
UN PANORAMA DE LA RECHERCHE UNIVERSITAIRE BELGE: le répertoire BICTEL/e des thèses électroniques et e-prints

Marjorie GOBIN
Université Catholique de Louvain (UCL)

Cadre du projet

Lancé au début de l'année 2002, le projet BICTEL/e, projet interuniversitaire belge, financé par la Bibliothèque Interuniversitaire de la Communauté française de Belgique asbl (BICfB), vise la constitution d'un répertoire électronique, commun aux neuf universités de la Communauté française de Belgique. Ce répertoire, que l'on appellerait en anglais une " *archive* ", a comme objectif de regrouper des thèses électroniques et " e-prints " (publications scientifiques électroniques), en texte intégral, des différentes institutions. Cette archive, qui s'inscrit dans un vaste mouvement visant la libéralisation de la communication scientifique, est librement accessible via une interface web : www.bicfb.be/bictel. Elle permet d'une part l'encodage des données par l'auteur, dans un répertoire local, propre à son institution, et d'autre part la recherche par l'utilisateur dans l'ensemble ou dans un seul de ces répertoires (fonction " parcourir ", recherche par mot-clés, recherche en texte intégral). Ces fonctionnalités répondent à l'objectif de diffusion et de communication des résultats de recherche parmi les communautés scientifiques.

Au niveau local, chacune des neuf institutions alimente son archive, d'un côté par des thèses défendues en son sein, d'un autre côté par des publications électroniques de ses chercheurs (articles, prépublications, documents de travail, interventions à des conférences, supports de cours...). Chaque institution scientifique est en effet responsable de la collecte, de l'organisation, de la validation et de la maintenance de sa production scientifique, selon le principe de l'auto-archivage. Le projet BICTEL/e offre aux neuf institutions l'infrastructure technique, toutefois la mise en œuvre pratique est prise en charge par chaque université, en fonction de ses particularités.

Les répertoires locaux, respectant les normes OAI, Open Archive Initiative, peuvent être dès lors " moissonnés ", (harvested). Leurs données

sont récoltées et rassemblées à un niveau supérieur pour constituer le répertoire commun à ces neuf institutions.¹ Dans le jargon de l'OAI, BICTEL/e est donc en même temps un fournisseur de données et un fournisseur de services. Pour les thèses, l'étape supérieure est l'enregistrement du répertoire des thèses BICTEL/e au réseau international NDLTD² (*Networked Digital Library of Theses and Dissertations*), regroupant les thèses électroniques de près de deux cents institutions et réseaux universitaires de par le monde.³ Pour le moment, il n'existe pas d'équivalent stricto sensu pour les e-prints. Toutefois, cette possibilité d'intégration à un ou plusieurs autres répertoires e-prints est envisageable par la conformité aux normes OAI. Les répertoires peuvent en effet être " moissonnés " par d'autres fournisseurs de services, par exemple au niveau d'archives thématiques.

Des travaux préliminaires ayant été réalisés par l'équipe informatique des Bibliothèques de l'Université Catholique de Louvain, l'UCL a été chargée de la mise en œuvre du projet, sous la direction de Charles-Henri NYNS, bibliothécaire en chef. Du point de vue des moyens humains, une graduée en informatique, Florence DARDENNE, et une documentaliste universitaire, Marjorie GOBIN, ont travaillé sur ce projet. De plus, un comité de pilotage dans lequel sont présents des représentants de la plupart des institutions universitaires membres de la BICfB suit le bon déroulement du projet.

A. Volet thèses électroniques

Depuis plusieurs années déjà, tant aux Etats-Unis, qu'en Asie ou en Europe, se mettent en place des répertoires de thèses électroniques.

