

Élèves bibliothécaires

Mise en place d'un circuit de traitement et de diffusion de thèses électroniques à l'Université de Cergy-Pontoise

Dossier d'aide à la décision

VINCENT Stéphanie

Sous la direction de Dominique MANIEZ
Professeur associé à l'enssib

Correspondant d'établissement
REYNAUD Natalie
Responsable de la Mission Formation et Thèses
Bibliothèque de l'Université de Cergy-Pontoise

Remerciements

Mes sincères remerciements s'adressent à toutes les personnes qui m'ont aidée dans le déroulement de cette année de formation :

Monsieur Jean-Paul DUCASSE, qui m'a accueillie au service ERAD de Lyon 2 et m'a apporté ses précieux éclairages;
Kim, Céline et Jérôme, qui m'ont révélé certains secrets du stylage et de l'informatique, à la faveur d'un stage passionnant;

Les personnes qui m'ont accueillie et soutenue dans mon établissement, et qui, sans encore bien me connaître, ont fait le pari de me confier cette mission ;

Monsieur Dominique Maniez, pour avoir su se glisser dans le costume de tuteur au dernier moment, avec une souplesse et une bienveillance particulièrement louables ;

La promotion 'FIBE N', baptisée du nom mérité de « Nsibien », pour m'avoir confirmé que l'on peut atteindre la qualité dans le travail et dans les relations humaines, simultanément ;

A François-Xavier, pour qu'il ait définitivement honte de moi...

Et à tous ceux qui m'ont soutenue, de quelque manière que ce soit.

REMARQUE LIMINAIRE

En accord avec l'ensemble de la promotion de la FIBE 2005-2006, je tiens à préciser que ce PPP a été rédigé dans un contexte particulièrement défavorable. Nous rappelons que les formateurs en charge de la formation initiale des bibliothécaires ont, pour la plupart, quitté leurs fonctions en cours d'année.

Ce flottement très préjudiciable n'a reçu aucune réaction de la direction de l'ENSSIB. La moitié des stagiaires ont ainsi rédigé leur PPP sans tuteur jusqu'à la nomination d'un représentant au jury peu de temps avant la remise des mémoires, provoquant une iniquité de fait entre les stagiaires.

Cette remarque a été insérée sur tous les PPP de la promotion.

Résumé :

Le mouvement pour les archives ouvertes s'affirme désormais au sein de la communauté scientifique, et le libre accès aux thèses est parallèlement devenu un enjeu majeur des universités, écoles d'ingénieurs et autres établissements d'enseignement supérieur et de recherche. Dans ce contexte, l'université de Cergy-Pontoise envisage la possibilité de diffuser ses thèses sous forme électronique, en vue de rendre plus visibles et accessibles ces documents scientifiques de valeur, souvent non publiés. Ce rapport étudie les projets en cours et les procédés techniques existant pour le dépôt, le signalement, la diffusion et l'archivage de ce type de documents, afin de soumettre une proposition applicable au sein de la Bibliothèque universitaire de Cergy-Pontoise.

Descripteurs :

Thèses et écrits académiques – Numérisation
Université de Cergy-Pontoise

Abstract :

The concept of open archives is from now on well known by the scientific community, while open access to theses has become a major preoccupation for institutes of higher education and research. Disseminating electronic theses is a key concern for the University of Cergy-Pontoise in providing visibility for and access to its scientific output that although mostly not published is from great value. The present work examines a review of the ongoing projects and techniques used to deposit, process (housing, classifying), disseminate and archive theses, in order to propose an achievable design within the university library of Cergy-Pontoise.

Keywords :

Dissertations -- Digitization -- France

Scholarly Web sites -- Directories

Toute reproduction sans accord exprès de l'auteur à des fins autres que strictement personnelles est prohibée.

Sommaire

INTRODUCTION.....	8
-------------------	---

PARTIE 1 : ENJEUX D'UNE MISE EN PLACE D'UN CIRCUIT DE TRAITEMENT ET DE DIFFUSION DE THÈSES ÉLECTRONIQUES9

1. LES ENJEUX DE LA DIFFUSION ÉLECTRONIQUE DES THÈSES.....	9
1.1. Pourquoi mettre en ligne les thèses ?.....	9
1.2. Un potentiel documentaire souvent sous-exploité.....	10
1.3. L'université de soutenance : l'opérateur ad hoc d'une mise en ligne .	11
1.3.1. La notion d'archive institutionnelle et de dépôt institutionnel.....	12
1.3.2. La notion d'archive ouverte.....	14
1.3.3. Dépôt compatible avec OAI.....	14
2. LES RECOMMANDATIONS MINISTÉRIELLES	15
3. MISE EN PLACE D'UN CIRCUIT DE DIFFUSION DES THÈSES ET QUESTIONS	

JURIDIQUES..... 18

3.1. Les aspects juridiques afférents aux thèses électroniques	18
3.2. Droit de diffusion versus droit d'auteur et commercialisation.....	19
3.2.1. Formulaire/contrat type	20
3.2.2. La question du droit de citation : les obligations du doctorant	20
3.3. Exemples de solutions retenues par des établissements français	21
3.3.1. Pas formulaire interne à l'établissement	21
3.3.2. Formulaire ne mentionnant que l'autorisation de diffuser	21
3.3.3. Contrat de diffusion détaillé	22
3.3.4. Attestation de conformité de la version électronique avec la version papier	22
3.3.5. Contrat exclusif de diffusion	23
3.4. Bilan des exemples retenus	24

PARTIE 2 : L'UNIVERSITÉ DE CERGY-PONTOISE : UN CONTEXTE PARTICULIER POUR CES ENJEUX.....25

1. PRÉSENTATION DE L'UNIVERSITÉ ET DE SES COMPOSANTES	25
1.1. La recherche à l'UCP.....	25

1.2.	<i>Les écoles doctorales</i>	26
1.3.	<i>La Bibliothèque universitaire</i>	28
1.3.1.	Présentation de la Bibliothèque universitaire	28
1.3.2.	Un défi pour la Bibliothèque universitaire	31
1.3.2.1.	Le choix de STAR.....	31
1.3.2.2.	Principes et caractéristiques de STAR.....	32
1.3.2.3.	Avantages et contraintes induits par l'adoption de STAR.....	33
2.	ANALYSE DE L'EXISTANT À L'UNIVERSITÉ DE CERGY-PONTOISE :	36
2.1.	<i>Le circuit actuel : un circuit papier</i>	36
2.1.1.	Le dispositif réglementaire national.....	36
2.1.2.	Le circuit actuel au niveau du SCD	37
2.2.	<i>Volumétrie des thèses à traiter</i>	38
2.2.1.	Production annuelle de thèses et question de la rétroconversion .	38
2.2.2.	L'autorisation de diffusion en ligne	40
3.	LE CHOIX DU FORMAT	41
3.1.	<i>Le format PDF</i>	42
3.2.	<i>Le langage XML</i>	43
3.3.	<i>Préconisations et contraintes induites</i>	44

PARTIE 3 : PROJETS EN COURS DANS LES ÉTABLISSEMENTS DE SOUTENANCE ET SCÉNARIIS ENVISAGEABLES À CERGY-PONTOISE 47

1.	PRÉSENTATION DES PROJETS EXTÉRIEURS EN COURS	47
1.1.	<i>Contexte international : projets des universités américaines (Virginia Tech) et québécoises (Laval)</i>	47
1.1.1.	Le Networked Digital Library of Theses and Dissertations (NDLTD).....	47
1.1.2.	La collection de l'Université de Laval	47
1.1.3.	Le dépôt institutionnel Papyrus de l'université de Montréal	48
1.2.	<i>En France : les projets à l'échelle locale et nationale</i>	49
1.2.1.	Lyon2-Cyberdocs	49
1.2.1.1.	Origines et principes du programme.....	49
1.2.1.2.	Fonctionnement de Cyberdocs	49
1.2.2.	CCSD : TEL Thèses-en-lignes.....	51

1.2.3. Thèses de l'Université Louis Pasteur : archive ouverte de l'ULP (E-print)	53
2. AXES DE RÉFLEXION, TÂCHES COMMUNES AUX DIFFÉRENTS SCÉNARI D'UN CIRCUIT ÉLECTRONIQUE DES THÈSES	55
2.1. <i>Réflexions préalables aux différents scénarii : les enjeux communs</i> ...	55
2.2. <i>Les étapes et éléments communs aux scénarii</i>	56
3. TROIS SCÉNARI POSSIBLES POUR CERGY-PONTOISE	58
3.1. <i>Implanter Cyberthèses : un projet intégré fonctionnant sur une logique de proximité</i>	58
3.1.1. Organisation.....	59
3.1.2. Coûts prévisibles	60
3.2. <i>Vers un réseau national de thèses électroniques : participation de l'UCP à TEL</i>	60
3.2.1. Organisation.....	61
3.2.2. Coûts prévisibles	62
3.3. <i>Vers une bibliothèque numérique : implantation d'une plate-forme ouverte</i>	62
3.3.1. Organisation.....	63
3.3.2. Coûts prévisibles	63
3.4. <i>Bilan des trois scénarii</i>	64
CONCLUSION	66
BIBLIOGRAPHIE	67
TABLE DES ANNEXES	78

Introduction

L'évolution des technologies et les propriétés d'Internet permettent aujourd'hui l'élargissement de la diffusion scientifique et la création d'un véritable réseau universitaire mondial. Cette possibilité concerne au premier chef les thèses, en tant que document scientifique validé : désormais produites dans leur quasi-totalité « nativement » sous format numérique¹, elles demeurent cependant non publiées dans la majorité des cas.

Or, la diffusion électronique ouvre incontestablement un champ de consultation beaucoup plus important que le support papier, notamment grâce aux protocoles d'échanges de données, qui rendent les archives électroniques visibles du monde entier.

Dès l'an 2000, le ministère de l'Education nationale avait pris la mesure de ces enjeux et du faible périmètre de la diffusion des thèses sous forme électronique à l'instigation des établissements. Cette volonté de valorisation est plus que jamais d'actualité : à la circulaire du 21 septembre 2000, incitant à généraliser la diffusion électronique des thèses, a succédé une nouvelle, datée de mars 2005, dans l'attente de l'arrêté du 7 août 2006 qui officialise une refonte du circuit national des thèses.

Plusieurs projets, notamment dans la mouvance des archives ouvertes institutionnelles, des universités, écoles d'ingénieurs et instituts de recherche, ont déjà entrepris la valorisation des thèses soutenues ou préparées en leur sein, en facilitant leur repérage et leur accessibilité. A chacun de ces projets correspondent des choix politiques, organisationnels et techniques, qui tiennent eux-mêmes compte de compétences disponibles et de ressources variées.

L'objet de ce travail est de mettre en exergue les différents choix possibles, et les implications de chacun d'eux, du point de vue juridique, technique et organisationnel, en vue de l'adoption d'un processus aussi adapté que possible aux réalités propres à l'Université de Cergy-Pontoise.

¹ Le terme « thèse électronique » renvoie à l'utilisation d'un support électronique (disquette, fichier, CD-ROM, site WEB, etc...) pour la production de la thèse, par opposition à la forme traditionnelle sur support papier, que les universités ont jusqu'à maintenant exigée de leurs étudiants.

Partie 1 : Enjeux d'une mise en place d'un circuit de traitement et de diffusion de thèses électroniques

1. Les enjeux de la diffusion électronique des thèses

1.1. Pourquoi mettre en ligne les thèses ?

Le préambule de la circulaire 05-094 du 29 mars 2005 concernant le dépôt, signalement, diffusion et archivage des thèses sous forme électronique, rappelle le caractère stratégique des thèses en matière de communication scientifique :

« Les thèses constituent des documents de première importance qui doivent être valorisés en tant que vecteurs essentiels de la communication scientifique, expression de l'activité de recherche dans les universités (et autres établissements habilités) et références pour la carrière des jeunes docteurs. La production des thèses aujourd'hui sous une forme nativement numérique permet d'envisager cette valorisation de façon renouvelée et plus ambitieuse. »

Le service ERAD de l'Université de Lyon ² relève plusieurs avantages de la thèse électronique, sur le plan de la présentation, de la formation des doctorants à de nouvelles compétences, et sur le plan de la diffusion, en soulignant que :

1. le **doctorant** est replacé dans sa position de **producteur et de diffuseur d'information scientifique** ;
2. le **repérage** et **l'accès rapide** au document proprement dit permettront d'une part au chercheur de gagner en **efficacité**, et d'autre part au monde de la recherche de s'ouvrir à un **public plus vaste** (non chercheur ou étranger) pas forcément familier du mode de signalement et de consultation des thèses en France ;

² Extraits de « Pourquoi diffuser sa thèse sur internet ? La diffusion sur Internet ou la ré-appropriation du travail de recherche par le chercheur ». Consultable sur le lien : http://theses.univ-lyon2.fr/pourquoi_diffuser.php [dernière consultation le 10 juillet 2006]

3. en assurant **un signalement et une diffusion quasi instantanés**, une **mise à jour immédiate des informations**, en facilitant la consultation gratuite du contenu des documents, la mise en ligne des thèses contribuera efficacement au **rayonnement de l'étudiant et de l'université**.

Outre ces bénéfices du dépôt électronique et de la diffusion en ligne, l'université de Laval met en exergue un autre avantage appréciable de cet avatar, par rapport à la forme papier, à savoir « assurer la pérennité du document (qui ne peut être détruit, endommagé ou volé) »³.

1.2. Un potentiel documentaire souvent sous-exploité

Jusqu'à récemment, seules quelques bases de données ou catalogues en ligne assuraient le signalement de certaines thèses en France⁴.

Ces catalogues permettent d'identifier une thèse par ses éléments administratifs et de localiser l'établissement de soutenance, mais le téléchargement de la version électronique du document est rarement proposé : l'utilisateur en est ainsi encore réduit, dans la majorité des cas, à effectuer une demande de reproduction auprès de l'établissement détenteur, qui lui délivrera une version papier de la thèse dans un délai plus ou moins long, et souvent moyennant une contrepartie financière.

Ce fonctionnement est contraignant pour les usagers, et pour les chercheurs en particulier : il est non seulement consommateur de temps, mais encore génère des aberrations (explosion du PEB par exemple, risque d'insatisfaction quant à la teneur réelle de la thèse, etc). Dans ce contexte, un simple accès en ligne au sommaire détaillé de la thèse (à défaut de son contenu) peut déjà présenter un énorme avantage (gains en terme de temps et de coût pour le lecteur).

La consultation en ligne présente aussi d'indéniables avantages en terme d'ergonomie de lecture et de modalités de recherches.

La diffusion des thèses électroniques vise donc avant tout à **améliorer la communication**. Les statistiques de consultation fournies par différentes expériences de diffusion menées à ce jour⁵ en montrent l'attrait pour les usagers:

³ Voir la page « Mémoires et thèses électroniques » à l'adresse suivante : <http://www.fes.ulaval.ca/sgc/mte> [Dernière consultation le 10 juillet 2006].

⁴ Un fascicule relié d'une centaine de pages, intitulé « Thèses et autres écrits académiques : identification et localisation », rédigé par Christophe Langlois et édité par la BNF, permet d'appréhender la complexité de cette démarche.

⁵ Voir notamment les différents décomptes rapportés dans le dossier consacré aux « Thèses à l'ère du numérique » dans Arabesques n° 43 juillet - août - septembre 2006 : Claire Douady évoque une moyenne de 700 visiteurs par jour pour le

l'archive du CCSD, qui comprend 4 000 thèses, comptabiliserait ainsi plus de 1 000 téléchargements par jour. Globalement, la consultation des thèses en ligne s'avère élevée, et sans commune mesure avec celle des thèses sur papier, évaluée quant à elle à une fois tous les dix ans en moyenne⁶.

1.3. L'université de soutenance : l'opérateur ad hoc d'une mise en ligne

Les thèses comme indicateurs qualitatifs de l'établissement

Dans la partie consacrée à la recherche de son contrat quadriennal, l'université de Cergy-Pontoise affiche sa volonté de **mettre en avant la qualité de la recherche**, qui passe par une évaluation efficace, comme garant de « la crédibilité et la pertinence de ses choix vis-à-vis de ses partenaires »⁷. Il est par ailleurs précisé que cette évaluation doit s'exercer à deux niveaux : « celui des équipes (voire des individus) et celui de l'établissement. »

La publication annuelle du classement des meilleures universités au niveau mondial a instauré, de fait, un climat de compétition entre les différentes institutions : il s'agit désormais de produire de travaux scientifiques de qualité, et ce en quantité, mais aussi d'être capable d'améliorer l'impact des productions scientifiques. Un point stratégique consiste également à **garantir aux institutions de recherche une certaine visibilité**, et ce jusqu'à l'échelle des différentes entités (laboratoires, équipes de recherche...) qui en font partie.

Or, les **thèses** peuvent être considérées comme un **indicateur pertinent**, parmi d'autres, d'une telle évaluation. Telle est l'observation déjà effectuée en 2002 par Ana Pavani et Joseph M. Moxley, ancien maître d'œuvre de l'ETD 1999, dans un article intitulé « Quelle est l'utilité des thèses sous forme numérique ? », extrait du *Guide sur la production et la diffusion des thèses sous forme numérique*⁸, proposé en ligne par l'Unesco :

« La qualité d'une université est visible dans les travaux intellectuels de ses étudiants. Les thèses et les mémoires montrent l'habileté d'une université à

projet de Limoges, et une consultation de l'ordre de 5000 fois en ligne, pour une thèse habituellement empruntée une dizaine de fois en bibliothèque. Jean-Paul Ducasse estime quant à lui à 6000 la moyenne mensuelle des visiteurs pour Cyberthèses. <http://www.abes.fr/abes/DocumentsWebAbes/abes/arabesques/Arabesques43.pdf>

⁶ Ces chiffres sont rapportés par Diane Le Hénaff-Stitelet et Catherine Thiolon dans un article paru en octobre 2005 dans *Documentalistes-Sciences de l'information*, « Gérer et diffuser des thèses électroniques ».

⁷ Extrait du volet recherche du contrat quadriennal : lien intranet <http://www.u-cergy.fr/rubrique598.html>

⁸ Voir le texte extrait du *Guide pour la production et la diffusion des thèses*, proposé par l'Unesco à l'adresse suivante <http://www.etdguide.org/> et plus précisément pour le chapitre concerné à l'adresse : <http://www.etdguide.org/content/2.1%20Why%20ETDs.htm> [dernière consultation le 12/07/06]

guider ses étudiants et à encourager le travail original. Un jour, lorsque les bibliothèques de thèses sous forme numérique seront plus répandues, les étudiant(e)s et les facultés jugeront la qualité d'une université en fonction de sa bibliothèque numérique.»

L'université de soutenance : au cœur du dispositif de diffusion

Les thèses, en tant que documents emblématiques des universités, et stratégiques à la fois pour l'image de marque des chercheurs et de l'établissement lui-même, semblent convoquer tout naturellement la **notion de diffusion institutionnelle**.

Il peut s'agir d'une diffusion a minima (uniquement une identification dans le catalogue), ou, plus avantageusement pour l'utilisateur, en texte intégral, sous différents formats, et avec des clefs de recherches plus ou moins sophistiquées.

Dans les paragraphes suivants seront définis quelques termes, et les enjeux afférents, de quelques modes de diffusion en pleine expansion, à savoir les archives institutionnelles, et plus spécifiquement les archives ouvertes.

1.3.1. La notion d'archive institutionnelle et de dépôt institutionnel

Un archive institutionnelle peut être définie comme une collection constituée de documents numériques, issus de la production intellectuelle des différentes communautés de recherche et d'enseignement au sein d'une institution.

Les dépôts institutionnels centralisent, préservent et rendent accessibles les connaissances (articles, versions numériques de thèses et mémoires, mais aussi tous autres types de documents) générées par les établissements universitaires et de recherche⁹.

A l'heure actuelle, les universités peuvent choisir entre créer une archive, ou renvoyer à d'autres bases de données ou archives thématiques existantes, elles-mêmes institutionnelles. Dans ce dernier cas, cependant, l'identité de l'université, et l'affiliation de ses chercheurs ou laboratoires, peuvent être plus ou moins immédiatement visibles.

Lorsque l'université opte pour une archive qui lui est propre, c'est-à-dire voulue, conçue et supportée par elle, elle encourage son personnel à déposer ses publications en les valorisant. Dans ce cas, l'institution ne prendra en charge que les productions scientifiques réalisées à son niveau. **Elle peut ainsi mettre en**

place un système de contrôle et de validation des données au moment du dépôt, et renforcer son « image », ainsi que la lisibilité de ses activités scientifiques.

Cela n'exclut pas un éventuel dépôt dans d'autres archives, et en outre, par le jeu de l'affichage de métadonnées¹⁰ normalisées (voir infra sur le protocole OAI-PMH, notamment), le signalement et/ou le lien vers l'archive originelle sur différents portails thématiques.

Ce type de fonctionnement présente trois mérites, qui sont autant d'enjeux pour l'établissement :

- Proximité signifie incitation, rapidité et meilleur suivi ;
- Pérennité grâce à l'aspect institutionnel (notions de politique suivie de conservation, et de mémoire institutionnelle) ;
- Importance de l'affiliation pour la lisibilité et le prestige, et dans ce contexte, **nécessité d'optimiser la qualité des métadonnées.**

Sur ce dernier aspect, l'article intitulé « Métadonnées pour une cyberinfrastructure de la recherche : Le cas des thèses françaises », publié en mai 2006 sur le site ARTIST¹¹, apporte un éclairage fondamental.

