad et l'ensemble des universités africaines doivent collaborer.

Pour les bibliothèques, le coût de la documentation ne cesse d'augmenter, en particulier pour les périodiques. Ainsi, plusieurs bibliothèques universitaires africaines francophones se sont vues obligées de cesser leurs abon-

nements aux périodiques du fait de la dévaluation monétaire et de l'accroissement du coût des documents. Les ressources électroniques leur offrent l'opportunité de contrecarrer ces contraintes budgétaires et de répondre aux besoins des chercheurs. Dans ce sens, les bibliothèques de l'enseignement supérieur du Sud doivent promouvoir la mise à disposition de ressources numériques gratuites et en accès libre, placées sous leur contrôle en garantissant la visibilité, la préservation et la conservation pérenne des travaux de recherche sans chercher à concurrencer l'édition commerciale.

Toutefois, les enjeux du numérique posent avec acuité la question de la formation des personnels des bibliothèques. La mise en place de bibliothèques électroniques requiert des savoir-faire spécifiques, des compétences nouvelles (culture du document structuré, pratique de nouveaux formats, application des normes documentaires et éditoriales, connaissance de modèles conceptuels de données et de leur description, gestion d'une chaîne de conversion de formats différents, etc.) qui s'apparentent à une véritable mutation de la fonction de bibliothécaire.

Mais si les bibliothécaires du Sud doivent fournir un effort pour acquérir des compétences en matière de création de collections numériques, celles-ci contribuent à la diffusion des savoirs africains et au désenclavement de la recherche au Sud. ●

Décembre 2008

Références

- Agence intergouvernementale de la francophonie, *Plateforme Cyberdocs : site de développement collaboratif*. <http://sourcesup.cru.fr/projects/cyberdocs> (consulté le 9 mai 2008)
- Frédéric Aliotti, *Étude de 2 chaînes d'édition numérique XML : projet de diffusion électronique de la production scientifique de l'INSA*, mémoire DESS Lyon-INSA, 2003. http://docinsa.insa-lyon.fr/these/doc/dess_2003.pdf
- Abdrahamane Anne, *Appropriation d'une plateforme d'édition électronique basée sur XML : Cyberdocs*, Dess Ingénierie documentaire, rapport de stage, Villeurbanne, Enssib, 2003.
- Christine Aubry, Joana Janik (dir.), *Les archives ouvertes, enjeux et pratiques : guide à l'usage des professionnels de l'information*, Paris, ADBS, 2005, 332 p. (Sciences et techniques de l'information).
- Sara Aubry, « Archives ouvertes : nouveaux modèles de publication et de diffusion et nouveaux outils », *Publier sur internet : séminaire Inria, 27 septembre-1^{er} octobre 2004, Aix-les-Bains*, Paris, ADBS, 2004 (Sciences et techniques de l'information), p. 209-237.
- Michèle Battisti, « Libre accès aux résultats de la recherche : une politique pour un renouveau de la publication scientifique ? Journées d'étude Inist », *Documentaliste-Sciences de l'information*, 2004, vol. 41, n° 4-5, p. 262-267.
- Viviane Boulétreau, Jean-Paul Ducasse, Marie-Pierre Gilliéron-Graber, « Cyberthèses en Europe », *Bulletin des bibliothèques de France*, 2001, n° 6, p. 122-129.
- Viviane Boulétreau, Jean-François Gauvin, Jean-Paul Ducasse, « La publication électronique des thèses : un exemple de collaboration franco-qubécoise », *Documentaliste-Sciences de l'information*, 1999, vol. 36, n° 6, p. 337-344.
- Thierry Chanier, *Archives ouvertes et publication scientifique : comment mettre en place l'accès libre et aux résultats de la recherche ?*, Paris, L'Harmattan, 2004, 186 p.
- Cyberthèses : publication et diffusion en ligne des thèses, portail francophone des thèses électroniques. <http://cybertheses.francophonie.org> (consulté le 9 mai 2008).
- Laurence Leroux, Marie-Laure Malingre, Alexandre Serres, *Maîtrise de l'information scientifique : la publication électronique des thèses*, École doctorale Rennes-II, 2007.
- Ministère des affaires étrangères, Système d'information scientifique et technique. www.sist-sciencesdev.net (consulté le 9 mai 2008).
- Organisation internationale de la francophonie, programme Cyberthèses. www.cybertheses.org (consulté le 9 mai 2008).
- Ana Pavani, Jose Canos, « Les raisons d'archiver les thèses sous forme numérique et des stratégies », *Guide sur la production et la diffusion des thèses sous forme numérique*, Paris, Unesco, 2002. www.etdguide.org/french/Chapitre2.doc (consulté le 9 mai 2008).
- Université Cheikh-Anta-Diop de Dakar, réseau des bibliothèques. www.bu.ucad.sn/cyberpac (consulté le 9 mai 2008).
- Université Lyon-2, *Didacticiel : produire et diffuser des documents structurés*. <http://theses.univ-lyon2.fr/didacticiel/> (consulté le 9 mai 2008).