Leur particularité est de proposer le texte intégral des thèses, en libre accès. Comme pour les auteurs d'e-prints, les bénéficiaires pour les nouveaux docteurs qui déposent leur thèse en ligne sont notamment la rapidité de diffusion de travaux de

recherche, une plus grande visibilité par l'accroissement considérable du public potentiel, l'enrichissement possible du document par des éléments multimédia (son, vidéo, image 3D, photos...). L'exemple sans doute le plus connu de ce type de répertoire est celui de l'Université technique de Virginie (Virginia Tech), intégré au répertoire international NDLTD.⁴

1. Logiciels retenus

Quelques indications d'ordre techniques : le logiciel de base qui est employé pour le répertoire des thèses est *ETD-db*⁵. Celui-ci est développé par Virginia Tech (USA) et est déjà largement utilisé par les membres du réseau NDLTD. L'emploi de ce logiciel *open source*, permet le référencement des thèses de nos universités dans le catalogue collectif mondial NDLTD.

Le choix du logiciel de *harvesting* s'est porté sur trois modules *ODL* (Open Digital Library) développés par un membre de la même institution.⁶ Ce choix a été fait en fonction des technologies utilisées (similaires à *ETD-db* et à *EPrints.org* qui, comme son nom l'indique, sera employé pour le répertoire des e-prints), du libre accès aux sources et surtout de la simplicité de l'outil.

Le logiciel d'indexation retenu est *Inktomi Enterprise Search*. Ce logiciel, payant, dispose de fonctionnalités que l'on ne retrouve pas dans les logiciels gratuits. Il offre un outil puissant d'indexation et de recherche en texte intégral ainsi qu'une interface web pour son administration.

2. Premières mises en œuvre techniques et élaboration d'un prototype

L'élaboration du premier prototype de site web pour les thèses électroniques a suivi les premières installations techniques et la configuration du serveur. L'idée était de développer au sein de l'institution pilote un prototype, de le proposer aux avis et commentaires des autres institutions, puis de l'installer chez elles, une fois les dernières modifications effectuées. La version définitive du site BICTEL/e fut approuvée lors de l'assemblée générale de la BICfB en février 2003.

Ce prototype de site devait tenir compte de la structure du projet : à un premier niveau, se trou-

vent les archives institutionnelles propres à chacune des neuf universités (accès aux thèses et/ou aux e-prints de chaque institution); à un niveau supérieur, le répertoire BICTEL/e regroupant la production scientifique des neuf universités de la Communauté française (thèses et/ou e-prints de l'ensemble des institutions).

3. Tests en conditions réelles - expérience de l'institution pilote, l'UCL

Parallèlement au développement technique, dès le lancement du projet BICTEL/e, un groupe de travail a été mis en place à l'UCL, sous la direction de Freddy DONCKELS (service du pro-recteur), afin d'encadrer la mise en œuvre de ce projet dans l'institution. Les réflexions ont porté notamment sur l'intégration de la procédure du dépôt électronique du texte intégral au processus en vigueur en matière de défense de thèse, sur les étapes du processus de soumission, sur les personnes relais à désigner, sur la procédure de validation à mettre en place, sur la question essentielle des droits d'auteur.

Des tests de soumission de thèses ont été réalisés en conditions réelles dès l'automne 2002. Ceux-ci ont bien entendu apporté des indications précieuses sur les éventuelles difficultés rencontrées par les utilisateurs et gestionnaires du site. L'expérience de l'institution pilote a permis de rédiger des documents de travail, à destination des autres institutions, reprenant des pistes de réflexion, et notamment les différentes étapes à mettre en place pour lancer le projet au sein de leur institution. Toutes ces pistes ont enrichi l'aide en ligne proposée sur le site.⁷

La mise en place d'un tel projet dans une institution implique en effet un grand effort de prise de contact avec les entités concernées : Administration de la recherche; représentants du corps scientifique; commissions de gestion des troisièmes cycles et représentants des doctorants; représentants des secrétariats des facultés; responsables des bibliothèques; service juridique...