Outre la notion essentielle d'interopérabilité et de qualité des métadonnées bibliographiques, les auteurs soulignent en effet, **dans un contexte de pilotage de la recherche**, le caractère éminemment stratégique d'autres types de métadonnées, telles que vocabulaires/terminologies et affiliations, **préalablement indexés et homogénéisés au sein de référentiels par des professionnels de la gestion documentaire**. A leurs yeux, ces métadonnées « doivent également s'avérer aptes à la manipulation par des outils infométriques en vue de produire des indicateurs et analyses scientifiques ou stratégiques ». Et de souligner que le service de l'évaluation impose aux chercheurs de fournir, tous les quatre ans, une information (sur leur production scientifique) déjà existante dans l'entrepôt documentaire, mais pas forcément exploitée en ce sens.

⁹Source : Papyrus Dépôt institutionnel numérique Université Montréal <https://papyrus.bib.umontreal.ca/dspace/udem/faq.jsp#Questcequundépôtinstitutionnel>, dernière consultation le 5 septembre 2006.

¹⁰ On entend par métadonnée un ensemble d'informations qui décrit une ressource informationnelle, soit externe, soit interne au fichier. Ainsi, dans un « Guide pour les revues numériques », Guylaine Beaudry et Gérard Boismenu en donnent l'illustration suivante : « Les informations bibliographiques, les résumés, les termes d'indexation, les abstracts, tout ce qui peut être un substitut au document original et qui libère les usagers potentiels de la nécessité de connaître à l'avance l'existence et les caractéristiques de ce document sont des métadonnées. ». Consultable sur le lien < <http://revues.enssib.fr/titre/6tecnic/3verselec/9meta.htm> > Dernière consultation le 5 septembre 2006.

¹¹ Voir sur cette question précise l'article intitulé « Métadonnées pour une cyberinfrastructure de la recherche : Le cas des thèses françaises », publié en mai 2006 sur le site ARTIST : http://artist.inist.fr/article.php3?id_article=332 [Dernière consultation le 12 septembre 2006].

Ce critère s'avère donc majeur dans le choix du support de collecte et de diffusion : **le système permettant l'indexation la plus fine et la plus communément usitée au sein de la communauté scientifique, au moins à l'échelle nationale, sera aussi à terme le plus efficace, et celui qui offrira le plus de possibilités pour les différents usages à venir** que pourraient envisager les partenaires du projet¹².

1.3.2. La notion d'archive ouverte

La notion d'archive ouverte désigne également « un espace où sont déposés, décrits et conservés des documents scientifiques ». Mais elle renvoie plus spécifiquement à la notion d'interopérabilité¹³, et donc à l'utilisation d'une chaîne de logiciels libres et de standards, qui dérivent des recommandations préconisées par le mouvement de l'Open Access¹⁴, terme désignant la mise à disposition libre de contenus numériques.

Le concept d'archive ouverte s'appuie en outre, généralement, sur le principe d'autoarchivage, avec des liens entre déposant/auteurs/gestionnaires et un protocole de communication.

1.3.3. Dépôt compatible avec OAI

Originellement, l'Open Archives Initiative (Initiative pour des archives ouvertes), généralement abrégée en OAI, désigne un projet visant à fédérer les banques informatisées d'archives d'articles scientifiques, de façon à ce qu'elles puissent être interrogées en une seule requête, quelle que soit leur localisation sur le réseau

¹² Lors d'un entretien avec le Pr. Tuong T. Truong, directeur de l'école doctorale de sciences et ingénierie, celui-ci nous a fait part de l'existence d'un référentiel propre aux unités de recherches, et dont la reprise dans la structuration des thèses, au sein d'un support de diffusion, représenterait un net avantage, dans cette même logique d'évaluation des résultats de la recherche produite au sein de l'école doctorale.

¹³ « L'interopérabilité est la capacité de deux systèmes à dialoguer et à interagir. L'interopérabilité technique permet à des systèmes de communiquer grâce à des protocoles et langages similaires ou pour lesquels il existe une procédure d'équivalence. L'interopérabilité sémantique est possible lorsque les métadonnées sont similaires ou comprennent des liens d'équivalences car elles représentent les mêmes concepts. L'interopérabilité syntaxique suppose que les métadonnées ont une syntaxe similaire ou qu'une procédure d'équivalence existe. L'interopérabilité organisationnelle suppose que les modes d'organisation soient similaires et que les personnes qui assurent la maintenance des données aient des fonctions et des qualifications comparables. » Définition donnée par François NAWROCKI dans "Le protocole OAI et ses usages en bibliothèque : Qu'est-ce que le protocole OAI ». Consultable en ligne : <http://www.culture.gouv.fr/culture/dll/OAI-PMH.htm> [Dernière consultation le 12 septembre 2006].

¹⁴ L'Initiative de Budapest pour le Libre Accès à la Recherche (BOAI - Budapest Open Access Initiative) marque en 2002 l'un des principaux jalons de ce mouvement, suivie par la déclaration de Berlin, en octobre 2003, qui prône l'accès libre aux archives scientifiques. Pour les signataires, de hauts responsables d'instituts européens de recherche, ce choix doit dynamiser le partage et la recherche d'informations, ainsi que la diffusion des connaissances, menacés selon eux par les éditeurs de revue aux archives payantes. Cet appel met l'accent sur l'emploi d'outil concrets : des formats d'échange et de recherche entre les différentes bases d'archives, tels que ceux développés par l'Open archives initiative (OAI) et des licences ouvertes empruntées au monde libre du logiciel libre. Pour un panorama sur la question, voir la synthèse établie par le SCD de la Réunion sur le lien suivant : <http://bu.univ-reunion.fr/ebibliotheque/oai.html> [Dernière consultation le 12 septembre 2006].

mondial. A cette fin, des **protocoles d'échanges** de données ont été définis de manière standardisée.

L'OAI permet à des fournisseurs de services de moissonner des métadonnées sur les sites de fournisseurs de données. L'implémentation technique de ce concept est définie dans l'[Open Archives Initiative Protocol for Metadata Harvesting](#) (OAI-PMH). Ce protocole d'échange, basé sur XML et le Dublin Core, permet l'échange de métadonnées entre fournisseurs de données et fournisseurs de services. Il permet de créer, d'alimenter et de tenir à jour, par des procédures automatisées, des réservoirs d'enregistrements qui signalent, décrivent et rendent accessibles des documents, sans les dupliquer ni modifier leur localisation d'origine¹⁵.

Il rend donc possible la valorisation d'une base documentaire existante en exposant ses objets documentaires, puisqu'il suffit d'ajouter une couche supplémentaire de métadonnées, selon des formats prescrits par l'OAI, soit au minimum le Dublin Core non qualifié.

Dans ce cas, l'établissement fait évoluer la chaîne de collecte et de traitement des productions scientifiques, en **constituant un catalogue institutionnel exhaustif, dont une partie** des travaux, respectant les règles de copyright et présentant un certain niveau de validation scientifique, **pourra être affichée en OAI.**

Cela permet à l'institution de conserver un contrôle sur la validation des productions déposées et sur les droits liés à leur affichage en Libre Accès.

2. Les recommandations ministérielles

Un ensemble de textes réglementaires encadrent le dépôt, le signalement, la diffusion des thèses en France. Leur application est confiée à la Direction de l'Enseignement Supérieur.

Le Ministère de l'Education nationale a, depuis le rapport Jolly¹⁶, manifesté sa volonté de valoriser et de pérenniser les thèses françaises grâce à leur diffusion par voie électronique. : la circulaire n° 2000-149 du 21 septembre 2000 visait une diffusion en ligne des thèses plus systématique. Le dispositif prévoyait alors le dépôt électronique, la conversion du document au format XML en vue d'un archivage pérenne, une diffusion

¹⁵ Extrait de "Le protocole OAI et ses usages en bibliothèque : Qu'est-ce que le protocole OAI ?", par François NAWROCKI, Ministère de la culture et de la communication, Bureau des politiques documentaires de la DLL. Consultable en ligne : <http://www.culture.gouv.fr/culture/dll/OAI-PMH.htm> [Dernière consultation le 12 septembre 2006].

¹⁶ <http://www.sup.adc.education.fr/bib/Acti/These/jolly/entete.htm> consulté le 05.07.2006

institutionnelle, en format HTML ou PDF, un signalement dans le SUDOC, et un archivage, soit local, soit au CINES, ou les deux.

En 2004, un rapport établi par la société Six et Dix¹⁷, sur mandat de la Direction de l'Enseignement Supérieur, tirait un premier bilan des expériences menées dans différents établissements après parution de cette circulaire. Ce rapport conclut à une application limitée du dispositif, et ce pour des motifs à la fois techniques et opérationnels¹⁸.

La **circulaire suivante (n° 05-094) du 29 mars 2005** tient compte de ces difficultés. Priorité est alors donnée à la rapidité de diffusion ; **la structuration en XML n'est plus considérée comme décisive dans la nouvelle circulaire**, qui préconise un dispositif plus souple que celui de l'an 2000. L'accent y est toujours mis sur le dépôt électronique, le signalement dans le SUDOC, l'archivage au CINES, et **la diffusion s'effectuera en priorité au sein de l'établissement**. Le nouveau circuit¹⁹ prévoit notamment que

- L'établissement habilité à délivrer le doctorat détermine la feuille de style²⁰ et les formats de production et de diffusion de ses thèses ;
- A des fins d'archivages, ce même établissement fournira, quels que soient les choix de traitement opérés en amont, une copie en format PDF ou XML ;
- Chaque thèse doit être accompagnée de métadonnées descriptives et de gestion, structurées en XML selon le schéma TEF (Thèses électroniques françaises)²¹ ;
- Il revient à l'Agence Bibliographique de l'Enseignement Supérieur (ABES) d'enregistrer le dépôt de chaque thèse, d'en extraire et d'en collecter les métadonnées, et de leur attribuer un identifiant permanent via une application baptisée STAR (Signalement des thèses, archivage et recherche)

La circulaire de 2005 introduisait par ailleurs de futures contraintes, notamment que la thèse électronique devienne à terme le document officiel et canonique, en lieu et place du document papier, contraintes qui devaient être officialisées dans un nouvel arrêté prévu pour la fin 2005.

¹⁷ Consultable sur le lien <http://www.sup.adc.education.fr/bib/Acti/These/rapportV5.doc> [consulté le 05.07.2006]

¹⁸ Le rapport relevait alors que 645 thèses étaient diffusées en ligne, dont 360 au format XML ; sur quatorze établissements ayant mis en place un dépôt mixte, huit utilisaient une chaîne de conversion XML. Le manque de moyens humains en regard des difficultés techniques était principalement mis en avant dans ce rapport pour expliquer la modestie de ces résultats.

¹⁹ Source : Arabesques, novembre-décembre 2005, n°40, p.2-5.

²⁰ Idéalement, le SCD serait habilité à définir cet feuille de style en concertation avec les doctorants pour définir au mieux leurs besoins, et avec le soutien éventuel des services informatiques de l'université, en vue d'une conversion aboutie.

²¹ TEF est une recommandation produite par un groupe d'experts ad hoc CG46/CN357/GE5 qui, dans le cadre de l'AFNOR, rassemble des documentalistes, des bibliothécaires et des informaticiens issus de divers établissements de l'Enseignement Supérieur et de la Recherche. Voir la présentation de la recommandation TEF sur le site de l'ABES : <http://websrv.abes.fr/abes/DesktopDefault.aspx?tabid=403> ainsi que la page <http://www.abes.fr/abes/documents/tef/index.html> [Dernières consultations le 13/07/2006]

Au final, l'**arrêté** attendu relatif aux modalités de dépôt, de signalement, de reproduction, de diffusion et de conservation des thèses ou des travaux présentés en soutenance en vue du doctorat, en date **du 7 août 2006** et paru le journal officiel n° 195 du 24 août 2006, ne tranche pas sur la nature du document officiel déposé par le doctorant : il est ainsi spécifié, au titre premier de l'arrêté, **que la thèse peut être déposée « soit sur support papier, soit sous forme électronique »**.

Cette formulation soulève plusieurs questions : ce choix de la forme est-il laissé à la discrétion de l'établissement de soutenance et des services concernés en son sein, ou bien revient-il au doctorant lui-même ? Et dans ce cas, une décision des conseils scientifique et d'administration, officialisant une unique chaîne de traitement (par exemple, électronique à l'exclusion du papier), pourrait-elle être remise en cause par un doctorant en désaccord avec ce choix ? En d'autres termes, l'université sera-t-elle tenue de proposer deux chaînes distinctes et concurrentes, en attendant que soit statué plus précisément sur cette question ? L'ABES donne l'interprétation suivante à cette alternative dans la FAQ de STAR²² : *« Dans le cadre d'un dépôt électronique, le seul cas où le doctorant peut être amené à fournir des exemplaires papier concerne le déroulement de la soutenance et uniquement si l'établissement ne se charge pas de cette reproduction papier. Les exemplaires papier sont destinés aux membres du jury personnellement. Ils n'ont pas vocation à être conservés ou diffusés après la soutenance. »*

Quoiqu'il en soit, même si l'UCP opte pour une chaîne de traitement tout électronique, une phase de transition, où coexisteront circuit papier et circuit électronique, semble s'imposer matériellement, ne serait-ce que pour éclaircir ces zones d'ombre.

Par contre, **dans le cas d'un dépôt sous forme électronique**, et dans ce cas uniquement, est confirmée l'obligation d'utiliser l'application STAR (article 10) et de jeux de métadonnées « conformes à la recommandation nationale TEF » (article 8). **L'arrêté, en précisant que le doctorant devra fournir sa thèse « sous forme numérique selon les prescriptions de l'établissement de soutenance » (art. 8) affermit en outre la position centrale et le caractère décisionnaire de l'université en la matière.** D'après le libellé du texte, le dépôt électronique dispense le doctorant de fournir des exemplaires papier au SCD, **ce qui suppose donc un mode de diffusion lui-même en ligne au sein de l'université.**

²² Voir la FAQ de STAR : <http://www.abes.fr/abes/DesktopDefault.aspx?tabindex=6&tabid=532#> Dernière consultation le 10 octobre 2006.

3. Mise en place d'un circuit de diffusion des thèses et questions juridiques

Les textes officiels et recommandations ministérielles sont listés dans la partie de la bibliographie consacrée à ce thème.

Seuls quelques aspects particuliers feront ici l'objet d'un développement : dans un premier temps, les points les plus communément problématiques rencontrés par les établissements ayant déjà mis en ligne, et dans un second temps, le type de solutions adoptées par certains de ces établissements.

3.1. Les aspects juridiques afférents aux thèses électroniques

La thèse est à la fois un document administratif, dont le dépôt est obligatoire (arrêté du 7 août 2006), et une œuvre, c'est-à-dire un document protégé par le droit d'auteur.

Le doctorant en tant qu'auteur de cette œuvre, jouit seul de l'ensemble des droits liés à la propriété intellectuelle sur ce travail, et pas le directeur de thèse, si l'apport de ce dernier s'est limité à soumettre des idées, selon la jurisprudence.

La réglementation relative à l'accès aux documents administratifs (arrêté du 25 septembre 1985, circulaire n°2000-149 du 21 septembre 2000) stipule quant à elle que les thèses doivent être consultables et mises à disposition dans les bibliothèques universitaires²³, lorsqu'elles répondent aux obligations suivantes : avoir reçu un avis favorable du jury, ne pas être confidentielle, et bénéficier de l'autorisation du président de l'université de soutenance.

La mise en ligne des thèses sous format électronique convoque quant à elle deux autres aspects du droit de la propriété intellectuelle, à savoir le droit de reproduction et le droit de représentation. Comme le rappelle l'article 11 de l'arrêté du 7 août 2006, « la mise en ligne de la thèse sur la toile est subordonnée à l'autorisation du nouveau docteur sous réserve de l'absence de clauses de confidentialité ». Le doctorant est donc tout à fait fondé à refuser une telle diffusion.

²³ Voir sur ces questions la synthèse proposée dans « Actualités du droit de l'information, n°20, décembre 2001 : http://www.adbs.fr/site/publications/droit_info/dec2001.pdf, et l'article d'Alain MARTER sur le site de l'enssib sur le lien suivant : http://www.enssib.fr/article2.php?id_cat=76&idx=39&cat=La+biblioth%C3%A8que [Dernière consultation le 14 septembre 2006]. Voir aussi, pour une synthèse récente des différents aspects, le mémoire de DCB 14 intitulé *Projet de mise en place d'un circuit de traitement et de diffusion des thèses électroniques au Service Commun de la Documentation de Paris III : aspect administratif et juridique*, par **BERNAUDIN, Anne-Claire, DEGEZ, Camille, DUPUICH, Lucile, FEUCHOT, Nicole, NGUYEN, Claire, PINET, Nicolas**.

3.2. Droit de diffusion versus droit d'auteur et commercialisation

L'enjeu pour les bibliothèques consiste donc, en premier lieu, à **obtenir de l'étudiant**, lors du dépôt de sa thèse, **le droit de diffusion** et éventuellement de reproduction, **via un contrat**. La plupart des étudiants tablent sur une diffusion large, et surtout rapide, de leurs travaux sous format électronique, en vue d'un accès facilité au marché de l'emploi. Toutefois, des réticences s'expriment parfois, notamment face au risque de compromettre une possible publication, ou des prises de brevets.

Il est donc important que l'accord de diffusion puisse tenir compte du plan de carrière, ou des projets ultérieurs de publication, en offrant **plusieurs possibilités** concernant :

- **l'étendue de la diffusion** :
 - diffusion à tout l'internet sans restrictions,
 - diffusion à tout l'internet, à l'exception de certaines parties,
 - diffusion limitée au campus ou à l'intranet de l'université,
 - diffusion différée dans le temps.
- **la rapidité de la diffusion** : Diffusion **immédiate ou différée** dans le temps ; une telle procédure devant être conçue comme exceptionnelle, eu égard à la complexité introduite par une telle retenue dans le processus usuel.

Le principe, rappelé par l'arrêté du 7 août 2006, étant que la décision de diffuser ou non revient aux doctorants, ces options ne doivent pas être considérées comme des contraintes, mais plutôt comme des possibilités d'accroître la participation des étudiants et directeurs plus réticents. **Elles doivent en outre s'accompagner d'actions de sensibilisation aux différents régimes de droits, notamment dans le cadre des formations déjà dispensées aux doctorants.**

Un certain nombre d'universités et d'établissements de recherche, sensibilisés à la question des archives ouvertes, promeuvent par ailleurs auprès des doctorants et chercheurs la notion de Creative Commons²⁴, à savoir des contrats-type (ou « licences ») inspirés par les licences libres, les mouvements open source et open access, visant à faciliter la mise à disposition d'œuvres en ligne (textes, photos, musique, sites web...).

²⁴ Le service Erad de Lyon 2 introduit notamment cette notion dans ses pages institutionnelles : <http://theses.univ-lyon2.fr/?q=fr/node/42> et renvoie au site des promoteurs français de Creative Commons : <http://fr.creativecommons.org/>

3.2.1. Formulaire/contrat type

Un formulaire type, ou contrat, doit donc être conçu au sein de l'université, dont l'ambition irait bien au-delà de l'autorisation de diffusion actuellement usitée (voir infra dans la partie « analyse de l'existant »).

Ce contrat doit idéalement :

- Mentionner toutes les implications qu'entraînent la diffusion sur support électronique des thèses et mémoires ;
- Garantir en retour au doctorant que l'établissement assurera l'intégrité du document (respect de la forme et du contenu de la thèse, à un instant donné mais aussi à plus long terme).²⁵

Cette notion d'intégrité et de sécurité du document amène également à s'interroger sur les formats retenus (voir infra, dans la partie consacrée à ce thème).

Par ailleurs, le droit de diffusion pour les cas de thèses publiées sous forme d'articles, pour lesquelles les droits de publication ont été cédés aux éditeurs, suppose qu'un *distinguo* puisse être fait entre l'article, version concentrée reconnue par les pairs, et la thèse, plus détaillée (puisque cette dernière fait état des résultats, et de la démarche scientifique qui a accompagné leur élaboration).

3.2.2. La question du droit de citation : les obligations du doctorant

La reproduction libre d'un texte, d'une image, photographie, interview, etc, est interdite par l'article L.122-4 du Code de la Propriété intellectuelle, s'il n'y a pas eu au préalable consentement de son auteur.

En revanche, lorsque l'œuvre a été rendue publique par son auteur, l'article L.122-5 alinéa 3 du Code de la Propriété intellectuelle permet d'en citer des extraits, sous certaines conditions : le nom de l'auteur ainsi que la source doivent être indiqués et il peut s'agir d'une "analyse", ou une "courte citation". **La jurisprudence a par contre exclu toute citation d'une image ou d'un son, même un simple extrait sonore ou le détail d'une image.**

Ce type de problème se pose régulièrement avec les thèses, qui emploient souvent des documents extérieurs, comme support d'analyse, ou comme illustrations.