Depuis le 15 mars 2003, la procédure de dépôt de thèses électroniques est opérationnelle à l'UCL. Elle rencontre déjà un beau succès, avec environ un tiers des thèses soutenues déposées en texte intégral dans le répertoire.⁸

4. Processus de soumission à l'UCL

Le dépôt du texte intégral dans le répertoire BICTEL/e n'a pas été rendu obligatoire par les autorités académiques, mais est vivement encouragé. Pour rappel, à l'UCL, seules les thèses de doctorat et d'agrégation peuvent être intégrées au répertoire BICTEL/e. D'autre part, il a été décidé que le dépôt de la version électronique d'une thèse devait être effectué obligatoirement avant la défense de la thèse. Ceci afin de s'assurer que la version présente dans le répertoire BICTEL/e est bien celle qui a été défendue. Le projet est en effet la constitution d'un répertoire institutionnel de thèses universitaires, dans un but de communication scientifique à des fins non commerciales, et ne remplit pas les fonctions caractéristiques d'un éditeur. A terme, la thèse électronique devrait en outre remplacer la thèse papier.

Pratiquement, voici les principales étapes du processus de soumission d'une thèse électronique à l'UCL. Quelques semaines avant la défense de thèse, le doctorant reçoit, par le gestionnaire de son doctorat,⁹ une information sur le projet des thèses électroniques. S'il souhaite participer à ce projet, il est invité à signer un *contrat de licence de diffusion d'une thèse électronique*. Ce document comprend plusieurs aspects : tout d'abord le doctorant donne l'autorisation à l'université de diffuser la thèse électronique via son répertoire institutionnel; il s'engage sur le fait qu'il détient bien les droits d'auteur sur son travail; enfin, l'auteur y précise le type de diffusion souhaité. Le principe de base du projet est la diffusion de la totalité de la thèse, en texte intégral, via internet. Toutefois, pour des raisons de droits d'auteur ou de confidentialité, des limitations à la diffusion peuvent être demandées : accès limité à l'intranet de l'institution, accès interdit à certains chapitres, diffusion différée dans le temps.

Afin d'éviter toute demande saugrenue, l'entité gestionnaire est chargée elle-même de faire signer au doctorant ce contrat de diffusion; de le renvoyer à l'Administration de la recherche qui se charge de le vérifier; de demander un accès au système pour le doctorant au Service central des bibliothèques et enfin, après la défense de thèse, de communiquer la validation de la thèse à ce dernier. Le Service central des bibliothèques crée le login (nom d'utilisateur et mot de passe) et l'envoie au doctorant. Une fois muni de ce login, le doctorant peut introduire ses données. Il lui suffit de remplir les différents champs relatifs aux données bibliographiques de la thèse (nom, prénom, adresse, département, grade du doctorat,

titre de la thèse, mots-clés, résumé, composition du jury...), puis de joindre le ou les fichiers PDF de la thèse. Aucune exigence n'est imposée quant à la forme du document.

Des fichiers multimédias (images, sons, vidéos) peuvent venir enrichir la thèse. Les formats suivants sont acceptés :

- Pour les images : JPEG (.jpg), CompuServe GIF (.gif), CGM Computer Graphics Metafile (.cgm)
- Pour les vidéos : MPEG (i.e., MPEG-1, MPEG-2) (.mpg), Audio Video Interleaved - Microsoft (.avi)
- Pour les fichiers audio : MP3, wav, midi
- Divers : Autocad, Arcview

La fiche de la thèse ainsi encodée demeure dans une zone tampon, jusqu'à l'annonce de la délibération par l'entité gestionnaire du doctorat. L'accès du doctorant est bloqué et la thèse apparaît dans le système, suivant les modalités prévues.

B. Volet E-prints

Dans le répertoire BICTEL/e, à côté des thèses électroniques, nous trouvons également les e-prints. Sous cette appellation générique d'e-prints, se cachent dix catégories de documents scientifiques électroniques, dont le texte intégral est librement consultable : livre; chapitre de livre; actes de colloque; contribution à un colloque; poster de colloque; document de travail; article de revue scientifique; article de presse; prépublication; support pédagogique.

Depuis quelques années, dans l'esprit de l'initiative de Budapest et la déclaration de Berlin¹⁰, se multiplient les exemples de répertoires de publications scientifiques électroniques.¹¹ Les physiciens de Los Alamos, spécialement Paul GINSPARG, ont initié le mouvement en 1991 avec la première archive ouverte de prépublications d'articles.¹² Depuis lors, des dizaines de répertoires de ce genre ont vu le jour.¹³

Il est important ici d'insister sur le caractère institutionnel du répertoire BICTEL/e. Il s'agit bien d'illustrer les résultats des recherches de

l'ensemble des disciplines pratiquées dans les institutions universitaires de la Communauté française de Belgique. En cela, BICTEL/e se différencie de répertoires thématiques, tels que Cogprint¹⁴ consacrés aux neuro-sciences, ou le répertoire du département d'électronique et d'informatique de l'Université de Southampton¹⁵, ou encore @rchiveSIC¹⁶, dédié à la Science de l'Information et de la Communication.

1. Le logiciel E-prints.org

Le logiciel *EPrints.org*¹⁷ est utilisé pour ce second volet du projet. Il s'agit également d'un logiciel *open source*, développé quant à lui par l'Université de Southampton (Angleterre), et lui aussi déjà employé par bon nombre d'institutions ou organismes scientifiques pour leurs répertoires de publications électroniques.¹⁸ La langue principale du répertoire des e-prints est le français, toutefois, une version anglaise, langue vernaculaire de la communauté scientifique, est également proposée.

2. Elaboration d'un prototype

La démarche pour le volet e-prints est similaire à celle qui a été présentée pour les thèses électroniques : préparation, élaboration et mise en place au sein de l'institution pilote d'un prototype testé au sein de celle-ci puis soumis à l'avis des autres participants, enfin rectifications éventuelles avant installation pour l'ensemble des universités. Par rapport au volet des thèses électroniques, la procédure de dépôt d'e-prints dans le répertoire BICTEL/e est fortement simplifiée.

La simplicité de l'outil devait répondre aux besoins des chercheurs désireux de déposer une publication dans le répertoire BICTEL/e. Le choix pertinent des types de documents pouvant être soumis, une liste des sujets adéquate, un processus de soumission clair, une procédure de validation efficace, sans être trop lourde, furent les préoccupations majeures de ce second volet.

L'aspect " validation " a fait l'objet de longues réflexions au sein du comité de pilotage du projet et du groupe de travail de l'UCL. Il en est ressorti que, dans la mesure du possible, il est préférable d'utiliser les circuits existants.

Ainsi, pour l'image des institutions, il est hors de question que " n'importe qui " puisse alimenter le répertoire des e-prints par " tout et n'importe quoi ". Seuls bien entendu les membres de la communauté scientifique de l'institution (professeurs, doctorants, chercheurs) ont le droit de déposer des documents dans le répertoire de son institution. Habituellement, dans les centres de recherche, une personne est déjà chargée de la gestion des publications. Ainsi, pour ne pas multiplier les intervenants, il suffit de lui créer un accès au répertoire.

Se pose plus généralement la question de la confiance que l'on peut accorder aux publications contenues dans ce genre de répertoire. En aucun cas, le projet BICTEL/e ne joue le rôle d'éditeur dans un sens classique. Il n'a par exemple pas de comité de lecture. Il est absolument nécessaire de donner aux visiteurs du répertoire une information claire à ce propos. Ainsi, la première étape d'un dépôt est de choisir parmi les dix rubriques proposées¹⁹ celle qui correspond au document puis de préciser si celui-ci a fait ou non l'objet d'un peer-review par ailleurs. En aucun cas, le lecteur ne doit être laissé dans l'ignorance du niveau de fiabilité de la publication.