Or, en diffusant des thèses litigieuses sur ce point de droit, l'université endosse par la même une responsabilité dans l'atteinte aux droits des auteurs originaux. Il est donc fondamental, là encore, de sensibiliser les doctorants à ces

²⁵ Le propriétaire d'un serveur endosse en effet la même responsabilité qu'un éditeur dans le mode de l'imprimé.

questions, et de **réclamer de leur part un engagement à respecter ce droit de citation.**

3.3. Exemples de solutions retenues par des établissements français

Les établissements déjà en lice pour une mise en ligne de thèses électroniques ont en général opté, **afin de remédier à ces difficultés légales, pour des formulaires signés par les doctorants au moment du dépôt de la thèse.** Ces solutions seront ici abordées depuis les formules les moins détaillées, aux dispositifs les plus contraignants.

3.3.1. Pas formulaire interne à l'établissement

La solution a minima consiste à se contenter du formulaire administratif national²⁶ commun à tous les circuits, et qui n'aborde pas la question de la diffusion en ligne. L'université Joseph Fourier de Grenoble²⁷, qui a fait le choix de l'autonomie du doctorant pour son dépôt électronique, renvoie au processus de dépôt dans TEL (l'archive Thèses En Ligne, subdivision de l'archive ouverte HAL). Le principe est ici que le doctorant, ayant opéré lui-même la mise en ligne, donne son accord de fait. Outre les exemplaires papier de la thèse et le formulaire administratif national, seule la page de compte rendu de dépôt proposée dans TEL lui sera donc réclamée par la scolarité.

3.3.2. Formulaire ne mentionnant que l'autorisation de diffuser

C'est actuellement la solution retenue par l'université de Cergy-Pontoise, mais qui s'avère peu performante à ce jour (voir infra dans l'analyse de l'existant). Ce formulaire n'est pas spécifique à la diffusion en ligne, mais concerne aussi les droits de reproduction du support papier.

Le SCD de Limoges²⁸ a, quant à lui, rédigé un formulaire d'enregistrement spécifique pour la version électronique (http://www.unilim.fr/scd/theses/Form_these_elec.pdf), qui comporte l'autorisation à signer par l'auteur ET par le jury pour la mise en ligne de la thèse. Ce formulaire s'appuie sur une charte du dépôt électronique approuvée en conseil²⁹.

²⁶Formulaire d'enregistrement de thèse soutenue : <http://www.abes.fr/abes/documents/theses/formulairetheses.doc>

²⁷Voir la page de consignes : <http://tel-ujf.ujf-grenoble.fr/thesard.htm>

²⁸<http://www.unilim.fr/scd/theses/difthese.htm>

²⁹<http://www.unilim.fr/scd/theses/Charte-these.pdf>

3.3.3. Contrat de diffusion détaillé

L'INSA de Lyon (projet CITHER, <http://docinsa.insa-lyon.fr/these/index.php?rub=04>) ne se contente pas d'une simple autorisation de diffusion accordée par le doctorant, mais réaffirme que l'auteur conserve tous ses droits d'auteur, et délimite l'étendue spatiale et temporelle de cette autorisation. Il est précisé que l'auteur peut y apporter des modifications ultérieurement. En outre, et c'est là une spécificité, le document précise les limites, à la fois techniques et juridiques, endossées par Doc'Insa :

- le doctorant, en tant qu'auteur de la thèse, est tenu pour seul responsable du contenu de son travail, ce qui évacue pour l'institution la question du droit de citation.
- **L'école se réserve également le droit de faire usage ou non de l'autorisation de diffuser qui lui est concédée.**

<http://docinsa.insa-lyon.fr/these/contenus/doc/contrat-auteur-these.rtf>

3.3.4. Attestation de conformité de la version électronique avec la version papier

Certains établissements, eu égard au caractère administratif (passeport pour le grade de docteur) et à la valeur de preuve de la thèse, doublent le contrat de diffusion proprement dit d'un **formulaire supplémentaire, devant garantir que la version électronique de la thèse est bien conforme à la version canonique de soutenance**. C'est le cas notamment de l'université de Marne-La-Vallée (UMLV)³⁰, de l'Ecole doctorale de Sciences-Po Paris³¹, et de l'université Lyon 2-Lumière.

Université de Marne-La-Vallée (UMLV)

L'université de Marne-La-Vallée (UMLV) propose un contrat de diffusion (http://www.univ-mlv.fr/bibliotheque/bib_ligne/theses/contrat_diffusion.rtf) qui traite explicitement de la question du droit de citation, notamment dans son article 5 : « L'Université se réserve le droit de ne pas faire apparaître des documents ou portions de documents inclus dans la thèse, pour lesquels les droits de reproduction et de représentation n'auraient pas été acquis ». L'UMLV réclame en outre de ses doctorants une attestation de conformité de la version électronique avec la version papier de soutenance.

³⁰ http://www.univ-mlv.fr/bibliotheque/bib_ligne/theses/theses.htm#depot

³¹ http://ecoledoctorale.sciences-po.fr/theses/documents_administratifs/index.htm

http://www.univ-mlv.fr/bibliotheque/bib_ligne/theses/attestation_conformite.rtf

Sciences-Po Paris

L'Ecole doctorale de Sciences-Po Paris propose également un formulaire d'enregistrement de thèse³², adapté du modèle national, sur lequel figure la mention d'autorisation ou non de diffuser en ligne. Le doctorant doit impérativement s'exprimer sur ce point³³ et, en cas d'acceptation, est tenu de compléter deux autres formulaires, en deux temps :

- une [Attestation de conformité entre la version imprimée et la version électronique](#) de la thèse lors du dépôt de fichier par le doctorant
- une [Attestation de conformité entre la version électronique et la version PDF](#) de la thèse, après finalisation sous forme de fichier PDF par l'Ecole doctorale, le doctorant vérifiant lui-même la conformité de cette version.

Université Lyon 2 Lumière

Une page d'informations³⁴ rappelle d'emblée que le dépôt sous forme électronique est obligatoire à l'université de Lyon 2. Le libellé de la convention de diffusion (http://www.cybertheses.org/files/2_convention_diff.pdf) rappelle celui de l'INSA: l'auteur conserve tous ses droits d'auteur, il peut toujours modifier cette autorisation initiale, et Lyon 2 se réserve également le droit de ne pas faire usage de l'autorisation concédée. Le service ERAD met l'accent auprès des doctorants sur le respect du droit de citation. Il réclame également la signature par ses doctorants d'une attestation de conformité de la version électronique avec la version papier de soutenance (http://www.cybertheses.org/files/1_conform_auto_diff.pdf). En outre, un formulaire supplémentaire a été prévu pour le cas de nouveau dépôt électronique après corrections éventuelles : http://www.cybertheses.org/files/3_conform_corr.pdf.

3.3.5. Contrat exclusif de diffusion

L'université de Paris IV³⁵ propose le contrat aux termes les plus contraignants (http://www.paris4.sorbonne.fr/fr/IMG/pdf/accord_de_diffusion.pdf), qui stipulent

³² http://ecoledoctorale.sciences-po.fr/doctorat/formulaire_enregistre_these_autoris_electron.rtf

³³ Le libellé sur la page d'information est le suivant : « Attention ce formulaire est obligatoire, même si vous ne souhaitez pas que votre thèse soit diffusée sur Internet, auquel cas, vous devez cocher la case NON ».

³⁴ <http://www.cybertheses.org/?q=fr/node/8> Dernière consultation le 10 octobre 2006.

³⁵ http://www.paris4.sorbonne.fr/fr/article.php?id_article=3368

que le doctorant lui cède ses droits patrimoniaux ; l'auteur conserve cependant ses droits pour l'édition papier.

En outre, **l'université réclame aux doctorants de fournir les autorisations des ayant-droits pour tout document joint à la thèse**, un modèle de lettre étant même mis à leur disposition par le service juridique pour ce faire.

Enfin, **la formation en vue de structurer la thèse est explicitement obligatoire pour toute personne désireuse de déposer.**

3.4. Bilan des exemples retenus

Quelle que soit le dispositif que choisira l'université de Cergy-Pontoise, une modification de l'imprimé actuellement usité s'avère nécessaire, dans la mesure où il ne constitue pas de véritable contrat. En outre, le caractère impératif de cette étape n'a semble-t-il pas été suffisamment portée à l'attention des doctorants jusqu'alors (voir infra dans la partie « Analyse de l'existant »).

Dans l'optique d'une refonte du formulaire, les exemples évoqués ci-dessus permettent de dégager **trois axes fondamentaux** :

- **L'autorisation de diffuser, ou non, doit être explicitement exprimée par le doctorant** ; afin que ce choix soit fait en toute connaissance de cause, les implications qu'entraînent la diffusion sur support électronique et les différents régimes de droits doivent être portés à sa connaissance
- **Le doctorant doit pouvoir disposer d'un droit de retrait** ; les procédés techniques choisis doivent donc permettre l'exercice d'une telle révocation. Inversement, il convient que les abus en la matière (multiples revirements...) soient encadrés.
- **L'établissement doit expliciter qu'il se réserve le droit de ne pas diffuser la thèse.** Ce peut-être pour des motifs juridiques (non respect du droit dans l'œuvre) ou techniques.

Partie 2 : L'Université de Cergy-Pontoise : un contexte particulier pour ces enjeux

1. Présentation de l'université et de ses composantes

L'université de Cergy-Pontoise (UCP) a été fondée par le décret n° 91 708 du 22 juillet 1991 pour faire face à l'accroissement de la population étudiante parisienne. L'établissement, implanté sur sept sites dans le Val-d'Oise (Cergy-Pontoise, Sarcelles et Argenteuil), est pluridisciplinaire et propose, à l'exclusion de la santé, une offre de formation allant du 1er au 3e cycle. 135 diplômes sont délivrés du 1er au 3e cycle, dont 29 en 3e cycle et 61 professionnalisés.

L'UCP compte 11 745 étudiants en formation initiale et continue, dont 1705 étudiants étrangers (soit 14,5 % de la population étudiante).

Après une croissance soutenue, et confrontée à partir de 1998 à une stagnation de ses effectifs, l'université a développé, en complément de ses formations classiques, des formations à finalité professionnelle en prise avec le milieu économique.

L'université de Cergy-Pontoise est ainsi composée de :

- Cinq unités de formation et de recherche (UFR), créées en 1996 : UFR droit, UFR économie et gestion, UFR langues, UFR lettres et sciences humaines et UFR sciences et techniques.
- d'un Institut de préparation à l'administration générale (IPAG) créé en 1996.
- et d'un Institut universitaire de technologie (IUT), créé en 1989 et rattaché à l'université de Cergy-Pontoise depuis 1991. L'IUT, qui regroupe huit départements, implantés sur plusieurs sites, est le seul institut de cette nature dans le département du Val-d'Oise.

1.1. La recherche à l'UCP

En dépit de la jeunesse de cet établissement, l'Université de Cergy-Pontoise (UCP) compte **20 équipes de recherche scientifique, dont 8 associées au CNRS.**

Sur les quatre dernières années, les publications dans des journaux scientifiques de nombreuses équipes de recherche de l'UCP ont augmenté de près de 30%. Parallèlement, le nombre de présentations lors de congrès internationaux a doublé, tout comme le volume de contrats signés avec l'industrie.

Pour assurer le développement de son potentiel de recherche, qu'elle souhaiterait rendre encore mieux identifiable, l'UCP a entrepris une structuration de son dispositif de recherche en **cinq pôles thématiques** :

- théories, modèles, information ;
- géomatériaux, mécanique, génie civil, environnement ;
- activités transculturelles en sciences humaines ;
- la recherche en droit ;
- les équipes expérimentales.

L'établissement, qui met par ailleurs l'accent sur la dimension internationale, notamment concernant les centres de recherche en sciences humaines, entend ainsi asseoir sa renommée, et être plus attractive.

Cette **volonté de valorisation** se traduit également par une stratégie d'ancrage des travaux des chercheurs dans le monde industriel. C'est déjà le cas de nombreux projets, parmi lesquels trois ont donné lieu à la création d'entreprises:

- deux projets en chimie :
 - SylChem (création d'une entreprise de production et de vente de réactifs chimiques innovants)
 - le projet Cellules- photovoltaïques, du laboratoire physicochimie des polymères et interfaces ;
- en biologie, le projet Enzgels (gels dynamiques enzymatiquement remodelés) de l'équipe de recherche sur les relations matrice extracellulaire cellules (ERRMECe)

1.2. Les écoles doctorales

L'université de Cergy-Pontoise (UCP) a mis en place **trois écoles doctorales** (voir annexe 1), habilitées en septembre 2000 par le ministère de l'Éducation nationale³⁶ :

- L'école doctorale **Sciences et ingénierie** : pluridisciplinaire, elle regroupe cinq DEA et onze équipes de recherche.

³⁶ Voir en Annexe 1 les effectifs des trois écoles doctorales.

- L'école doctorale **Droit et sciences humaines** : également pluridisciplinaire, elle regroupe cinq DEA et six équipes de recherche ;
- Enfin l'école doctorale **Économie et mathématiques Paris-Ouest** regroupe sur Cergy quatre DEA et deux équipes de recherche.

A l'heure actuelle, du fait de leur création récente, ces trois écoles doctorales fonctionnent de manière plutôt autonome les unes par rapport aux autres.

Aucune d'entre elles n'a à ce jour instauré une collaboration systématique à une archive institutionnelle, et, a fortiori, la constitution d'une telle archive. Si un bon nombre de laboratoires constituants possèdent déjà une certaine pratique du dépôt dans des archives ouvertes (thématiques par exemple)³⁷, il n'y a pour l'instant pas uniformisation des usages, que ce soit entre les différents laboratoires au sein d'une même école doctorale, ou de manière transversale entre écoles doctorales.

Cependant, un entretien informel avec les directeurs des écoles doctorales de Sciences et techniques d'une part, et de Sciences humaines d'autre part, a montré un réel intérêt pour une telle démarche. Les contours d'un tel projet restent encore à affiner : à ce stade, il n'avait pas encore été pris de décision concernant les modalités de dépôt des thèses sous forme électronique³⁸, ni sur la récupération et l'enrichissement des métadonnées (par qui, comment)³⁹.

Par ailleurs, conformément à leurs missions d'accompagnement et de soutien aux doctorants, les trois écoles doctorales financent déjà des formations, dispensées par des intervenants différents pour chaque école doctorale.

Si le dépôt des thèses sous forme électronique devait devenir la norme, et ne plus être seulement optionnelle comme à l'heure actuelle, **il pourrait s'avérer nécessaire de repenser la structure et la teneur de ces formations**, dans le sens d'une sensibilisation aux questions de droit et d'une harmonisation des pratiques des doctorants. Il s'agira cependant, bien évidemment, de tenir compte de leurs besoins spécifiques en formation, selon leurs niveaux et leurs domaines de compétences.

³⁷ Ainsi, on pouvait relever courant juillet 2006 que 8 documents avaient été déposés par des doctorants au sein de l'archive TEL, aucun ne relevant des sciences humaines, ce qui tendrait à montrer une plus grande familiarité des chercheurs et doctorants relevant des sciences dites "dures", à l'UCP comme ailleurs, avec les réservoirs et archives ouverts. Voir sur le site de TEL, en sélectionnant l'Université de Cergy-Pontoise comme organisme de délivrance: http://tel.ccsd.cnrs.fr/index.php?halsid=0cda8ce8868f04f733fac17fc04e32b5&action_todo=search&s_type=advanced

³⁸ Voir la partie infra sur le circuit actuel, le paragraphe intitulé "thèses déjà disponibles sous format électronique".

³⁹ Toutefois, la mise en production de STAR (voir infra la partie 2, paragraphe 1.3.2.1), donnera sous peu une impulsion significative sur ce point, puisque la répartition des tâches pour le dépôt et la validation, préétablie dans ce système, prévoit une implication importante des écoles doctorales.

La sensibilisation des doctorants aux notions de droit d'auteur (notamment sur la question du droit de citation) et de publication pourrait être notamment considérablement accentuée lors de ces formations.

Enfin, il s'agit de réfléchir, à l'échelle de l'université, et en concertation avec les écoles doctorales, au type de **communication institutionnelle** qu'il s'agirait de mettre en place à **destination des doctorants**.

A ce stade, il est important de préciser que l'administration du site web de l'université, refondu courant 2005, revient au service communication, et que dans un souci de cohérence, les différentes entités et services ne disposent pas de sites propres, mais de rubriques, au sein du site de l'UCP. Ainsi, les responsables de rubriques sont appelés « administrateurs restreints », car leur action concerne une ou plusieurs rubriques, mais pas l'intégralité du site, par comparaison à l'administrateur général qu'est le webmaster⁴⁰.

Qu'il s'agisse de la bibliothèque ou des écoles doctorales, une discussion doit donc s'établir avec le service communication pour déterminer le type et l'ampleur des informations devant être délivrées à l'intérieur de ces rubriques, et des distinctions éventuelles entre site internet et site intranet. Les spécificités de chaque école doctorale et des informations à délivrer aux doctorants (modalités de formation, type de fichier attendu, etc.) devront en outre être prises en compte, sachant que les relations entre les trois écoles doctorales restent encore informelles. Ceci afin de réfléchir à l'opportunité de fédérer cette communication, ou de segmenter les informations selon les besoins.

1.3. La Bibliothèque universitaire

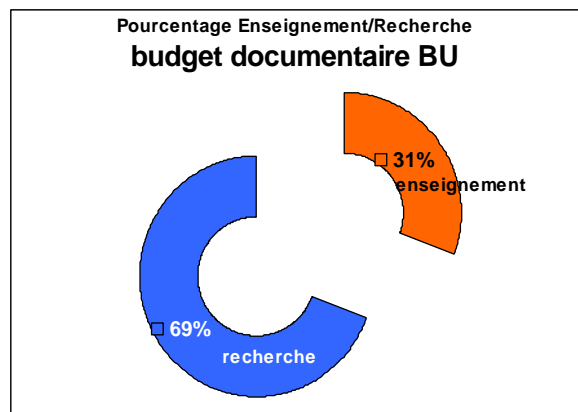
1.3.1. Présentation de la Bibliothèque universitaire

Statuts et missions

Selon ses statuts votés en octobre 1996, le service commun de la documentation, baptisé Bibliothèque universitaire de Cergy-Pontoise, a été créé pour servir la communauté des étudiants, enseignants et chercheurs de l'université de Cergy-Pontoise. Ses collections sont développées en fonction de l'évolution des

⁴⁰ Des administrateurs des rubriques ou des pages sont désignés par les chefs de services, après formation au logiciel SPIP, et après demande de création d'un compte auprès du service communication. Les procédures d'alimentation des rubriques d'information sont elles-mêmes codifiées.

enseignements et de la recherche dans une multiplicité de domaines, puisque l'université est elle-même pluridisciplinaire. Ainsi, la ventilation du budget des dépenses documentaires engagées par la Bibliothèque universitaire peut être représentée comme suit en 2005 :



Conformément à ses missions d'étude, de recherche, d'enseignement et de formation continue⁴¹, la Bibliothèque Universitaire de Cergy Pontoise (BUCP), sous l'autorité du directeur du SCD:

- constitue, développe et gère les ressources documentaires de l'Université en fonction des besoins de ses lecteurs ;
- organise et facilite l'accès de ses lecteurs à des ressources documentaires **sur tous supports** ;
- participe à la formation de ses lecteurs à la recherche d'information et de documentation sur tous supports ;
- **contribue à la production de l'information scientifique et technique de l'UCP, notamment par sa participation aux activités d'animation culturelle, scientifique et technique de cette dernière.**

Données matérielles

La Bibliothèque universitaire a réparti ses collections sur cinq sites principaux, qui correspondent à des unités documentaires, à l'exception notable de la bibliothèque « centrale », dite des Cerclades, pluridisciplinaire. Ces différents sites sont relativement éloignés géographiquement les uns des autres, hormis les bibliothèques de Saint-Martin, des Cerclades et des Chênes, toutes trois implantées sur un campus intégré dans l'agglomération de Cergy.

Les collections comptent environ 180 000 ouvrages, dont 22 % en langue étrangère. L'accroissement annuel est de 8 000 à 10 000 documents. 90 % des

⁴¹ Définies par le décret n° 85-694 du 4 juillet 1985, consultable sur le lien suivant : <http://www.dsi.cnrs.fr/rmlr/textesintegraux/volume1/143-d85-694.htm> [Dernière consultation le 7 septembre 2006].

ouvrages sont en libre accès et les collections sont réparties comme suit sur les cinq sites principaux :

Site	Nombre de documents	Domaine
Cerclades	125 000	pluridisciplinaire
Chênes	8 000	économie et droit
Saint-Christophe	3 000	gestion 1er cycle et transports
Saint-Martin	12 000	sciences dures
Neuville	13 000	sciences et techniques
Sarcelles	1 000	génie électrique, services et réseaux de communication, techniques de commercialisation
Argenteuil	1000	génie de la production, gestion logistique et transport

La bibliothèque est abonnée à environ 1200 titres de revues⁴² et propose une documentation électronique abondante, avec 1 944 revues électroniques, dont 1 890 revues scientifiques et techniques dans le cadre de l'adhésion de la bibliothèque au consortium Couperin.

Ce réseau documentaire mobilise un effectif total de 59 titulaires, stagiaires ou contractuels sur postes vacants, dont 40 personnes pour le seul site des Cerclades.