Si les résultats de recherche peuvent se traduire par des publications telles que des livres, des articles dans des revues scientifiques ou dans la presse, en passant par des interventions à des colloques ou à des conférences (actes, posters), le type de document le plus déposé est sans conteste, et assez logiquement, le " working paper ", support de travail et d'échange avec les autres spécialistes du domaine, préambule à de futurs articles scientifiques.²⁰

Enfin, l'une des tâches essentielles a été la constitution d'une liste de sujets, limitée pour des raisons de clarté à moins de deux cents entrées, mais illustrant bien les thèmes de recherche pratiqués en Belgique francophone. Cette liste, inspirée de la *Library of Congress Classification* et de sa traduction par l'Université Laval, a dû subir quelques modifications, pour répondre aux besoins des neuf universités impliquées dans le projet.

3. Premières mises en oeuvre à l'UCL

Après des tests du prototype entrepris au sein de l'institution pilote, d'abord de manière artificielle, puis en conditions réelles auprès de centres de recherche déjà habitués à déposer leurs travaux sur Internet, une version de production du répertoire fut proposée à la rentrée 2003.

Les réactions sont positives : le répertoire contient déjà plusieurs dizaines d'e-prints. Les chercheurs ont en effet bien saisi l'intérêt de faire figurer leurs publications dans ce type de répertoire institutionnel, répondant aux normes OAI, et cela même s'ils déposent dans d'autres répertoires sectoriels. Le multipostage n'est pas un problème, au contraire la visibilité des travaux de recherche n'en est qu'accrue.

C. La question des droits d'auteur

Si les chercheurs sont généralement enthousiastes à l'idée de participer au projet BICTEL/e, ils restent néanmoins prudents face au problème des droits d'auteur.

Le service juridique de l'Administration de la Recherche reçoit de nombreuses questions à ce propos, spécialement en ce qui concerne les thèses électroniques. Les doctorants ont notamment peur que la mise en ligne de leur thèse constitue un obstacle pour une publication ultérieure.

Juridiquement, il n'y a aucune incompatibilité entre les deux, dans la mesure où le doctorant ne donne à l'UCL qu'une licence non exclusive, ce qui lui laisse les mains libres pour publier ailleurs sous quelle que forme que ce soit. Les doctorants qui présentent des thèses à articles sont confrontés à un autre problème. Dans ce cas, ils ont souvent déjà cédé leurs droits sur cette " œuvre " et doivent demander l'autorisation à l'éditeur de la déposer dans le répertoire BICTEL/e, ce dernier ayant comme fin la diffusion scientifique, sans but commercial. La question de la possibilité du plagiat peut être également soulevée. Il s'est avéré qu'il est nettement plus risqué de plagier une thèse électronique, répertoriée de manière précise et diffusée à grande échelle, qu'une thèse dont seuls quelques exemplaires papier ont été tirés.²¹

Une Foire Aux Questions en ligne²² a été mise sur pied pour donner une première réponse à toutes ces questions. Si besoin, une juriste apporte son soutien aux chercheurs dans leurs démarches auprès des éditeurs chez qui ils auraient déjà publié. A ce propos, le projet anglais RoMEO²³ synthétise les positions des éditeurs au sujet de ce type de répertoire. De manière générale, les demandes individuelles, motivées, sont acceptées par les éditeurs. Mais jusqu'à présent, aucun accord global n'a pu être conclu.