Ce site, qui constitue la bibliothèque centrale et pluridisciplinaire, compte un conservateur responsable du site, outre les deux conservateurs qui pilotent le SCD depuis les Cerclades, un bibliothécaire (un autre poste attendant la nomination d'un candidat issu du concours), 2 ingénieurs d'étude, 12 bibliothécaires adjoints et assistants de bibliothèques, 15 magasiniers en chef et magasiniers spécialisés, 5 agents et adjoints techniques. En 2005, pour assurer complètement les missions qui lui sont confiées, le SCD recourait en outre aux services de 2 CES et de 4 moniteurs étudiants, affectés aux Cerclades.

Un magasinier se charge, en collaboration avec le personnel du CRI, de la maintenance du parc de micro-ordinateurs des Cerclades (une soixantaine de postes répartis entre les postes professionnels, postes publics, postes de formation).

⁴² 850 titres sont proposés en sciences humaines, dont plus de 300 en droit et sciences politiques, près de 200 titres en économie/gestion et une centaine de titres en lettres et arts. En sciences et techniques, la bibliothèque est abonnée à 350 revues. Plus de la moitié sont des publications scientifiques étrangères.

1.3.2. Un défi pour la Bibliothèque universitaire

Si l'on reprend le libellé des missions de la bibliothèque universitaire énoncées plus haut, en tant que "contribution à la production de l'information scientifique et technique de l'UCP", les thèses électroniques relèvent donc bien de ce champ.

Le dispositif intégrant une chaîne électronique implique de nouvelles fonctions par rapport au rôle traditionnellement endossé par chaque SCD dans un circuit papier (voir sur ce point l'analyse de l'existant pour Cergy-Pontoise, typique du traitement classique): formation et assistance technique des étudiants, gestion des fichiers informatiques, et notamment indexation et aménagement des accès à ces derniers, contrôles techniques et éventuellement édition papier à la demande.

Ces différentes fonctions peuvent être assumées par les écoles doctorales ou par divers opérateurs techniques locaux, qui restent souvent à identifier, mais dont le rôle est évidemment majeur dans cette organisation.

Or, dans un article du BBF⁴³ datant de 2001, Ghislaine Chartron relevait déjà la place essentielle que pouvait, ou devrait, occuper la bibliothèque universitaire dans ce nouveau dispositif :

« N'appartient-il pas aussi aux bibliothèques de se positionner sur la formation et sur les techniques d'édition électronique, au lieu de se limiter aux traditionnelles fonctions de catalogage ? Au risque, nous semble-t-il, de voir leur rôle diminué, alors qu'il pourrait être élargi dans les reconfigurations actuelles de la documentation scientifique. »

Dans l'environnement des thèses électroniques devront nécessairement être conçus des normes (sur les métadonnées identifiant les instances par exemple), des partenariats entre les différents acteurs du circuit, des modes et des chaînes de travail différents. **Cette redistribution**, qui fournit aussi l'opportunité de développer des outils communs, **peut conférer une position favorable à la bibliothèque** si celle-ci parvient à redynamiser les liens avec les différents services : ce faisant, **la bibliothèque peut également y gagner une meilleure visibilité au sein de la communauté universitaire.**

1.3.2.1. Le choix de STAR

Le cadre réglementaire⁴⁴ concernant les thèses impose qu'une version validée de ce document soit déposée et que soient assurés :

⁴³ Consultable sur le lien suivant : <http://bbf.enssib.fr/sdx/BBF/frontoffice/2001/02/document.xsp?id=bbf-2001-02-0132-013/2001/02/fam-critique/critique&statutMaitre=non&statutFils=non>

⁴⁴Cf. les chapitres 2 et 3 de la partie 1

- d'une part son archivage pérenne
- d'autre part son signalement dans le catalogue collectif SUDOC.

Ce sont là les buts recherchés dans le développement de l'application de dépôt, diffusion et archivage STAR. Cet outil, géré par l'ABES sur mandat du ministère délégué à l'Enseignement supérieur et à la Recherche, permet de centraliser l'information au niveau national, ce qui offre de nouveaux débouchés aux thèses (en terme de recherche d'information) et assure l'existence d'au moins une version officielle de chaque thèse, disponible et accessible.

La préparation de l'arrêté du 7 août 2006 a servi de toile de fond à la conception de l'outil STAR, qui cadre donc parfaitement avec les différentes obligations instaurées par cet arrêté. C'est pourquoi la Bibliothèque universitaire l'a choisi dans le cadre d'un futur circuit électronique des thèses, dont l'architecture et la gestion demeurent à la discrétion de l'établissement. La bibliothèque a ainsi anticipé ces obligations réglementaires, et demandé, dès le mois de mai 2006, à intégrer le premier cercle d'établissements utilisateurs, qui compte une trentaine de membres⁴⁵.

La première réunion de lancement, à laquelle participera la BU de Cergy-Pontoise, doit se tenir le 12 octobre prochain, et fait suite à une phase d'expérimentation lancée en juin 2006 sur six sites pilotes⁴⁶.

Dans la distribution des rôles prévus dans STAR, il peut-être stratégique pour la bibliothèque, à l'origine de l'implantation de STAR, que son personnel soit désigné "correspondant STAR" au sein de l'UCP : l'ABES prévoit en effet de laisser le libre choix à l'université pour désigner un correspondant parmi les différents acteurs du circuit.

1.3.2.2. *Principes et caractéristiques de STAR*

STAR est une plate-forme de dépôt des thèses; c'est une application nationale légère (accès aux formulaires de saisie par internet), qui **assure l'acheminement des thèses depuis l'établissement de soutenance vers des outils et/ou acteurs** intervenant à différentes étapes dans le cycle de vie de la thèse.

Du point de vue technique, les caractéristiques précises de l'application sont assez peu connues ; il est juste précisé qu'aucune application logiciel supplémentaire n'est nécessaire pour utiliser STAR, un accès Internet suffisant pour accéder à l'espace de

⁴⁵Ce déploiement s'opère dans un premier temps dans des établissements ne disposant pas d'outil local de gestion des thèses électroniques, l'ABES travaillant à la mise au point d'une version permettant les importations depuis les systèmes locaux, pour les établissements disposant déjà d'un tel dispositif. Le but est d'éviter à terme les doubles saisies.

⁴⁶ Il s'agit de Bordeaux-I, INP de Toulouse, INSA de Lyon, Limoges (qui était déjà pilote dans SPARTE), Paris-V et Strasbourg-I.

travail dédié, conçu pour chaque établissement participant. Dans cet espace seront déposés les fichiers des thèses (telles qu'elles ont été validées par les établissements de soutenance) et renseignées les métadonnées associées (en format TEF). Par contre, les données concernant le workflow et les modalités d'implantation de la chaîne sont désormais détaillés dans une "foire aux questions" régulièrement mise à jour⁴⁷.

Les fonctionnalités essentielles de STAR y sont notamment énoncées:

- une validation globale des fichiers et des métadonnées
- **archivage** : export vers le CINES
- **signalement automatique des thèses dans le catalogue SUDOC** via une conversion des métadonnées en Unimarc :
- **indexation** : métadonnées exportées vers des portails (SUDOC, locaux...)
- **diffusion** : plate-formes/archives nationales (CINES, CCSD), ou locales, après renvoi à l'établissement de soutenance d'une version de diffusion de la thèse et des métadonnées dans le format souhaité (PDF ou HTML)
- **suivi de la thèse** : identifiant permanent associé à la thèse; résolveur de liens pour l'adresse URL vers la version canonique.

Le principe de travail repose ainsi sur **une saisie collaborative**, qui implique les différents acteurs de la chaîne de traitement des thèses, validée en bout de chaîne par un administrateur respectant les recommandations de TEF.

1.3.2.3. *Avantages et contraintes induits par l'adoption de STAR*

Un outil de workflow pertinent, mais partiel

L'avantage principal de STAR consiste en la définition d'un espace collaboratif de travail (ou workflow) : Concrètement, le workflow décrit le circuit de validation, les tâches à accomplir entre les différents acteurs d'un processus, les délais, les modes de validation, et fournit à chacun des acteurs les informations nécessaires pour la réalisation de sa tâche. Il est en cela un outil de pilotage opérationnel.

Le workflow profilé dans STAR prend en compte les différents acteurs de la chaîne de traitement des thèses que l'on retrouve dans tout établissement de soutenance, à savoir les écoles doctorales/les services de scolarité, le centre de ressources informatiques (CRI), la bibliothèque et le doctorant lui-même. Cette

⁴⁷ "FAQ de STAR", <http://www.abes.fr/abes/DesktopDefault.aspx?tabid=532.dernière> mise à jour le 27 septembre 2006,

clarification dans la répartition des rôles s'avère tout à fait positive en soi, puisqu'elle rend possible une meilleure communication et une mutuelle prise en compte du travail des différents acteurs.

Ainsi sont prédéfinis quatre rôles, chacun avec ses attributions et son identifiant :

- **Rôle DEPO** : l'école doctorale, le service de scolarité. Le premier à intervenir dans la chaîne, au moment du dépôt du sujet de la thèse. C'est le DEPO qui crée l'identifiant pour le doctorant (DOCT), qui renseigne la plupart des métadonnées de gestion, et qui effectue le dépôt physique du fichier thèse.
- **Rôle DOCT** : le doctorant. Le doctorant peut saisir lui-même les résumés de sa thèse et ses propositions de mots-clé. META a également la possibilité de faire ces saisies si l'établissement ne souhaite pas que les doctorants interviennent dans STAR.
- **Rôle METAB** : la bibliothèque qui saisit les métadonnées descriptives, les données bibliographiques (autorités...).
- **Rôle VALID** : le valideur final, qui intervient en dernière instance (après DEPO et METAB) et valide l'ensemble des données saisies, déclenchant ainsi l'envoi au CINES du fichier thèse (pour archivage) et le signalement dans le catalogue Sudoc.

Grâce à ce workflow préétabli et réparti en amont de cette application, la redistribution du circuit entre les différents acteurs, nécessaire à la faveur du passage aux thèses électroniques, a déjà été pensée et éprouvée.

Restent malgré tout deux aspects à définir à l'échelle de l'UCP :

- D'une part, **la répartition des identifiants** (dont le nombre est limité) et notamment de l'identifiant VALID (administrateur) : chaque établissement se voit attribuer, par l'ABES, des identifiants, correspondant chacun à un ou plusieurs rôles (DEPO, METAB et VALID), soit 3 identifiants maximum par établissement. Il s'agit donc de déterminer si la bibliothèque devient dépositaire uniquement du rôle METAB, ou bien également du **rôle stratégique de valideur final**.
- D'autre part, si STAR tend à automatiser un certain nombre d'opérations (notamment le signalement dans le SUDOC pour la bibliothèque), l'arrivée des thèses électroniques génère par contre de nouvelles tâches et enjeux aux différentes étapes administratives de la vie de la thèse.

dernière consultation le 9 octobre 2006.

Le circuit ‘thèses électroniques’ fait notamment intervenir **deux étapes supplémentaires, dont la responsabilité devrait par nature incomber aux secrétariats des écoles doctorales**, mais qui supposent des modifications profondes dans les pratiques quotidiennes de travail :

- 1) une **étape de vérification des normes** : il s’agira de vérifier le formatage de la thèse électronique, pour qu’elle soit bel et bien lisible (conforme à PDF/A, etc...) ;
- 2) un **dépôt de thèse sous forme électronique, au moment de la soumission** : le rôle DEPO prévoit que les écoles doctorales reçoivent et sauvegardent de manière temporaire les fichiers. Reste à déterminer avec le service informatique selon quelles modalités les transferts peuvent-ils s’effectuer, et si les espaces de travail dédiés dans STAR offrent une telle possibilité de sauvegarde temporaire.

Par contre, le workflow de STAR ne couvre pas l’ensemble de la chaîne depuis le dépôt jusqu’à la mise en ligne. L’objectif du workflow consiste en effet à valider chaque étape, depuis le premier dépôt du document numérique, et de permettre de déterminer pour chaque thèse les opérations déjà effectuées et celles restant à faire (dépôt complet, attente de validation, traitement en cours, type de diffusion, archivage...).

Il doit donc recouvrir **le dépôt du matériel informatique, la conversion (éventuellement), le signalement, la diffusion, l’archivage et enfin le contrôle et les statistiques**. Ici, seule une partie du workflow est prise en compte, à partir de la validation du jury, et avant la diffusion en ligne.

Reste à déterminer avec les administrateurs de STAR si des développements sont possibles pour élargir le champ de ce workflow propre à STAR, ou si un autre outil doit être adopté/développé au sein de l’UCP.

Ni diffusion, ni conversion : un simple intermédiaire

Cette application **n’assure pas la conversion des fichiers électroniques dans un format de données (XML et/ou PDF)** et n’a qu’une fonction d’**intermédiaire** entre l’établissement de soutenance et les différents acteurs précédemment cités.

Ce qui crée une confusion, c’est que STAR "met à disposition de l’établissement la thèse validée dans sa version de diffusion (format PDF ou HTML) avec un identifiant pérenne ainsi que les métadonnées".

Or, **STAR n’est pas non plus une plate-forme de diffusion** : l’outil peut permettre d’exporter les fichiers vers la plate-forme choisie pour ce faire (serveur ou archive ouverte) mais ne comporte pas de module de diffusion propre. Il est d’ailleurs bien précisé dans la FAQ de STAR que « l’archivage et la diffusion sur un serveur de

l'établissement et/ou un serveur national ne sont pas obligatoires. Chaque établissement est responsable de sa chaîne de conversion ».

Par ailleurs, STAR fonctionne sur le principe d'une gestion centralisée et délocalisée : l'avantage pour les établissements est de les décharger du suivi technique des serveurs (l'archivage pérenne étant assuré à Montpellier au CINES). Mais cela présente aussi un système moins facilement paramétrable au niveau local, seule l'administration des utilisateurs relevant de la maîtrise réelle de l'université.

C'est donc à l'UCP d'arbitrer sur le type de diffusion dont elle souhaite faire bénéficier ses thèses (archive ouverte, serveur local, archive nationale...) et c'est sur cet aspect en particulier - l'articulation possible entre STAR et un mode de diffusion aussi pertinent que possible pour l'UCP - que vont s'orienter les scénarii de la dernière partie (voir le chapitre 2 de la troisième partie).

2. Analyse de l'existant à l'Université de Cergy-Pontoise :

2.1. Le circuit actuel : un circuit papier

2.1.1. Le dispositif réglementaire national

Avant la promulgation de l'arrêté du 7 août 2006⁴⁸, qui abroge celui du 25 septembre 1985⁴⁹, le dépôt de la thèse s'effectuait en deux étapes :

- dépôt d'un formulaire donnant le résumé et les mots-clés déposés par le candidat avant la soutenance auprès du service de doctorat de son établissement.
- enregistrement de 3 exemplaires papiers après soutenance par la bibliothèque de l'université :

- le premier pour archivage;
- le deuxième pour la consultation sur place;
- et le troisième est transmis, sur autorisation du jury, aux Ateliers nationaux de reproduction des thèses (ANRT) pour être microfiché.

L'ANRT se chargeait de diffuser les microfiches auprès des universités du domaine scientifique correspondant à la thèse, et commercialisait à la demande des exemplaires de la thèse.

⁴⁸ Voir en partie 1 le paragraphe « Recommandations ministérielles ».

⁴⁹ <http://www.sup.adc.education.fr/bib/Acti/These/textregl.htm#25sept> [visité le 22 mai 2006]

Avec la promulgation du nouvel arrêté, qui distingue le circuit papier (dans le titre II) du circuit électronique (titre III), les dispositions concernant le circuit papier sont intégralement reprises, et les tâches de l'ANRT réaffirmées. Seule précision nouvelle et d'importance : « L'atelier national de reproduction des thèses de l'université Lille III peut en outre, sur la commande d'un établissement, assurer sur un autre support la reproduction des thèses soutenues dans celui-ci. », ce qui semble désigner non seulement le papier, mais aussi des fichiers électroniques, au format PDF ou XML par exemple.

2.1.2. Le circuit actuel au niveau du SCD

Trois exemplaires sont directement fournis à la bibliothèque par chacun des secrétariats des Ecoles doctorales.

Premier exemplaire	Deuxième et troisième exemplaires	
envoyé à l'ANRT	Exemplaire de prêt / magasins	Exemplaire pour archivage / consultation sur place
	création d'une notice minimale dans le module d'acquisition du SIGB Dynix-Horizon	Création d'une notice bibliographique, dans un délai d'environ 1 mois (d'abord dans le Sudoc, puis récupération dans Horizon)
	Cotation, qui reflète dans sa structure : <ul style="list-style-type: none"> • <i>la discipline</i> : TS pour scientifique, TD pour le droit, etc. • <i>au code</i> de l'université de Cergy-Pontoise : CERG • <i>à l'année d'enregistrement</i> de la thèse. Pour une thèse de sciences de l'année 2005 : TS CERG 2005	

Les thèses sont cataloguées par une même personne pour l'ensemble du SCD (Natalie Reynaud, responsable du projet) ; l'indexation des thèses est confiée aux responsables de chaque domaine thématique. Les thèses scientifiques (sciences dites dures) sont par exemple indexées par les acquéreurs des domaines sur les différents sites.

2.2. Volumétrie des thèses à traiter

Toutes les données chiffrées présentées dans cette partie résultent d'un état de fonds (exemplaires papier, électroniques, bordereaux de thèses...) pratiqué en juillet 2006 ; une quinzaine de thèses a été cataloguée entre temps, cependant les tendances de fond demeurent inchangées.

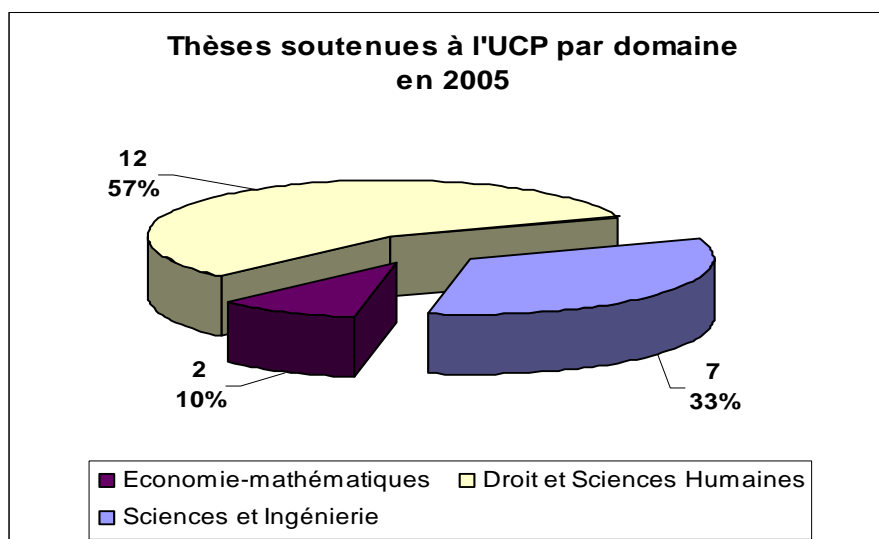
2.2.1. Production annuelle de thèses et question de la rétroconversion

La production annuelle de thèses à l'UCP

En moyenne, une vingtaine de thèses sont produites au sein de l'université de Cergy-Pontoise chaque année, avec une tendance à la hausse (plutôt 30 thèses) depuis trois ans.

Le graphique ci-dessous, élaboré à partir du fichier alimenté depuis 1993 de manière exhaustive par la personne chargée du catalogage, illustre la répartition par école doctorale des thèses de l'UCP soutenues en 2005. Ce fichier source répertorie également les thèses confidentielles ; les nombres au-dessus des pourcentages correspondent aux unités pour chaque école doctorale :

Graphique 3 : Répartition par école doctorale des thèses soutenues en 2005



En l'absence d'une étude plus poussée au cas par cas pour chaque thèse produite dans son format électronique natif, cette répartition par école doctorale donne déjà une bonne indication de ce à quoi s'attendre sur le type de fichiers électroniques et sur les traitements à prévoir pour chacun d'entre eux : **l'utilisation de LaTeX est extrêmement répandue (voire majoritaire) parmi les doctorants de sciences et ingénierie, parfois employée par des doctorants en économie-mathématiques,**

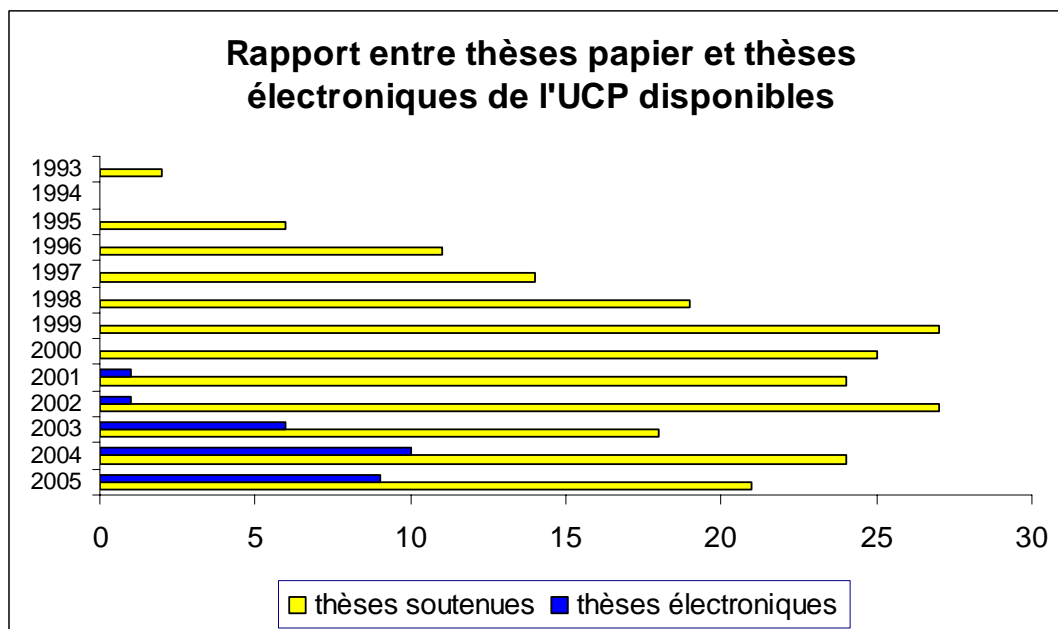
dont certains utilisent également **MathType**. Ce sont également les thèses de ces deux domaines qui comportent le plus de tableaux et figures (sans compter les cas de volumineux fichiers de calculs....).

Les thèses issues de l'école doctorale de Droit et sciences humaines sont pour leur part essentiellement produites avec des logiciels de traitement de texte, de type Word ou OpenOffice. Cependant, avec le développement d'une offre de formation en géographie (ouverture récente d'un doctorat), on peut s'attendre à la multiplication et une complexification des illustrations même au sein de ces thèses. Comme on le précisera plus avant dans ce rapport, cette proportion a une influence directe sur le choix le plus adéquat du format en vue d'une conversion et d'une diffusion en ligne (voir infra « Le choix du format »).

Les thèses déjà disponibles sous format électronique

Le dépôt des thèses est donc aujourd'hui uniquement papier.

Il est cependant demandé aux doctorants de fournir, avec les exemplaires papier, une disquette contenant les résumés afin de les incorporer à la notice de catalogage, ainsi qu'un CD-ROM comportant la version électronique de la thèse. Sur ce dernier point, nous sommes encore loin de l'exhaustivité, puisqu'en 2005, seulement 27 thèses nous étaient parvenues sous forme électronique sur les 249 soutenues au sein de l'université depuis 1993.



Il paraît donc évident que l'une des premières actions à entreprendre, dans le cas où le scénario d'un traitement et/ou d'une mise en ligne des thèses principalement sous format électronique serait retenu, consiste à rappeler auprès des doctorants : de manière rétrospective, pour ceux dont les coordonnées sont encore

disponibles, et avec les réserves d'usage sur les chances d'atteindre l'exhaustivité, et surtout, de manière prospective, auprès des futurs rédacteurs ou des doctorants en train de finaliser leur thèse.

Question de l'antériorité et de la rétroconversion

Depuis 1993, la bibliothèque est devenue dépositaire d'environ 250 thèses.

Par rapport à la ventilation par domaine des thèses soutenues à l'UCP en 2005 (graphique 3), la proportion représentée par les thèses « scientifiques » est largement plus importante dans la totalité de la collection de thèses que conserve la bibliothèque sous format papier (voir le graphique 4, dans l'annexe 2).

En l'état actuel des choses, l'université de Cergy-Pontoise dispose donc de 27 titres de thèses dans un format électronique, qui puissent être immédiatement traitées dans le cadre d'un test, quel que soit le système finalement retenu. De manière assez étonnante, ces 27 thèses disponibles sous format électronique se répartissent dans des proportions à peu de choses près comparables à l'ensemble du corpus (voir le graphique 5, dans l'annexe 2).

Il est à noter que pour l'instant, la version actuelle de STAR ne permet pas de reverser a posteriori les thèses dont nous disposons déjà sous forme de CD-ROM.

L'intervention dans STAR se comprend donc de manière prospective, et pas pour l'existant à ce jour.

Traiter l'antériorité suppose donc que la bibliothèque se dote d'une plateforme de conversion (en format PDF ou XML), ou bien qu'elle fasse appel à un prestataire extérieur, tel que l'ANRT, qui a déjà fait une proposition en ce sens. Cependant, ce dernier cas implique :

- la *notion de coût* : bien que raisonnable, le devis incite à déterminer d'une part si l'on entend numériser tout, ou partie, de la collection papier pour laquelle nous ne disposons pas d'une version électronique, et d'autre part les critères d'une telle sélection ;
- la *notion de format* : l'ANRT propose en effet une conversion au format PDF ou XML, mais le graphique montrant la proportion prédominante de thèses issues de l'école doctorale sciences et ingénierie « scientifiques » infléchit nécessairement la question du choix du format (voir infra).

2.2.2. L'autorisation de diffusion en ligne

A l'université de Cergy-Pontoise, les imprimés relatifs aux droits et à la propriété intellectuelle des doctorants (à l'époque support papier uniquement pris en compte) n'ont été mis en circulation qu'à partir de 2003, et ce de manière non systématique (les imprimés apparaissent dans les bordereaux en 2003 d'abord en chimie, biochimie et génie).

S'enquérir par écrit des désirs des doctorants sur les modalités de communication de leurs thèses ne se généralise qu'en 2004, et c'est seulement à partir de cette date qu'est mentionné spécifiquement le cas de la diffusion en ligne.

Parmi les 249 doctorants répertoriés, 43 ont donc signé un imprimé précisant les conditions de communication :

- Parmi ceux-là, 12 n'ont pas été consultés sur la question de la diffusion en ligne ;
- Sur les 31 consultés sur cette question précise, seuls 11 ont explicitement formulé le souhait de voir leur thèse diffusée en ligne ;
- Sur ces 11 doctorants, **9 ont effectivement fourni le CD-ROM de leur thèse.**
- Quatre autres doctorants ont bien fourni un CD-ROM, mais n'ont pas coché la case adéquate, peut-être par oubli, et un dernier a fourni le CD-ROM de sa thèse, mais cette dernière doit rester confidentielle jusqu'en 2008.

Déterminer clairement la volonté de diffusion ou non du doctorant est donc une étape qui, dans l'optique d'un circuit uniquement papier, a longtemps été négligée. Or, l'arrêté du 7 août 2006 stipule bien, dans son article 5, que la reproduction et la diffusion sont soumises à l'accord de l'auteur.

L'accent doit donc être désormais placé sur ce point, et il s'agit de mener en priorité un travail de sensibilisation, d'information des doctorants, et de normalisation des pratiques, dans le recueil à la fois de supports et des autorisations, et ce en coopération, au premier chef, avec les secrétariats des Ecoles doctorales.

3. Le choix du format

En tant que document administratif donnant lieu à la délivrance d'un titre (le doctorat), les thèses électroniques doivent être archivées avec un soin tout particulier. La circulaire n° 05-094 du 29 mars 2005 considère d'ailleurs « un

archivage pérenne, qui garantisse un accès durable à l'information par delà les mutations technologiques » comme l'un des quatre enjeux majeurs devant être impérativement assurés. L'archivage pérenne vise à garantir une triple fonction⁵⁰ :

- conserver le document ;
- le rendre accessible ;
- en préserver l'intelligibilité.

Or l'obsolescence technologique, phénomène inéluctable, est un facteur qui rend déterminant le choix des normes d'encodage.

Il importe de recourir à un format normalisé, reconnu, qui permettra la récupération des données dans l'avenir et qui assurera la conservation du formatage original.

Par ailleurs, les usages offerts par les différents formats, même normalisés, ne sont pas les mêmes. Il s'agit donc également de définir les usages que l'on envisage pour les documents électroniques :

- à des fins d'archivage ou de diffusion d'une part ;
- d'autre part, concernant la diffusion proprement dite, quels types de services l'on souhaite offrir, avec quelle ergonomie, quelles latitudes (impression ou non), etc...

Les outils actuellement en phase de production utilisent principalement deux formats : le format PDF et le format XML.

3.1. Le format PDF

Créé en 1993, le format PDF (ou Portable Document Format) décrit de manière standardisée tous les éléments constitutifs d'une page : il représente non seulement les données contenues dans un document, mais également l'aspect précis de ce dernier⁵¹.

Ainsi, un fichier au format PDF se compose d'une série de pages dont chacune contient du texte, des spécifications de polices, des marges, une mise en page, des éléments graphiques et des couleurs de fond et de texte. Le fichier peut aussi comporter d'autres éléments tels que des métadonnées, des hyperliens et des champs de formulaires.

⁵⁰ Voir sur cette question l'article de synthèse « L'archivage pérenne des documents numériques » in La Gazette du CINES n° 20 ; consultable en ligne : <http://www.cines.fr/textes/gazette20.pdf> [Dernière visite le 10/07/2006]

⁵¹ Extension du langage PostScript, le PDF le supplante en assurant le rendu de textes et d'images complexes, non plus seulement sur imprimante, mais également à l'écran, quel que soit le matériel utilisé.

Le PDF a pour avantage de préserver l'apparence originale du document ; il est, de plus, habituellement facile de produire un document PDF dans un environnement contrôlé. **Toutefois, le PDF est davantage un format d'impression sur demande qu'un format de visualisation ou d'exploitation.**

La simplicité de ce format – dont le sigle fait référence à " format de document transférable " en français – permet de garantir qu'un fichier au format PDF peut être affiché et imprimé avec la même fidélité sur n'importe quelle plate-forme (Unix, Microsoft Windows, Mac OS, etc.). Cependant, **il se contente de restituer à l'écran ce que l'on retrouve sur papier.**

En outre, comme **l'information d'un document PDF n'est pas structurée, les possibilités de recherche sont moindres.**

Enfin, bien que les spécifications du PDF soient accessibles à tous, permettant le développement d'outils dédiés à la création ou à la visualisation de ce format, ce dernier n'en est pas moins propriétaire, créé par la société Adobe : il n'y a donc aucune assurance quant à la pérennité de l'information. La création d'une norme récente (norme ISO 19005) pour un format d'archivage, le PDF-A⁵², vise à remédier à cette faiblesse en apportant des garanties accrues.

Cependant, le format d'archivage PDF-A implique pour l'instant un certain nombre de contraintes (utilisation obligatoire de l'UNICODE, non prise en charge de fichiers type JPEG, vidéos, aucune restriction d'accès au contenu du document)⁵³ pouvant s'avérer défavorables à l'usage souhaité dans le cadre de la diffusion de thèses électroniques.

3.2. Le langage XML

XML, abréviation de Extensible Markup Language, est un langage de balisage (ou codage de données) qui permet de décrire la structure logique des documents. C'est une représentation textuelle de données structurées, selon une syntaxe normalisée. Développé sous l'égide du World Wide Web Consortium (W3C) depuis fin 1996, XML est un dérivé simplifié du SGML, (Structured Generalized Markup Language), dont il conserve le vocabulaire standard ainsi que la méthode de description de structure.

⁵² Présentation du format sur le site de la société Adobe : <http://www.adobe.com/fr/products/acrobat/pdfs/pdfarchiving.pdf> [Dernière visite le 10/07/06]

⁵³ Voir l'analyse de la *National Digital Information Infrastructure and Preservation Program* de la *Library of Congress* sur le lien suivant : <http://www.digitalpreservation.gov/formats/fdd/fdd000125.shtml> [Dernière visite le 10/07/2006]

XML a été conçu pour permettre d'échanger et de stocker des données, indépendamment des programmes ou des processus qui les produisent ou qui les utilisent, **ce qui garantit sa lisibilité dans la durée.**

Par ailleurs, il s'agit d'un **format non-propriétaire**, déjà utilisé par l'industrie, le domaine tertiaire (banques) et la documentation en ligne dans le service public et privé.

Dans le principe, XML permet de séparer le contenu de la présentation, puisqu'il définit le fond indépendamment de la forme. Un document XML se décompose en effet en deux parties : le document proprement dit, obligatoirement présent, et une description formelle optionnelle, appelée Document Type Definition (DTD). Cette description contient les règles syntaxiques que doit respecter le document. Le langage a une grammaire, qui décrit les structures syntaxiques permises et les mots réservés.

Concrètement, cela signifie par exemple que la structuration d'une thèse, pour contraignante qu'elle soit à l'étape de production, permettra de séquencer son contenu et d'identifier des objets scientifiques spécifiques contenus dans le document, offrant ainsi des possibilités d'indexation et de recherche infiniment plus puissantes (recherches au sein des figures, des citations...).

Différentes DTD existent : EAD (Encoded Archive Document), pour les documents d'archive, Docbook pour les documents électroniques et TEI (Text Encoded Initiative), pour les données textuelles.

XML est donc un standard d'archivage et d'échange de documents qui permet à la fois une communication des documents sous un format riche et peu soumis aux évolutions techniques à venir, et procure de puissantes possibilités de recherches dans les données. Il offre, de plus, des possibilités importantes d'interopérabilité.

3.3. Préconisations et contraintes induites

Ces différents aspects seront pris en compte lors de l'élaboration des scénarii possibles pour le traitement et la mise en ligne des thèses à l'Université de Cergy-Pontoise. *Pour une synthèse sur les avantages et contraintes respectifs de chaque format, voir le tableau reproduit dans l'annexe 3.*

XML apparaît clairement comme le format le plus pertinent pour un projet de diffusion électronique des thèses, car il permet un archivage pérenne tout en autorisant une diffusion en PDF ou HTML. Il **combine les avantages des deux formats, en terme de diffusion et de conservation, tout en laissant ouvertes les possibilités de retraitement ultérieur.**

Néanmoins, parce qu'il impose une structuration forte, et, pour ce faire, l'usage d'une feuille de style⁵⁴, le format XML suppose une implication de l'auteur et l'observance d'un certain nombre d'exigences. Ce qui peut être ressenti, à juste titre, comme une contrainte forte par le doctorant.

Toutefois, l'usage d'une feuille de style peut aussi représenter un gain de temps considérable pour l'auteur, pour peu que ce dernier ait reçu une formation ou, a minima, des indications adéquates. En effet, la feuille de style présente au moins deux avantages pour le doctorant : elle lui apporte d'une part la garantie de la **portabilité** de sa mise en page, élément non négligeable pour quiconque travaille dans un contexte de partage des ressources (salles de libre accès informatique...). D'autre part, un certain nombre de fonctionnalités essentielles dans la manipulation de textes de longueur importante (telles que la génération et la mise à jour automatiques de sommaires ou tables des matières...), sont fondées sur l'utilisation de styles. Or, dans un modèle de documents, ces styles peuvent être préformatés selon les besoins du rédacteur (d'une thèse par exemple), et sont alors plus simples d'utilisation.

Le choix du format XML comme élément structurant impliquerait donc un certain nombre de modifications dans les pratiques actuelles des doctorants et des divers acteurs de la chaîne, et supposerait notamment :

- **Sur le plan technique / matériel :**
 - la mise à disposition de feuilles de style,
 - de nouvelles modalités de dépôt,
 - l'installation d'une chaîne de conversion,
 - la définition de sites de diffusion,
 - des principes et modalités de conservation.
- **Du point de vue de la communication :**
 - des sessions de formation en vue d'une utilisation plus avancée des logiciels de traitement de texte ;
 - une sensibilisation aux problèmes de droits d'auteur ;
 - des campagnes d'information sur les enjeux pour la communauté universitaire

54 A noter que dans son rapport d'activité de 2005, dans la partie consacrée aux thèses électroniques, l'ABES soulignait que « l'utilisation des feuilles de styles ne garantit aucunement (et même le plus souvent brouille) le principe de séparation du contenu d'avec sa présentation. La conclusion à tirer de cette réflexion est que seule la génération native du XML (TEI, Docbook,...) permet cette indispensable séparation ». Une feuille de style sans DTD et sans sensibilisation des doctorants n'est donc pas une condition suffisante pour un traitement correct des documents, relève l'ABES.

Cela suppose un investissement non négligeable à court terme, en temps comme en moyens. Mais s'il s'agit de choisir la solution la plus ouverte, permettant au mieux une réévaluation des options à la lumière des évolutions futures, sans perte de données et sans coût prohibitif, la faveur revient indubitablement au choix du format XML à plus long terme.

En outre, même si l'implantation d'un circuit fondé sur la conversion en format PDF peut sembler moins perturbatrice des pratiques en cours, et s'avère effectivement moins consommatrice de temps, le choix de l'un ou l'autre des formats impliquera **de toute façon une refonte des modalités de dépôt**.

Il faut par ailleurs considérer que le choix de l'un ou l'autre des formats peut ne pas être exclusif : ainsi dans bon nombre d'établissements ayant choisi le XML comme format de prédilection, le format PDF sert parfois de solution de repli ou de dernier recours, lorsque la conversion XML s'avère trop complexe, voire impraticable.

Partie 3 : Projets en cours dans les établissements de soutenance et scénarii envisageables à Cergy-Pontoise

Cette partie présente uniquement une sélection d'exemples, à partir desquels seront repris des éléments de scénario. Pour une approche moins fouillée, mais plus globale, voir l'annexe 4 (*Panorama de l'offre actuelle d'archives ouvertes pour les thèses*).

1. Présentation des projets extérieurs en cours

1.1. Contexte international : projets des universités américaines (Virginia Tech) et québécoises (Laval)

1.1.1. Le Networked Digital Library of Theses and Dissertations (NDLTD)

Les premiers projets de thèses en ligne sur Internet sont nés sur le continent nord-américain, notamment à l'université de Virginia Tech qui a engagé une réflexion sur le sujet dès 1987. Depuis 1997, les thèses y sont déposées uniquement sous format électronique. Le projet de Virginia Tech se prolonge dans le Networked Digital Library of Theses and Dissertations (NDLTD) (<http://www.ndltd.org/index.en.html> [dernière visite le 21/06/2006]). Ce projet de recherche, d'une durée de trois ans, a débuté en septembre 1996 et a été subventionné par le Département de l'éducation des États-Unis. Il avait pour but d'établir un réseau de thèses accessibles sous forme électronique. Il fédère aujourd'hui près de 200 projets de diffusion des thèses en texte intégral par Internet.

1.1.2. La collection de l'Université de Laval

L'Université de Laval a lancé en 1998 un projet de mise en ligne des thèses et mémoires de ses étudiants. La Collection a été lancée en octobre 2002 ; elle ne prend en considération que les thèses et mémoires rédigés et déposés sur support

électronique depuis 2002, et compte ainsi 256 documents consultables en texte intégral en format PDF (Portable Document Format) protégé et XHTML (eXtensible Hypertext Markup Language). Ces documents sont consultables sur le lien suivant : <http://www.theses.ulaval.ca/>

L'Université de Laval est la première à avoir développé l'aspect formation aux outils bureautiques et à l'utilisation d'une feuille de style, en présentiel et en ligne. La Bibliothèque de l'Université Laval est également à l'origine d'**un système personnalisé de dépôt institutionnel intitulé Archimède** (archimede.bibl.ulaval.ca). Inspiré du modèle DSpace, Archimède est orienté vers les communautés de recherche, et entièrement développé en « open source ». Le système est conforme à OAI (Open Archive Initiative), utilisant des métadonnées basées sur le Dublin Core, et permet de faire des recherches dans les métadonnées aussi bien que dans le texte intégral.

Une distribution « open source » de Archimède est également disponible⁵⁵, avec notamment une interface en français. Une nouvelle version (2.0) d'Archimède⁵⁶ vient d'être mise au point et présentée au colloque ETD 2006 en juin dernier.

1.1.3. Le dépôt institutionnel Papyrus de l'université de Montréal

Papyrus (<https://papyrus.bib.umontreal.ca/dspace/>) utilise l'application open-source Dspace⁵⁷, développée au MIT de concert avec Hewlett-Packard. C'est un logiciel qui permet d'entreposer, d'indexer, d'archiver à long terme, et de rendre accessibles, tous types de documents électroniques.

Il est conçu pour supporter une grande variété de formats, et permettre un dépôt quasi illimité, géré de manière ventilée, par collection et communauté. La notion de « communauté » désigne dans Papyrus une unité de recherche ou d'enseignement de l'université (un département, un centre de recherche ou une faculté). Pour chaque communauté ou sous-communauté sont définis : qui peut soumettre un document, quels sont les différents mécanismes de validation, ou quel type de contenu retrouver dans une collection donnée.

Dspace propose l'attribution d'une URL permanente à chaque document électronique, ainsi qu'une fonction « Documents vedettes » qui permet de connaître, chaque mois, les dix documents les plus consultés dans le dépôt

⁵⁵ <http://www.bibl.ulaval.ca/archimede/pages/telecharger.html>

⁵⁶ <http://archimede.bibl.ulaval.ca/archimede/pages/home/index.jsf>

⁵⁷ <http://dspace.org/>

institutionnel. Cette plate-forme fonctionne avec le logiciel de gestion de base de données PostgreSQL, qui permet de gérer des informations concernant l'organisation (communautés...), les métadonnées, l'annuaire des e-people et de pratiquer la maintenance et les sauvegardes. Son environnement logiciel comprend aussi le serveur Apache, Java, et Tomcat pour le moteur de servlet.