¹ L'Initiative pour une Archive Ouverte (OAI) a défini des normes d'échange de métadonnées (date, auteur, titre, nom du journal, etc.), rendant compatibles entre eux les documents respectant ces balises. Voir par exemple : <http://www.dlib.vt.edu/projects/OAI/> ; <http://www.inist.fr/oa/spip/> , <http://www.ukoln.ac.uk/distributed-systems/oa/collection-description/whitepaper.html> ; G. CHARTON, *Les archives ouvertes dans la communication scientifique*, janvier 2003 : enjeux, OAI, typologie, exemples, ressources <http://www.urfist.jussieu.fr/urfist/archives-ouvertes.htm> ; Stevan HARNAD, *Lecture et écriture scientifique " dans le ciel " : Une anomalie post - gutenbergiene et comment la résoudre* : problématique générale http://www.text-e.org/conf/index.cfm?fa=printable&ConfText_ID=7

² <http://www.ndltd.org/>

³ <http://rocky.dlib.vt.edu/~etdunion/cgi-bin/OCLCUnion/UI/browse.pl>

⁴ <http://scholar.lib.vt.edu/theses/> ; <http://www.ndltd.org/>

⁵ Pour des informations techniques, voir <http://scholar.lib.vt.edu/ETD-db/developer/>

⁶ Ces trois modules ODL sont : DB-Union (" harvester "), DB-Browse (outil pour parcourir les éléments " récoltés ") et BrowseUI (interface web pour DB-Browse). Pour plus d'information, voir <http://oai.dlib.vt.edu/odl/>

-
- ⁷ <http://edoc.bib.ucl.ac.be:81/ETD-db/help/>
- ⁸ Sur un total de 95 thèses défendues entre le 15 mars et le 15 septembre 2003, 31 ont été déposées en texte intégral dans le répertoire BICTEL/e.
- ⁹ " entité gestionnaire du doctorat " : terme générique pour désigner la ou les personnes du secrétariat facultaire ou départemental, ou encore de l'école doctorale, qui gère l'organisation des défenses de thèses pour les doctorants de l'entité concernée et qui assure le suivi des démarches administratives (ex : annonce de la défense, transmission du diplôme au recteur pour signature).
- ¹⁰ http://www.inist.fr/oa/spip/article.php3?id_article=38
- ¹¹ CARL, Institutional Repositories Pilot Project, portail de ressources en lignes :
<http://www.carl-abrc.ca/projects/ir/> Liste des répertoires compatibles OAI :
<http://www.openarchives.org/Register/BrowseSites.pl>
- ¹² ArXiv X/ P. GINSPARG: <http://xxx.lanl.gov>, <http://fr.arXiv.org>
- ¹³ http://www.inist.fr/oa/spip/rubrique.php3?id_rubrique=7
- ¹⁴ <http://cogprints.ecs.soton.ac.uk/>
- ¹⁵ <http://eprints.ecs.soton.ac.uk/>
- ¹⁶ <http://archivesic.ccsd.cnrs.fr/>
- ¹⁷ <http://www.eprints.org/>
- ¹⁸ Pour la liste des utilisateurs d'*Eprints.org*, voir <http://software.eprints.org/> Exemples de répertoires : CogPrints, de l'Université de Southampton ; Papyrus de l'Université de Montréal; ANU, Australian National University; LU:research , Université de Lund; Università degli studi di Firenze ...
- ¹⁹ Livre, book ; chapitre de livre, book chapter ; actes de colloque, conference proceedings ; contribution à un colloque, conference paper; poster de colloque, conference poster; document de travail, working paper; article de revue scientifique, journal article; article de presse, newspaper/magazine article; pré-publication; preprint; support pédagogique, educational material. Les champs proposés par le formulaire d'encodage varient quelque peu suivant la nature des documents.
- ²⁰ A l'UCL, plusieurs centres de recherche avaient déjà l'habitude de déposer leurs documents de travail en PDF sur leur site.
- ²¹ Voir à ce propos : JISC Plagiarism Advisory Service
http://online.northumbria.ac.uk/faculties/art/information_studies/Imri/Jiscpas/site/jiscpas.asp
- ²² http://edoc.bib.ucl.ac.be:81/ETD-db/help/prep_faq.html
- ²³ <http://www.lboro.ac.uk/departments/ls/disresearch/romeo/index.html> <http://www.sherpa.ac.uk/index.html>

* * *