1.2. En France : les projets à l'échelle locale et nationale

1.2.1. Lyon2-Cyberdocs

1.2.1.1. *Origines et principes du programme*

Cyberthèses, issu d'une collaboration engagé en 1998 par les universités de Montréal et de Lyon II, soutenue par le Fonds francophone des inforoutes pour la publication électronique des thèses, est un programme visant à diffuser en ligne des thèses en s'appuyant sur la norme SGML. Le programme a développé pour ce faire une plate-forme logicielle (Cyberdocs) permettant de transformer un document stylé issu d'un traitement de texte en un document structuré en SGML, conforme à la DTD TEI Lite. Le programme consistait également à mettre en place un serveur commun aux différents établissements partenaires permettant l'indexation de leurs thèses. Celles-ci sont consultables sur le site de Cyberthèses : < <http://theses.univ-lyon2.fr/> >

A partir de 2004, une nouvelle version de la plate-forme logicielle, basée sur XML, a été lancée selon les mêmes principes : **l'appui sur des normes existantes, et le développement d'une plate-forme logicielle libre et open source.** Tout établissement peut ainsi télécharger les sources, installer la plate-forme sur ses propres ordinateurs, et les adapter à ses besoins. De la documentation régulièrement actualisée concernant le fonctionnement de Cyberdocs, le téléchargement de ses sources et l'installation de la plate-forme, sont disponibles sur le site de développement de Cyberthèses⁵⁸.

1.2.1.2. *Fonctionnement de Cyberdocs*

Cyberthèses utilise actuellement la plate-forme baptisée Cyberdocs, soit un ensemble de logiciels coordonnés par des pilotes : OpenOffice.org, une machine virtuelle Java, des parseurs XML, des processeurs XSL, des processeurs XSL-FO,

⁵⁸consultable sur le lien : <http://sourcesup.cru.fr/cyberdocs/>

un moteur de servlets (Tomcat), une servlet pour l'indexation et la recherche (SDX). Tous ces logiciels sont accessibles gratuitement sur Internet.

La plate-forme Cyberdocs se divise elle-même en trois modules : conversion, gestion et publication.

Le module de conversion

Il prend en entrée un fichier structuré⁵⁹ au format Microsoft Word, converti au préalable en un fichier OpenOffice, lequel par passage dans le module donne naissance à un fichier dit « pivot » en XML, formé selon la DTD TEI Lite. A partir du « fichier pivot » sont produits simultanément des fichiers pour la consultation (HTML, XHTML) et l'impression (PDF). La conversion est automatisée, et sa durée dépend de la taille de la thèse, et des performances de la machine (depuis moins d'une minute, la plupart des cas, à huit minutes pour une thèse comportant de nombreuses figures). Un fichier de métadonnées XML, essentiellement du Dublin Core (standard composé de quinze éléments qualifiés) doit être joint au fichier source.

Le module de gestion OGET

Ecrite en langage de script PHP, OGET est une **interface Web** conçue par le service ERAD en vue de **gérer les thèses aux différentes étapes du circuit suivi par la thèse**⁶⁰ : à la fois base de données centralisant les informations sur une même thèse, et entrepôt de documents primaires, OGET permet d'en gérer la diffusion et de faire le point sur l'état d'avancement des travaux (stylage, conversion...) à destination de tous les acteurs de la conversion.

Chaque thèse se voit attribuer un espace unique dans OGET, où sont saisies toutes les informations la concernant, mais aussi les documents sources et les documents issus du stylage, permettant ainsi un archivage des copies électroniques des thèses, avec leurs métadonnées.

De plus, OGET permet également un suivi statistique et une identification rapide de thèses, ou groupes de thèses, grâce à un moteur de recherche multicritères,

⁵⁹ Pour obtenir ce résultat, le document source doit être structuré au préalable à l'aide d'une feuille de style. Les éléments de structure présents dans un document traitement de texte peuvent être : de l'enrichissement typographique (caractères gras ou italiques, couleurs, polices particulières, etc.) ; des structures déjà gérées par les traitements de texte (tableaux, illustrations, liens hypertextes, notes de bas de page, etc.) ; des styles (de paragraphes ou de caractères) ayant une sémantique commune à tous les documents (Titre 1, Titre 2, etc.) ; des styles ayant une sémantique particulière pour la plate-forme Cyberdocs (auteur, titre du document, année de publication, etc., qui génèrent également des métadonnées).

⁶⁰ Pour une présentation complète des objectifs et fonctionnalités d'OGET, voir l'article « Gérer et diffuser des thèses électroniques : un choix politique pour un enjeu scientifique » in Documentaliste – Sciences de l'Information 2005, vol. 42, n° 4-5, p. 280-282.

alimenté par les multiples informations saisies aux différentes étapes du traitement⁶¹.

Le module de consultation

C'est une application Web dynamique, bâtie sur la plate-forme SDX, qui permet à l'administrateur d'ajouter et de supprimer les documents de la base documentaire. Il permet également à l'utilisateur d'effectuer des recherches simples ou complexes. Ainsi, le site permet la consultation en plein texte de thèses de doctorat soutenues à l'Université de Lyon 2, et propose une sélection à partir de la liste des titres, ou bien par mots dans un formulaire de requête, pour retrouver l'intégralité ou les parties des documents répondant aux critères de recherche. L'utilisateur accède au chapitre des documents correspondant à sa requête, et les termes de la recherche sont mis en évidence lors de la consultation. L'espace de consultation permet également d'accéder à la version imprimable du document (en PDF notamment).

Les évolutions envisagées

Actuellement, Cyberdocs ne permet pas de transformer en XML les documents LaTeX de manière entièrement satisfaisante: ceux-ci sont donc convertis au préalable en format PDF pour entrer dans la chaîne de traitement (gestion et publication). De nouveaux développements de la plate-forme sont prévus dans les mois à venir, à savoir d'une part l'intégration des documents natifs LateX, qui pourront être convertis en documents XML de type Docbook ou TEILite, et d'autre part l'intégration des documents structurés sur le modèle de la DTD Docbook.

1.2.2. CCSD : TEL Thèses-en-lignes

Les systèmes documentaires comportant des thèses peuvent être couplés à des systèmes moins spécifiques accueillant parallèlement des articles scientifiques, mémoires, cours, archives ouvertes, etc. (voir l'annexe 4).

Le CCSD, Centre pour la Communication Scientifique Directe (CNRS), gère ainsi le projet « Thèses-En-Ligne », TEL.

Développé initialement à partir du logiciel de gestion éditoriale Eprints, dont il reprend les principes de fonctionnement, le système TEL du CCSD est dorénavant intégré dans le système générique HAL (Hyper articles en ligne). Ce système propose l'autoarchivage en ligne des thèses afin d'en faciliter l'accès.

Le dépôt

⁶¹ La possibilité de créer une version « générique » d'OGET, est envisageable, c'est-à-dire qui soit elle-même transportable dans des contextes différents de celui de l'université Lyon Lumière 2, et utilisable par une institution différente avec sa propre interface

Le dépôt dans TEL suppose l'emploi d'un format visualisable, PDF, ou éventuellement PS (postscript), en un seul fichier. Seul ce fichier « visualisable » est nécessaire, mais les documents source dans leur format d'origine (LateX, HTML, etc.) ou compressés (.tar,gz, zip...) peuvent également être déposés sur la plate-forme. Pour les dépôts en format DOC, RTF ou PS, TEL se charge automatiquement de créer un format PDF.

L'auteur, après s'être enregistré, peut librement déposer ses travaux. Le dépôt peut également être effectué par une personne du laboratoire (un bibliothécaire par exemple) ayant l'accord de l'auteur pour le faire. Le document est ensuite vérifié par le CCSD (contrôle sur les informations administratives et bibliographiques fournies). Par contre, aucune évaluation scientifique n'est effectuée: il est ainsi précisé qu' « il est demandé au déposant de télécharger la version qui contient compte de toutes les corrections demandées par le jury ; le cas échéant, si ces corrections sont absentes du premier dépôt, une seconde version doit impérativement être déposée pour remédier au problème. »⁶²

La diffusion

Les documents sont mis à disposition gratuite en ligne et le plus souvent, la thèse est disponible en ligne dans les 24 heures suivant le dépôt.

Outre la recherche (simple et avancée), l'interface standard de TEL permet une consultation par domaine scientifique, par année de soutenance, mais également par établissement, ou par collection (voir le paragraphe « vues et tampons »).

Une caractéristique de TEL consiste en effet dans la définition formelle et précise des relations entre auteurs et affiliations, qui doivent être correctement déclarées dans la procédure de dépôt. Ce dispositif facilite les vues institutionnelles⁶³.

Grâce à ce dispositif, HAL peut notamment fournir des services de nature plus administrative, comme la constitution de listes de publication (par auteur, laboratoire, institution, etc.) dans des formats multiples⁶⁴. Ceci illustre l'objectif premier du CCSD : une archive ouverte conçue pour faciliter le pilotage de la recherche.

⁶² Voir les informations sur le dépôt des thèses : http://www.ccsd.cnrs.fr/rubrique.php3?id_rubrique=14 et plus généralement le guide du dépôt et du bon usage de HAL : http://www.ccsd.cnrs.fr/rubrique.php3?id_rubrique=15

L'accent a été particulièrement mis sur une structuration solide de métadonnées, comme par exemple l'association auteurs-laboratoires-institutions : tout auteur doit être rattaché à un laboratoire (ou organisme de délivrance), lui-même affilié à une institution. L'utilisation de listes à valeurs fixes pour des informations comme les titres des journaux est imposée et une liste de 2000 laboratoires français a été incorporée dans Hal depuis Labintel et d'autres sources, afin d'assurer l'homogénéité des informations correspondantes (en particulier pour les affiliations).

⁶⁴ Pour que ces listes puissent être complétées dans le cas où le fichier du plein texte n'est pas disponible, il est possible de déposer sur HAL des « notices bibliographiques » ne contenant que des références de publication.

TEL peut être moissonnée via le protocole OAI-PMH avec deux formats de métadonnées : du Dublin Core non qualifié et un schéma spécifique au CCSD.

Les documents déposés dans HAL/TEL bénéficient de la pérennité de cette base, y restent de façon permanente et reçoivent une adresse web stable; ainsi, comme une publication dans un journal scientifique habituel, ils peuvent faire l'objet d'une citation dans un autre travail. Le projet TEL diffuse actuellement 2488 thèses (fin juin 2006) et le dépôt de thèses anciennes est également fortement encouragé (certaines datent ainsi de 1986).

<http://tel.ccsd.cnrs.fr/>

Le contrôle et l'usage des collections : vues et tampons

HAL propose une modalité de certification/ validation par l'apposition de « tampons électroniques », sans modification du document d'origine. L'opération se fait sous l'unique responsabilité du « tamponneur », qui détermine une « valeur » pour le tampon. Par contre, cette validation ne se conçoit que comme une 'couche supplémentaire', les documents déposés par d'autres ne pouvant pas être retirés par le tamponneur (ce serait contraire aux principes généraux de fonctionnement de la base et à ceux de OAI). HAL met à jour automatiquement des extractions automatiques contenant tous les documents (et notices) ayant recueilli un même tampon, qu'on appelle « vues par tampons ». Elles sont personnalisables par l'insertion d'un logo, des choix graphiques, etc. Un responsable de laboratoire ou d'établissement peut ainsi demander la création d'un tampon avec un nom institutionnel, afin de construire facilement une liste visible d'articles reconnus officiellement, en mettant un lien vers cette liste depuis le site internet officiel de son unité.

Il n'y a pas de limitation sur le nombre de tampons de HAL, ni sur le nombre de tampons distincts qu'un document donné peut recevoir. Le tamponnage se fait pour chaque document, version par version.

1.2.3. Thèses de l'Université Louis Pasteur : archive ouverte de l'ULP (E-print)

L'université de Strasbourg Louis Pasteur a choisi de créer une bibliothèque virtuelle en février 2005, dans le cadre du projet EPPUN (Espaces pédagogiques pour les universités numériques). Un nouveau portail documentaire a été constitué grâce à l'acquisition du logiciel MetaLib (Société Ex-Libris) et l'installation d'un

serveur de liens SFX. Une subdivision de cette bibliothèque numérique, consacrée aux thèses en ligne, a été créée avec le logiciel d'archives ouvertes GNU EPrints comme support de gestion et de diffusion, tandis que Cyberdocs est employé en amont comme module de conversion (<http://eprints-scd-ulp.u-strasbg.fr:8080/>).

Fonctionnalités et caractéristiques du logiciel Eprint

GNU Eprints est un logiciel serveur permettant de mettre en ligne tous types de documents. Ce logiciel permet de soumettre, explorer et gérer des documents, le tout par une interface web, et propose deux fonctionnalités essentielles :

- La possibilité pour les chercheurs de mettre en ligne leurs travaux sans intermédiaire (auto-archivage) ; ceci implique aussi la possibilité de gérer les étapes de révision d'un document et donc de maintenir sur un même serveur plusieurs documents retraçant plusieurs étapes d'élaboration d'un document scientifique.
- La mise en oeuvre d'un protocole sophistiqué (non-propritaire) de description et référencement des documents.

Trois autres fonctionnalités peuvent également être mentionnées :

- Eprints permet d'attacher plusieurs fichiers de formats variés (PDF, Word, HTML, etc.) à une fiche descriptive et de gérer les droits d'accès pour chaque fichier.
- Il permet aux auteurs de gérer le niveau de publicité de leur production (gestion à trois niveaux : public, utilisateurs enregistrés sur l'archive et administrateur)
- Il permet aux utilisateurs enregistrés de publier des commentaires sur les documents publiés.

Les fonctions de recherche dans la base sont :

- formulaire de recherche basique (tous champs confondus)
- formulaire de recherche avancée sur treize champs séparés, avec correspondances sur un des mots, tous les mots ou phrase exacte. Après adaptation à l'ULP, la recherche avancée propose également une recherche « plein texte », par type de document, par établissement et par composante.

En outre sont mis en valeur les "derniers ajouts" (dernière semaine). L'ensemble est complété par un système d'alerte proposé aux utilisateurs enregistrés, leur permettant d'obtenir, à une fréquence qui leur convient, la liste des dernières publications dans le domaine qui les intéresse.

Les résultats de la recherche proposent le téléchargement du document en PDF ainsi que le résumé complet et des et des métadonnées sur le document.

GNU EPrints⁶⁵ permet de créer des archives supportant le protocole OAI rapidement et facilement (les archives ainsi créées sont conformes au protocole OAI, mais doivent être déclarées par le fournisseur pour être accessible).

La configuration requise est la suivante :

- UNIX ou Linux
- Langage Perl
- Serveur APACHE avec module Perl
- Base de données Mysql

Distribué sous licence GNU/GPL, EPrints est librement téléchargeable et gratuitement utilisable⁶⁶.

2. Axes de réflexion, tâches communes aux différents scénarii d'un circuit électronique des thèses

Dans les parties précédentes ont été évoquées les contraintes à la fois juridiques (obligations réglementaires et droit d'auteur), organisationnelles (notion de workflow, et répartition des tâches dans le circuit), techniques (formats et outils). Un certain nombre d'enjeux de type stratégique, organisationnel, puis fonctionnel ont déjà été suggérés, mais doivent être détaillés avec la formulation des différents scénarii, et ce, avant tout choix d'une solution technique.

2.1. Réflexions préalables aux différents scénarii : les enjeux communs

Chaque définition d'objectif ou d'enjeu ouvre une alternative, elle-même déclinable en plusieurs options ; il est donc nécessaire de définir ce sur quoi l'UCP entend mettre l'accent.

Ainsi, **quels objectifs principaux vise-t-on au travers de cette refonte ?** S'agit-il :

- de **faciliter le signalement des thèses et le fonctionnement administratif**

⁶⁵ Site général Eprints : <http://www.eprints.org/>

⁶⁶ Son développement technique est assuré par deux personnes au sein de l'université de Southampton (Christopher Gutteridge, assisté de Mike Jewell). Le support technique est assuré à travers une liste de discussion Eprints-tech. Pour des détails sur l'aspect technique : GNU Eprints : <http://software.eprints.org/> ; Documentation technique : <http://software.eprints.org/documentation.php> ; Archives de listes de discussion : <http://software.eprints.org/maillist.php>

- d'améliorer **la notoriété et la visibilité scientifique** de l'UCP ? Auquel cas l'accent sera plutôt placé sur la qualité de la diffusion...

Si l'accent est effectivement mis sur une diffusion large et rapide, encore faut-il **définir les priorités en terme d'image pour l'établissement** :

1. **quel est le destinataire privilégié** : les laboratoires en interne, la communauté scientifique internationale, le lecteur qui se déplaçait en salle pour consulter les microfiches...
2. **selon quelle architecture** : adaptation d'une plate-forme de diffusion en interne avec forte empreinte de l'université (proximité institutionnelle, prestige...) ou délégation à l'extérieur (structure moins lourde, mais identification moindre, et évolution et gestion du serveur éventuellement moins souple).
3. souhaite-t-on favoriser **une mise en ligne rapide et exhaustive** des thèses, **ou prévoir un traitement plus complexe pour ménager un accès de qualité à des documents structurés** ?
4. en conséquence, quels sont les **services associés attendus** : recherche textuelle ou pas (renvoie aux principes de restitution, et donc d'indexation...), impression ou pas, téléchargement ou pas...⁶⁷

Il s'agit en outre de **délimiter l'ampleur et la stratégie de développement** que l'on souhaite donner au projet : une refonte du système s'avère nécessaire, mais une révision globale du système d'information institutionnel est-il envisageable à la faveur de ce projet ?

De même, le corpus initial concerne les thèses de Cergy-Pontoise, mais ce projet ne peut-il s'étendre à d'autres documents scientifiques de nature différente (littérature grise par exemple). Souhaite-t-on traiter la question des thèses, ou anticiper (comme le font déjà un certain nombre d'universités) sur une généralisation de la pratique de l'archivage et de l'édition numériques au sein des établissements ?

Tous ces éléments (objectifs, ampleur, nature...) doivent être enfin mis en regard des moyens humains disponibles et mobilisables à court, moyen et long termes.

2.2. Les étapes et éléments communs aux scénarii

⁶⁷ A la faveur de ce changement d'architecture du circuit, d'autres services pourraient d'ailleurs être envisagés, comme la reproduction des thèses papier destinés à la bibliothèque par le service de reprographie de l'université. Il s'agit d'une pratique de plus en plus répandue dans les universités, et c'est par exemple déjà le cas à Lyon 2.

Que l'on choisisse l'une ou l'autre des chaînes de traitement existantes, les seuls points de divergences apparaissent au moment de la conversion (si elle a lieu), et peuvent se répercuter sur les modes d'archivage et de diffusion. Un certain nombre d'éléments seront identiques, quelle que soit la solution qui sera finalement retenue, et notamment, quel que soit le mode de diffusion privilégié. Ainsi :

- des **actions d'information et de promotion** sont à envisager à destination des étudiants et des personnels en vue de les préparer aux modifications à venir. Des outils doivent être élaborés : feuillets, rubriques sur internet, présentant les enjeux et bénéfices de la publication électronique des thèses, et présentation détaillée du nouveau circuit des documents.
- que l'on choisisse une conversion en PDF ou en XML, il faut songer à mettre à disposition de feuilles de style,
- des **outils de formation** doivent également être mis à la disposition des **doctorants et surtout des personnels engagés dans le nouveau circuit**. Les formations à destination des seuls doctorants, qui gagneraient à intégrer un volet juridique, n'ont pas été prises ne compte dans l'évaluation du coût de chaque scénario, dans la mesure où elles sont déjà assurées et assumées financièrement par les Ecoles doctorales.
- le circuit de dépôt des documents électronique, et le traitement documentaire des documents (indexation, signalement) sont également des constantes quelle que soit l'architecture retenue ;
- tous les modes de diffusion envisagés sont compatibles avec OAI, et assureront donc **toutes** une visibilité internationale aux thèses.

Par contre, des points de divergences vont apparaître si l'on choisit :

1. une chaîne de conversion, laquelle implique :
 - l'intervention d'un technicien informatique pour installer la chaîne de conversion, et éventuellement assurer un suivi,
 - un environnement informatique spécifique : Linux ou Unix (alors que le reste de l'université tourne sur Windows),
 - pour la bibliothèque, une vérification plus approfondie des fichiers (conformité de la structuration des données fournies par le doctorant, et non plus seulement lisibilité du fichier),
 - des sessions de formation spécifiques, en vue d'une utilisation plus avancée des logiciels de traitement de texte ;

2. une plateforme de diffusion intégrée (et non déléguée à l'extérieur), qui impliquerait :
 - l'emploi d'un serveur, et une personne chargée de l'administrer (soit au sein de la bibliothèque, ou au niveau de l'université),
 - et des développements techniques (pour implanter par exemple Dspace ou EPrints).

3. Trois scénarii possibles pour Cergy-Pontoise

3.1. Implanter Cyberthèses : un projet intégré fonctionnant sur une logique de proximité

Choisir le programme Cyberthèses revient à se doter d'un système complètement intégré, et modulable, qui comporte :

- Un système de workflow (OGET) englobant toutes les étapes depuis le dépôt jusqu'à l'archivage
- une **plate-forme de conversion** vers le langage XML (**Cyberthèses**) accessible en ligne et libre de droits
- une **plateforme de diffusion (SDX)**

En outre, la cellule Lyon 2 ERAD propose des formations⁶⁸ pour le personnel chargé de la conversion et met à disposition un serveur de thèses hébergé à l'Université Lyon III – Jean Moulin.

Avantages de ce scénario :

- Proximité institutionnelle avec les différents acteurs et réactivité (par exemple : Rapidité de mise en ligne ou de retrait, en cas de demande du doctorant)
- Rapidité de mise en œuvre à l'échelle de l'établissement
- Excellente visibilité/ identification institutionnelle pour la diffusion : les thèses de l'UCP
- Production de documents structurés (XML) : meilleures performances de recherche pour l'utilisateur et pérennité de l'archivage pour l'institution
- Archivage local suppose maîtrise de l'université sur sa production
- Intégration possible dans un autre programme (voir l'exemple de l'ULP, qui utilise Cyberthèses dans le cadre d'une archive ouverte).

Ses inconvénients :

- Forte implication humaine nécessaire : technicien informatique, styleur, administrateur de base ; nécessité de doter cette nouvelle équipe des compétences nécessaires ;
- Un projet limité dans un premier temps en terme de corpus, d'ampleur du public visé et de collaboration avec d'autres instances de l'université ;
- La question encore délicate de la compatibilité avec STAR : pas de passerelle encore entre les métadonnées collectées dans OGET et celles de STAR (modèle TEF), mais en développement⁶⁹ ;
- Archivage local, qui suppose un serveur / de l'espace serveur, avec un usage finalement limité par rapport aux potentialités.

3.1.1. Organisation

Dans ce scénario, le dépôt électronique de la thèse est effectué, à l'instar du format papier, dans les secrétariats de chaque école doctorale. Puis, la Mission formation et thèses se charge de collecter le document au format électronique et de le convertir en XML. La plate-forme de conversion en XML est téléchargée sur 2 postes⁷⁰.

Le travail de reprise probable sur une thèse (entre quelques heures et quelques jours et une semaine) est à mettre en regard de la production annuelle de l'Université (environ une trentaine de thèses). La conversion en XML par une seule personne risque d'être difficile, aussi le recrutement d'un moniteur à temps plein, pour assister le responsable, semble adéquat. La personne chargée de styler devra être capable de manipuler des logiciels de type Word, des interfaces web, et d'alimenter des bases de données.

La formation du personnel ne devrait pas excéder 2 jours, et peut être assurée en interne, puisque la familiarisation avec le système est déjà effective pour l'une des deux bibliothécaires impliquées dans le projet. L'une ou l'autre (responsable de mission ou pas) devra recevoir une formation plus approfondie lui permettant d'assurer la veille technologique du service.

⁶⁸ dont j'ai justement pu bénéficier, dans le cadre de ma formation à l'enssib et en vue de ce projet.

⁶⁹ OGET et STAR ne répondent pas à la même définition du dépôt. Le premier assure un dépôt initial du matériel informatique ; le second assure un dépôt post conversion. Par ailleurs, OGET, développé en Open source, s'installe localement, tandis que STAR est une application 'client léger'. Par contre, tous deux présentent des fonctions redondantes, telles que l'export vers les dispositifs de signalement, diffusion et conservation.

⁷⁰ Idéalement, ces postes devraient être à proximité, et supporter une gamme variée de versions de logiciels de traitement de texte ou autre (ex : LaTeX et MathType) afin d'être en accord avec les usages des doctorants.

3.1.2. Coûts prévisibles

Sur le plan matériel :

Il faut « un espace disque d'environ 200 Mo pour une installation complète, incluant OpenOffice.org, PHP, SDX, et java. » précise selon la documentation en ligne de Cyberdocs⁷¹. Selon les options choisies, la place occupée par le document final peut être plus ou moins importante : pour évaluer la mémoire vive du serveur, il faut compter 50 Mo par document augmenté de 1,25 Mo par page.

La chaîne sera installée sur 3 postes à la bibliothèque, qui devront être équipés d'un navigateur Web récent supportant les normes HTML, CSS et javascript.

Achat d'un serveur dédié : 5000 €

Equipe :

Dans le cas du recrutement d'une personne pour le stylage, en catégorie C à temps complet (en tablant sur des traitements d'une demi-journée par thèse, en moyenne) ou moniteur étudiant, le coût serait d'environ 24 000 € par an.

Installation et paramétrage de la plateforme pour permettre une diffusion Internet : 4 jours technicien x 10 €/h : 400 €/par technicien

Coût d'investissement total : 30 000 €

Concernant le budget de fonctionnement prévisionnel pour la première année, le calcul est difficile, dans la mesure où la formation des doctorants est pour l'instant assurée par les Ecoles doctorales (peut être maintenu ainsi, pour peu que des modifications soient intégrées par les personnes en charge de ces formations); à ce stade donc, coût non comptabilisé pour la bibliothèque. Quant au coût engendré par le personnel du service, il correspond à 1/3 ETP de la responsable (Cat A, déjà en place) et pour la bibliothécaire (déjà en place) un 1/2 temps (pour l'élaboration et la distribution des documents dans une première phase) à 1/3 temps à terme

3.2. Vers un réseau national de thèses électroniques : participation de l'UCP à TEL

La plate-forme TEL-CCSD part d'une logique de centralisation des données à l'échelle nationale, de documents de même nature, en l'occurrence les thèses. La soumission se fait directement en ligne, sans installation de logiciel sur des machines, selon une logique d'autoarchivage, et la gestion est ensuite assurée par le CCSD lui-même. L'outil de diffusion est conçu être facilement paramétré par un établissement donné (logo, couleurs, requêtes...). Tout comme le système STAR,

⁷¹ Voir le lien : <http://sourcesup.cru.fr/cyberdocs/fr/installation/materiel.html>

TEL ne comprend pas de plate-forme de conversion : sa vocation est de conserver et diffuser les fichiers tels que déposés avec leurs métadonnées associées.

Il est à noter que la proximité avec STAR est manifeste : la FAQ de STAR ne cite nommément que TEL et la future plate-forme de diffusion du CINES comme débouchés de diffusion naturels ; les « autres solutions » envisageables sont laissées à l'appréciation de chaque établissement. Par ailleurs, les conditions à moyen terme d'une exportation automatisée des thèses vers le système HAL sont à l'étude, avantage non négligeable en terme de gestion quotidienne.

En outre, dans la mesure où TEL permet la récupération de collections entières déposées sur son archive, cela permet d'envisager ce choix comme une solution de diffusion transitoire, à court terme, avant éventuel passage à une autre plateforme (ou un autre système d'information au sein de la bibliothèque), et ce sans déperdition excessive d'énergie.

Avantages de ce scénario :

- Coût humain beaucoup plus faible (quasi externalisation)
- Traitement beaucoup plus léger des documents convertis en PDF (moindre technicité)
- Extrême rapidité de la mise en ligne (24 heures)
- Structure de l'archive mettant l'accent sur des vues institutionnelles
- Archive pluridisciplinaire mais spécialisée dans les thèses
- Insertion dans une logique nationale
- La complémentarité avec STAR semble assurée
- Bien que ce ne soit pas sa vocation, TEL permet de récupérer toute une collection si l'on souhaite opérer des traitements différents.

Ses inconvénients :

- Moindre maîtrise de la gestion par l'université
- Visibilité assurée, mais lien identitaire avec l'université également moindre
- Fichiers en PDF : performance moindre en terme de recherches textuelles et de pérennité des données
- Les problèmes liés à l'autoarchivage : multiplicité de versions, exhaustivité éventuellement difficile à atteindre.

3.2.1. Organisation

Avec ce scénario doit être pris en compte la spécificité du système proposé par TEL(fondé sur l'autoarchivage) : la gestion des processus de validation/

révision n'est pas totalement performante, faute d'un réel workflow intégré depuis le dépôt, et le système procède plus d'une modération qu'un contrôle. Or, la valeur administrative de la thèse impose à la fois que le dépôt atteigne l'exhaustivité, et qu'il s'agisse ensuite de la version canonique du document. Or, la multiplicité des versions accessibles d'une même thèse (établissement de soutenance, laboratoire d'accueil et école doctorale sont susceptibles de diffuser le même travail dans des versions différentes) peut créer une confusion pour un usager en recherche d'information. C'est ce qu'entendent contrebalancer :

- d'une part le système des « tampons » (voir p. 51)
- d'autre part le système STAR, qui collecte la version finale de la thèse

Il paraît donc nécessaire que le dépôt soit plutôt opéré par un bibliothécaire. Ainsi, le dépôt électronique de la thèse serait effectué, là encore, dans les secrétariats de chaque école doctorale. C'est ensuite la Mission formation et thèses qui prendrait en charge la collection du document, son éventuelle transformation en PDF, et son dépôt sur TEL.

Aucun personnel supplémentaire, ni formation particulière, n'est à prévoir dans ce cas.

3.2.2. Coûts prévisibles

Les coûts à prévoir dans ce scénario sont particulièrement minimes :

Pas d'investissement matériel à prévoir, ni en terme de personnel, les effectifs de la Mission formation et thèses devant suffire à mener à bien les processus.

La définition d'une feuille de style et la formation à son utilisation, même dans le cas de fichiers PDF, est toujours à prévoir ; cependant, la remarque concernant la prise en charge actuelle par les Ecoles doctorales s'applique ici aussi.

Le seul coût éventuellement envisageable concerne le temps passé pour le paramétrage initial des requêtes et de la vue institutionnelle Cergy dans TEL (un à deux jours). Mais le CCSD propose d'apporter un soutien technique sur ce point.

3.3. Vers une bibliothèque numérique : implantation d'une plate-forme ouverte

Dans ce scénario, inspiré notamment de l'exemple de l'ULP⁷², le corpus des thèses sera la première pierre d'une bibliothèque électronique qui rassemblera à terme les

⁷² Certes, la mise à disposition d'une archive ouverte à destination de l'ensemble de la communauté universitaire, et au-delà du grand public, présente en soi un intérêt certain. Mais l'ULP n'est pas la seule dans ce cas, puisqu'on peut citer également l'Archive ouverte de Lyon 2, celle de l'Université l'ambition de sa bibliothèque numérique pédagogique va bien au-delà de la simple conservation et diffusion des de Nanterre (toutes deux également fondées sur Eprints) ou

documents existants dans les différentes composantes de l'université de Cergy-Pontoise. La mise en ligne des thèses mobilisera de nombreux services universitaires. Elle apportera donc au SCD de nouvelles collaborations et une meilleure visibilité de ses actions au sein de la communauté universitaire.

Avantages de ce scénario :

- Proximité institutionnelle avec les différents acteurs et réactivité
- Excellente visibilité/ identification institutionnelle pour la diffusion, et vitrine prestigieuse pour l'établissement
- Compatibilité avec tous les formats envisageables
- Archivage local suppose maîtrise de l'université sur toute sa production
- Un projet de forte ampleur à terme : ampleur des corpus, et des publics visés (laboratoires...)

Ses inconvénients :

- Mise en œuvre nécessairement lente à l'échelle de l'établissement
- Forte technicité nécessaire pour la mise en place;
- Effet doublon possible des métadonnées de gestion de l'archive avec les métadonnées TEF nécessaires dans STAR ;
- Communication beaucoup plus lourde à prévoir au démarrage ;
- Archivage local, qui suppose un serveur / de l'espace serveur, avec un coût de maintenance non négligeable.

3.3.1. Organisation

Pour la partie concernant les thèses, le processus centralisé décrit pour TEL (soumission et validation par le bibliothécaire) semble le plus adéquat. Pour d'autres types de documents et avec l'ouverture à d'autres instances dans l'université, un fonctionnement de type décentralisé pourrait être envisagé.

Sur les aspects organisationnels par exemple de Dspace, voir le document ci-après établi au sein de la MSH-Alpes : https://dspace.msh-alpes.prd.fr/bitstream/1801/439/1/Dspace_Novembre2004.ppt . Coûts prévisibles

3.3.2. Coûts prévisibles

Les coûts à prévoir dans le cas de l'implantation d'une archive ouverte dans l'université à partir de la bibliothèque sont extrêmement variables.

encore ALADIN de la MSH-Alpes (Dspace). Ce qui distingue à ce jour l'ULP, c'est que cette archive est particulièrement intégrée et mise en valeur au sein d'un véritable système d'information, les différentes « briques documentaires » étant fédérées par un portail.

D'après un document du MIT intitulé « Business planning for a Dspace service »⁷³, leurs dépenses annuelles pour la création (c'est-à-dire aussi le développement ex nihilo, comprenant aussi les accroissements de volume des serveurs, etc) d'une archive Dspace était de 285 000 \$.

Cependant, le document précise aussi que ce coût peut être minimisé dans le cas d'un démarrage modeste⁷⁴ : le logiciel est gratuit et en open; matériellement, un PC suffisamment puissant suffit, et une équipe de bibliothécaires peut s'en charger, à condition qu'une personne à plein temps et qu'un technicien à mi-temps soient mobilisés. De plus, les développements peuvent être mutualisés avec d'autres établissements.

Sur le plan humain, outre l'équipe déjà constituée au sein de la Mission thèses, un technicien / administrateur du système serait nécessaire à temps plein dans les phases initiales de développement, puis à mi-temps dans la suite du processus, et ceci même en tablant sur une implantation modeste. Sur le plan matériel, un serveur serait a minima nécessaire (soit là encore 5000 €).

3.4. Bilan des trois scénarii

Le scénario le plus rapidement exécutable, et qui représente également un coût particulièrement minime, correspond à un partenariat avec l'archive TEL du CCSD. Dans une logique de mise en production de STAR, c'est la solution apparemment la plus pertinente.

Cependant, c'est aussi une solution dont l'évolutivité est moindre, et qui risque de nécessiter une remise à plat dans quelques années.

A contrario, ce scénario peut aussi être envisagé comme un choix transitoire, en attendant de développer une autre solution sur d'autres plate-formes, en fonction des tendances qui pourraient se dégager à l'avenir.

Les deux autres solutions, qui reposent toutes deux sur une logique d'open access, supposent quant à elles des traitements plus lourds, et donc plus coûteux, mais ce sont aussi celles qui offrent le plus de garantie de pérennité et d'évolutivité.

Il faut distinguer à ce propos le programme Cyberthèses, conçu au sein d'une université, et qui peut être mis en place de manière assez rapide et avec un coût raisonnable, d'une solution d'archive ouverte/ de dépôt institutionnel, qui implique

⁷³ Voir le fichier sur le lien : <http://www.nercomp.org/data/media/walker.ppt> Dernière consultation le 12 octobre 2006.

⁷⁴ Pour l'implémentation de Dspace par exemple, voir : <http://dspace.org/implement/index.html>. Pour EPrints, voir <http://www.eprints.org/documentation/tech/eprints-docs.pdf>.

des délais d'implantation plus longs, et donc des coûts ad hoc, mais offre aussi une gamme de services quasi illimitée.

Quoi qu'il en soit, le critère de coût ne peut être seul déterminant, dans la mesure où l'évolutivité des systèmes, et les choix politiques de l'établissement, en terme d'image notamment, sont eux-mêmes des composantes majeures du choix (cf. la partie « Réflexions préalables au scénarii »).

Le tableau ci-après propose une vue synthétique des forces de chaque système en fonction de différents critères (les critères sont notés de 1 à 3, la meilleure solution se voit attribuer 3, la moins pertinente 1) :

Critères	Scénario 1 : Cyberthèses	Scénario 2 : TEL	Scénario 3 : archive institutionnelle
Faisabilité	2	3	1
Innovation	2	1	3
Evolutivité	2	1	3
Coûts	2	3	1
Facilité de mise en œuvre	2	3	1
Impact sur l'organisation	2	1	3
Délais de mise en œuvre	2	3	1
TOTAL	14	15	13

Voir aussi le tableau synthétique des avantages et inconvénients de chaque scénario en annexe 5.

Conclusion

La mise en ligne des thèses au sein de l'université est un projet qui concerne l'ensemble des acteurs de l'établissement, et suppose donc, pour aboutir, que les différents services concernés se mobilisent et se coordonnent. En effet, le nouveau circuit des thèses implique à la fois une formalisation des relations avec les écoles doctorales, et **un circuit documentaire clairement balisé**, depuis l'avis de soutenance jusqu'à l'archivage, en passant par la répartition des documents imprimés auprès des services concernés (doctorant, bibliothèque...). Les avantages sur le plan institutionnel sont bien clairs pour l'université, tant sur le plan de la dynamique interne de ses composantes, qu'en terme de services rendus à ses usagers. Les choix techniques ne font donc, dans ce cas, que parachever et faciliter des choix avant tout politiques et stratégiques.

Dans ce nouveau circuit, la bibliothèque universitaire peut endosser une position pivot, et ainsi franchir le pas vers un nouveau paradigme (l'édition électronique), tout en demeurant dans le cadre de ses missions traditionnelles (acquérir, conserver, signaler, diffuser).

Au-delà des compromis nécessaires en vue de la mise en place d'un circuit performant, et de l'interactivité entre différents métiers (administratifs, bibliothécaires, techniciens, chercheurs), un élément central qui sous-tend cette étude est par ailleurs la notion d'évaluation. Une meilleure visibilité des différents maillons de la chaîne, tout comme la mise en valeur de la qualité de la recherche sont des questions actuellement fondamentales.

Les thèses ne sont qu'une expression parmi d'autres de la vitalité d'une université. Mais elles permettent déjà de penser l'articulation entre des considérations locales, un contexte national, et une échelle internationale, de plus en plus prégnante.

Bibliographie

SUR LES DOCUMENTS NUMÉRIQUES ET LA DIFFUSION DES THÈSES ÉLECTRONIQUES

Ouvrages et contributions généraux

BEAUDRY, Guylaine, BOISMENU Gérard. « Conception d'un portail de production, de diffusion et de gestion de publications électronique : Étude de faisabilité ». 2000.

[en ligne] <<http://www.erudit.org/documentation/etude/accueil.html>> Consulté le 17 juillet 2006

BOULETREAU, Viviane ; DUCASSE, Jean-Paul, « La diffusion électronique des thèses : un enjeu politique et non pas technique ». In *Comprendre les usages de l'Internet*. GUICHARD, Eric (dir.). Paris, éd. rue d'Ulm, 2001. P. 118-122.

FOX, Edward A.; MCMILLAN, Gail ; SULEMAN, Hussein; GONÇALVES, Marcos A. ; LUO, Ming. « Networked Digital Library of Theses and Dissertations (NDLTD) ». In ANDREWS, Judith (ed.) and LAW, Derek (ed.) *Digital Libraries: Policy, Planning and Practice*. Aldershot : Ashgate, 2004. p. 167-188.

JACQUESSON, Alain ; RIVIER, Alexis. *Bibliothèques et documents numériques: concepts, composantes, techniques et enjeux*. Paris : Éd. du Cercle de la Librairie, 2005.

MCMILLAN, Gail. « Implementing ETD Services in the Library ». In *Electronic Theses and Dissertations: A Sourcebook for Educators, Students, and Librarians*. Edward A. Fox, Joseph M. Moxley, Christian R. Weisser, editors. New York: Marcel Dekker, 2004. p. 319-329.

Articles

BAILEY JR., Charles W. *Electronic Theses and Dissertations: A Bibliography*. 2005. [en ligne] <<http://www.escholarlypub.com/digitalkoans/2005/07/08/electronic-theses-and-dissertations-a-bibliography/>> Consulté le 17 mars 2006.

DEBOIN, Marie-Claude. « Publier et diffuser sa thèse électronique sur Internet : conseils aux doctorants accueillis par le Cirad » ; In *CCSD @rchiveSIC* [Archive Ouverte en Sciences de l'Information et de la Communication] [en ligne] < http://archivesic.ccsd.cnrs.fr/docs/00/06/26/99/PDF/sic_00001652.pdf > Consulté le 10 avril 2006.

JACOBS, Neil. « A summary of the position with respect to electronic theses in selected countries in selected countries in Europe : Collated from questionnaire responses in advance of an international workshop on e- theses». Amsterdam, January 2006. 19 p. [en ligne] < <http://www.surf.nl/download/Overview%20of%20all%20countries.pdf> > Consulté le 10 avril 2006.

JOHNSON, Richard K. « Institutional repositories: partnering with faculty to enhance scholarly communication ». In *D-LibMagazine*, Volume 8, Number 11, November 2002. [en ligne] < <http://www.dlib.org/dlib/november02/johnson/11johnson.html> > Consulté le 17 juillet 2006

LE HENAFF, Diane ; THIOLON, Catherine, « Gérer et diffuser les thèses électroniques : un choix politique pour un enjeu scientifique », *Documentaliste - Sciences de l'information*, octobre 2005, Volume 42, n° 4-5, p.272-282. Consultable sur *CCSD @rchiveSIC* [en ligne] < http://archivesic.ccsd.cnrs.fr/docs/00/06/27/11/PDF/sic_00001689.pdf > Consulté le 13 juillet 2006.

PLATE-FORMES DE DÉPÔT, DE PRODUCTION ET/OU DE DIFFUSION POUR LES THÈSES ÉLECTRONIQUES

Etudes et synthèses sur les différents systèmes

ALLIOTI, Frédéric. *Etude de deux chaînes d'édition numérique XML. Projet de diffusion électronique de la production scientifique de l'INSA.* 2003. 61 p. DESS d'Ingénierie documentaire : Enssib, [en ligne] <<http://www.enssib.fr/bibliotheque/documents/dessid/rsaliotti.pdf>> Consulté le 17 mars 2006

AVENTURIER, Pascal . « *Les thèses en ligne à l'Inra. Rencontres 2005 professionnels de l'IST Nancy 20-22 juin 2005* ». 2005. 32 p. [en ligne] <

http://72.14.221.104/search?q=cache:ZPGbyKvDVwcJ:www.inist.fr/rencontresIST/IMG/pdf/M.Pascal_AVENTURIER_.pdf+%22.+Rencontres+2005+professionnels+de+l%E2%80%99IST+Nancy%22&hl=fr&gl=fr&ct=clnk&cd=1 > Consulté le 17 juillet 2006

BOURGUIGNAT, Christelle, GILLES, Benjamin GUSLEVIC, Thierry, EMMANUEL, Marine, MAURY, Brigitte. *Mise en place d'une chaîne de traitement des thèses électroniques. Service Commun de Documentation de l'Université de la Méditerranée (Aix-Marseille 2).*

Mémoire de conduite de projet : Enssib : 2003

[en ligne] < <http://www.enssib.fr/bibliotheque/documents/dcb/M-2003-PRO-15-Aix-Marseille2-vol1.pdf> > Consulté le 17 mars 2006

BUDAPEST OPEN ACCESS INITIATIVE. *A Guide to Institutional Repository Software v 3.0.* [en ligne] < <http://www.soros.org/openaccess/software/> > Consulté le 17 mars 2006

SALAH, Soraya. *Mise en ligne en texte intégral des thèses de l'Université Paul-Sabatier . 2004. 77 p.*

Projet professionnel personnel : Enssib : 2004

[en ligne] <<http://www.enssib.fr/bibliotheque/documents/ppp/pppsalah.pdf>>
Consulté le 17 mars 2006

SUNYACH, Laurence. *Projet de mise en place d'un circuit de traitement et de diffusion des thèses électroniques au Service Commun de la Documentation de Paris III : aspects techniques. 2005. 78 p.*

Projet professionnel personnel : Enssib : 2004

[en ligne]< <http://www.enssib.fr/bibliotheque/documents/ppp/ppp-sunyach-M.pdf> >

Castore

Castore [en ligne] < <http://castore.blogspot.com/> > Consulté le 10 avril 2006.

Castore. « Présentation de la plate-forme » [en ligne] < <http://www.emn.fr/x-info/bn/plateforme.htm> > Consulté le 10 avril 2006.

DUMAS, Cédric. « CASTORE, une plate-forme de bibliothèque numérique de littérature grise » In *Biblioacid*, volume 2, Numéro 3, Octobre 2005, p. 10-18.

Equipe projet Castore [en ligne] < <http://adullact.net/projects/castore/> > Consulté le 10 avril 2006.

Cyberthèses

AVENTURIER, Pascal ; VOLLAND-NAIL, Patricia. « Les Thèses en ligne à l'INSA ». 2001. 28 p. [en ligne] < http://www.ile-de-science.org/pages/events/colloque20010301/PAventurier_PVollandNail_Presentation.pdf > Consulté le 17 mars 2006

ANNE, Abdrahamane. « Appropriation d'une plate-forme d'édition électronique basée sur XML : Cyberdocs ». 2003. 99 p.
DESS d'Ingénierie documentaire : Enssib : 2003
[en ligne] < <http://www.enssib.fr/bibliotheque/documents/dessid/rsanne.pdf> >
Consulté le 17 mars 2006

ANNE, Abdrahamane. « Cyberthèses : une solution à la visibilité de la science africaine ? ». 2004. 7 p. In *Actes de la conférence sur la publication et la diffusion électronique*. Dakar, Sénégal, 1-2 septembre 2004
[en ligne] < http://www.codesria.org/Links/conferences/el_publ/Abdrahamane_Anne.pdf > Consulté le 17 mars 2006

BOULETREAU, Viviane ; DUCASSE, Jean-Paul ; GILLIERON-GRABER, Marie-Pierre. « Cyberthèses en Europe », *BBF*, 2001, n° 6, p. 122-125
[en ligne] < <http://bbf.enssib.fr> > Consulté le 17 mars 2006

CRU. « Informations pour le téléchargement des sources Cyberdocs » [en ligne] < <http://sourcesup.cru.fr/cybertheses/> > Consulté le 30 mars 2006

Cyberthèses. « Projet Cyberthèses » [en ligne] < <http://cyberdocs.univ-lyon2.fr> > Consulté le 30 mars 2006

Université Lyon 2. « Thèses électroniques » [en ligne] < <http://theses.univ-lyon2.fr/index.php> > Consulté le 2 octobre 2006

Dspace

BASS, Michael J., et al. « DSpace - a sustainable solution for institutional digital asset services - spanning the information asset value chain: ingest, manage,

preserve, disseminate ». 2002.

[en ligne] < <http://dspace.org/technology/functionality.pdf> > Consulté le 17 juillet 2006

CLABAULT, Gérard. "DSpace : Un Système d'Archivage d'Objets Numériques".2004.In *A L A D I N* [Accès Libre aux Archives du Dépôt Institutionnel Numérique de la MSH-Alpes]

[en ligne] < <http://hdl.handle.net/1801/440> > Consulté le 20 juillet 2006

DSPACE [en ligne] < <http://www.dspace.org/> > Consulté le 10 avril 2006.

DSpace for MIT Faculty and Researchers [en ligne]

<<http://libraries.mit.edu/dspace-mit/>> Consulté le 20 juillet 2006

JONES, Richard. "DSpace vs. ETD-db: Choosing Software to Manage Electronic Theses and Dissertations." In *Ariadne*, no. 38, 2004. [en ligne] <

<http://www.ariadne.ac.uk/issue38/jones/> > Consulté le 10 avril 2006.

LECOCQ, Fabrice. "International Workshop on D-Space - Bangalore - 7-11 mars 2005".2005.In *A L A D I N* [Accès Libre aux Archives du Dépôt Institutionnel Numérique de la MSH-Alpes]

[en ligne] < <http://hdl.handle.net/1801/442> > Consulté le 20 juillet 2006

RAPHAËL, Laurence. "Dspace : Aspects organisationnels". 2004. In *A L A D I N* [Accès Libre aux Archives du Dépôt Institutionnel Numérique de la MSH-Alpes]

[en ligne] < <http://hdl.handle.net/1801/439> > Consulté le 20 juillet 2006.

EPrint

EPrint [en ligne] < <http://www.eprints.org/> > Consulté le 10 avril 2006.

Eprints handbook [en ligne] < <http://www.eprints.org/documentation/handbook/> > Consulté le 10 avril 2006

HEIDEN, Serge ; MOUNIER, Pierre. « *Mise en place d'un serveur d'auto-archivage Eprints à l'ENS Lettres et Sciences Humaines : présentation* » [En ligne]

< <http://eprints.ens-lsh.fr/eprints.pdf> > Consulté le 15 septembre 2006

WARNER Simeon. « Eprints and the Open Archives Initiative ». In *Library Hi Tech*, 21, no. 2, 2003, p. 151-158.

STAR

Agence Bibliographique de l'Enseignement Supérieur (ABES). « Manuel d'expérimentation de STAR». [en ligne] <
<http://www.abes.fr/abes/DesktopDefault.aspx?tabindex=4&tabid=555> > Consulté le 15 septembre2006

Agence Bibliographique de l'Enseignement Supérieur (ABES). « L'outil STAR : chaîne de dépôt national des thèses électroniques». [en ligne] <
http://www.sup.adc.education.fr/bib/intro/reuDir/2006/Presentation_STARmgx.pdf
> Consulté le 15 septembre2006

Agence Bibliographique de l'Enseignement Supérieur (ABES). « STAR : journées réseau 2006». [en ligne]
<http://www.abes.fr/abes/DocumentsWebAbes/reseau/journees_reseau/06_Star.ppt
> Consulté le 15 septembre2006

Agence Bibliographique de l'Enseignement Supérieur (ABES). « STAR carrefour des thèses électroniques». [en ligne] <
<http://www.abes.fr/abes/DesktopDefault.aspx?tabindex=3&tabid=428> > Consulté le 15 septembre2006

THESES EN LIGNE ET ASPECTS JURIDIQUES

Textes législatifs encadrant les thèses

Code de l'Education, partie législative, livres VI à VIII

Arrêté du 3 septembre 1998 relatif à la charte des thèses (J.O. n°210 du 11 septembre 1998, p.13867) :

<http://www.legifrance.gouv.fr/WAspad/UnTexteDeJorf?numjo=MENR9802320A>

Circulaire n° 2000-149 du 21 septembre 2000

Ministère de l'Education, de l'enseignement supérieur et de la recherche.

« Diffusion électronique des thèses » In *Bulletin Officiel de l'Education Nationale* n°34 du 28 septembre 2000. [en ligne] <

<http://www.education.gouv.fr/bo/2000/34/sup.htm> > Consulté le 10 avril 2006

Circulaire n° 05-094 du 29 mars 2005

Ministère de l'Éducation, de l'enseignement supérieur et de la recherche.

« Dépôt, signalement, diffusion et archivage des thèses sous forme électronique », circulaire n° 05-094 du 29 mars 2005 [en ligne] <

<http://www.sup.adc.education.fr/bib/Acti/These/circulaire.rtf> > Consulté le 17 mars 2006.

Arrêté du 7 août 2006 relatif aux modalités de dépôt, de signalement, de reproduction, de diffusion et de conservation des thèses ou des travaux présentés en soutenance en vue du doctorat (MENS0602085A, J.O n° 195 du 24 août 2006, page 12471, texte n° 24):

<http://www.sup.adc.education.fr/bib/Acti/These/textregl.htm#7aout> > Consulté le 5 septembre 2006.

Arrêté du 7 août 2006 relatif à la formation doctorale (MENS0602083A, J.O. n° 195 du 24 août 2006 page 12468, texte n° 22):

<http://www.legifrance.gouv.fr/WAspad/UnTexteDeJorf?numjo=MENS0602083A> > Consulté le 5 septembre 2006.

Recommandations AFNOR relatives aux métadonnées des thèses électroniques Thèses électroniques françaises (TEF). [en ligne] <

<http://websrv.abes.fr/abes/DesktopDefault.aspx?tabid=403> > Consulté le 18 juillet 2006

Recommandations ministérielles pour les thèses

Agence Bibliographique de l'Enseignement Supérieur (ABES). « Guide pour la rédaction et la présentation des thèses à l'usage des doctorants ». [en ligne] <

<http://www.sup.adc.education.fr/bib/Acti/These/guidoct.rtf> > Consulté le 13 juillet 2006

Agence Bibliographique de l'Enseignement Supérieur (ABES). « Guide pour la diffusion et l'archivage électroniques des thèses ». [en ligne] <

<http://www.sup.adc.education.fr/bib/Acti/These/guiduniv.doc> > Consulté le 13 juillet 2006

CREPPY, Rachel. « Diffusion des thèses ». [en ligne] <

<http://www.sup.adc.education.fr/bib/Acti/These/theses.htm> > Consulté le 14 septembre 2006

DUCLOS-FAURE, Danièle, CREPPY, Rachel. «*Les thèses électroniques : dépôt, signalement, diffusion et archivage : Réunion 2006 des directeurs des bibliothèques des établissements d'enseignement supérieur : MENESR-DES-SDBD-B2 25 et 26 janvier 2006* ». [en ligne] < http://www.sup.adc.education.fr/bib/intro/reuDir/2006/B2_diff.pdf > Consulté le 13 juillet 2006

Droit d'auteur et thèses électroniques

ANDREW, Theo. « *Intellectual Property and Electronic Theses* ». London: JISC, 2004. [en ligne] < http://www.thesesalive.ac.uk/archive/IP_etheses.pdf > Consulté le 17 mars 2006

BAUDRIER, Liliane ; GILLES, Benjamin ; PONCIN, Olivier. *Le chercheur, l'enseignant et l'étudiant face au droit d'auteur.* Villeurbanne : 2003. 109 p. Diplôme de conservateur de bibliothèque : Enssib : 2003. [en ligne] < <http://www.enssib.fr/bibliotheque/documents/dcb/M-2003-RECH-03-baudrier.pdf> > Consulté le 17 mars 2006

BERNAUDIN, Anne-Claire, DEGEZ , Camille, DUPUICH, Lucile, FEUCHOT, Nicole, NGUYEN, Claire, PINET, Nicolas. *Projet de mise en place d'un circuit de traitement et de diffusion des thèses électroniques au Service Commun de la Documentation de Paris III : aspect administratif et juridique.* Villeurbanne : Enssib, 2005. Mémoire de projet : Enssib : juin 2005.

FORMATION ET ACCOMPAGNEMENT DES DOCTORANTS

Bibliothèque de l'E.N.S. Lyon, « Rédaction d'articles ou d'une thèse : guide. » [en ligne] < <http://www.ens-lyon.fr/Bibli/Technique/redactionpubli.php> > Consulté le 17 mars 2006.

Bibliothèque Universitaire de Lille 1, « Enjeux et mise en place de la diffusion électronique des thèses » [en ligne] < <http://www.univ-lille1.fr/bustl/index.php?idSection=2&idPartie=2&idTopic=4&idPage=3> > Consulté le 17 mars 2006.

Cyberthèses, « Didacticiel : Produire et diffuser des documents structurés » [en ligne] < <http://theses.univ-lyon2.fr/didacticiel/> > Consulté le 17 mars 2006.

DEBOIN, Marie-Claude. « Publier et diffuser sa thèse électronique sur Internet : conseils aux doctorants accueillis par le Cirad » in CCSD @rchiveSIC [Archive Ouverte en Sciences de l'Information et de la Communication] [en ligne] < http://archivesic.ccsd.cnrs.fr/docs/00/06/26/99/PDF/sic_00001652.pdf > consulté le 10 avril 2006.

Ecole doctorale de Sciences Po Paris. « Rédiger et présenter une thèse » [en ligne] < http://ecoledoctorale.sciences-po.fr/theses/presenter_these.htm > Consulté le 10 octobre 2006

Ministère de la jeunesse, de l'éducation nationale et de la recherche, « Guide pour la rédaction et la présentation des thèses à l'usage des doctorants » [en ligne]. < <http://www.sup.adc.education.fr/bib/Acti/These/guidoct.pdf> > Consulté le 10 avril 2006

Université de Limoges, Service Commun de la Documentation . « Rédiger et styler sa thèse » [en ligne] < <http://www.unilim.fr/scd/theses/> > Consulté le 10 avril 2006

THÈSES ET ARCHIVES OUVERTES

OAI : généralités, moissonnage des thèses

FILY Marie-Françoise. *Introduction au concept d'archive ouverte.* Montpellier : Cirad-Dic, 2005. 16 p. In CCSD @rchiveSIC [en ligne] < http://archivesic.ccsd.cnrs.fr/docs/00/06/26/47/PDF/sic_00001523.pdf > Consulté le 18 juillet 2006

FOULONNEAU Muriel. *Collaborer pour de nouveaux services culturels en ligne. Le protocole OAI, protocole de collecte de métadonnées de l'initiative des archives ouvertes.* Paris : Ministère de la culture et de la communication, 2005. 59 p. [en ligne] < http://www.culture.gouv.fr/culture/mrt/numerisation/fr/technique/documents/guide_oai.pdf > Consulté le 18 juillet 2006

NAWROCKI François. *Le protocole OAI et ses usages en bibliothèque.* Paris, Ministère de la culture et de la communication, 2005. [en ligne] < <http://www.culture.gouv.fr/culture/dll/OAI-PMH.htm> >

SULEMAN, Hussein, EDWARD A. Fox. « Leveraging OAI Harvesting to Disseminate Theses. » *Library Hi Tech* 21, no. 2, 2003, p. 219-227.
< <http://pubs.cs.uct.ac.za/archive/00000018/01/lht2002ra.pdf> > Consulté le 13 juillet 2006

SUR LES NORMES ET LES STANDARDS

Sur les métadonnées des thèses

AFNOR, Groupe CG46/CN357/GE5. « Les métadonnées des thèses électroniques françaises : TEF », mars 2006 [en ligne] < <http://www.abes.fr/abes/documents/tef/recommandation/index.html> > Consulté le 10 avril 2006

ARTIST. « *Métadonnées pour une cyberinfrastructure de la recherche : le cas des thèses françaises* », mai 2006. [en ligne] < http://artist.inist.fr/article.php3?id_article=332 > Consulté le 17 juillet 2006

GOMEZ de REGIL Rosa Maria. « *Métadonnées des thèses numériques : formation enssib : Les enjeux de la normalisation des métadonnées, Villeurbanne 8 juin 2006* ». Villeurbanne : Doc'Insa, juin 2006, 30 p.

Généralités sur les métadonnées

BATTISTI, Michèle. « Décrire, stocker et retrouver l'information sur son site Web : les métadonnées ». In *Documentaliste*, vol. 36, no 1, janvier-février 1999, p. 27-28. [en ligne] < http://www.adbs.fr/uploads/journees/519_fr.php > Consulté le 10 avril 2006

BEAUDRY, Guylaine ; BOISMENU, Gérard. « La production d'une version électronique : les métadonnées », Expertise de ressources pour l'édition de revues numériques, [en ligne] < <http://revues.enssib.fr/titre/6tecnic/3verselec/9meta.htm> > Consulté le 30 mars 2006

CHARTRON, Guylaine. « Standards-Normes-Documents numériques : Introduction générale », Urfist-Paris, janvier 2000, [en ligne] < <http://www.ccr.jussieu.fr/urfist/presse/standard/coursintro.htm> > Consulté le 25 mars 2006

Ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche. « Métadonnées, normes et standards ». *Educnet*, juin 2006 [en ligne] < <http://www.educnet.education.fr/dossier/metadata/default.htm> > Consulté le 17 juillet 2006

MOREL-PAIR, Catherine. « Dublin Core et métadonnées : principes, formats et utilisations », Rencontre des professionnels de l'IST, juin 2004, [en ligne] < <http://www.inist.fr/rencontresIST/2004/pdf/interventions/metatas.PDF> > Consulté le 10 avril 2006

Table des annexes

GLOSSAIRE ET ABRÉVIATIONS.....	79
ANNEXE 1 : LES DOCTORANTS DE L'UCP.....	81
ANNEXE 2 : VOLUMÉTRIE DES THÈSES ET VENTILATION PAR DOMAINES	83
ANNEXE 3 : AVANTAGES ET INCONVÉNIENTS DES DEUX PRINCIPAUX FORMATS	85
ANNEXE 4 : PANORAMA DE L'OFFRE ACTUELLE D'ARCHIVES OUVERTES POUR LES THÈSES	86
ANNEXE 5 : TABLEAU SYNTHÉTIQUE DES AVANTAGES ET INCONVÉNIENTS DES SCÉNARIIS	95

GLOSSAIRE et ABRÉVIATIONS

CMS : Content Management System, système de gestion de contenu. Il s'agit d'une famille de logiciels de conception et de mise à jour dynamique de sites web, conçus pour un travail collaboratif et permettant de structurer le contenu. Ils fournissent une chaîne de publication (workflow, voir infra), avec par exemple possibilité de publier (mettre en ligne le contenu) des documents. Certains CMS incluent le contrôle de version.

DTD : Document Type Definition, ensemble de règles de construction d'un document XML.

ETD : e-theses and dissertation, thèses et mémoires en ligne

Fichier TEI : fichier XML conforme à la DTD TEI Lite

Fichier Word : Fichier informatique au format du logiciel Word de Microsoft

OAI: Open Archives Initiative

OCR : Optical character recognition, ou reconnaissance optique de caractères, logiciel conçu pour transcrire des images de textes tapuscrits ou manuscrits (acquis le plus souvent via des scanners) en texte, et de le sauvegarder sous une forme pouvant être exploité dans un traitement de texte pour enrichissement, et stocké dans une base de données ou du moins, sur un support sûr et exploitable par un système informatique.

OO: OpenOffice.org

SDX : Système Documentaire XML. Logiciel permettant la recherche et la consultation de documents XML

Stylage : Opération consistant à appliquer des styles aux différentes parties d'un document conformément des règles définies dans un modèle de document.

TEI : Text Encoding Initiative

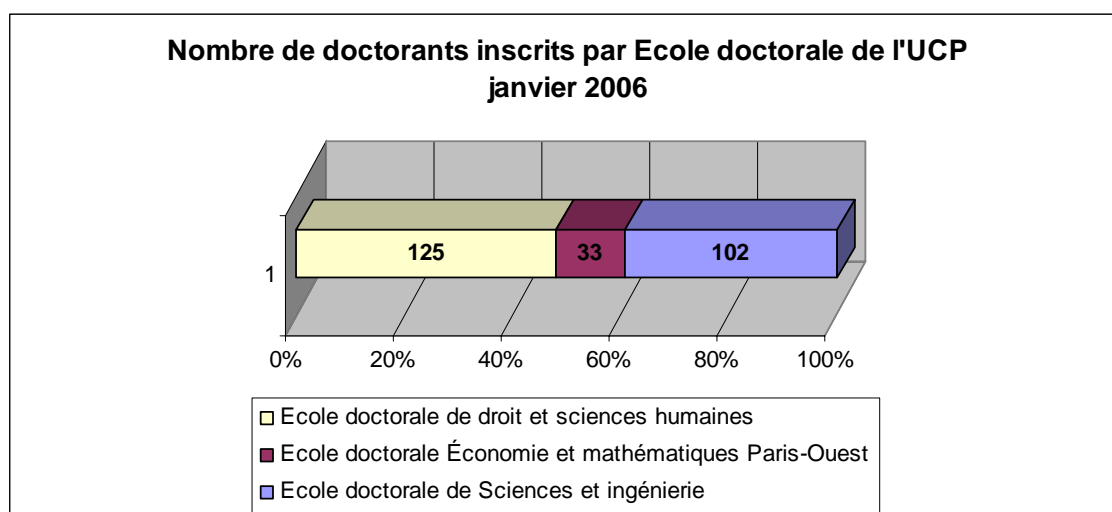
XML : eXtensible Markup Langage

XSL : eXtensible Stylesheet Langage

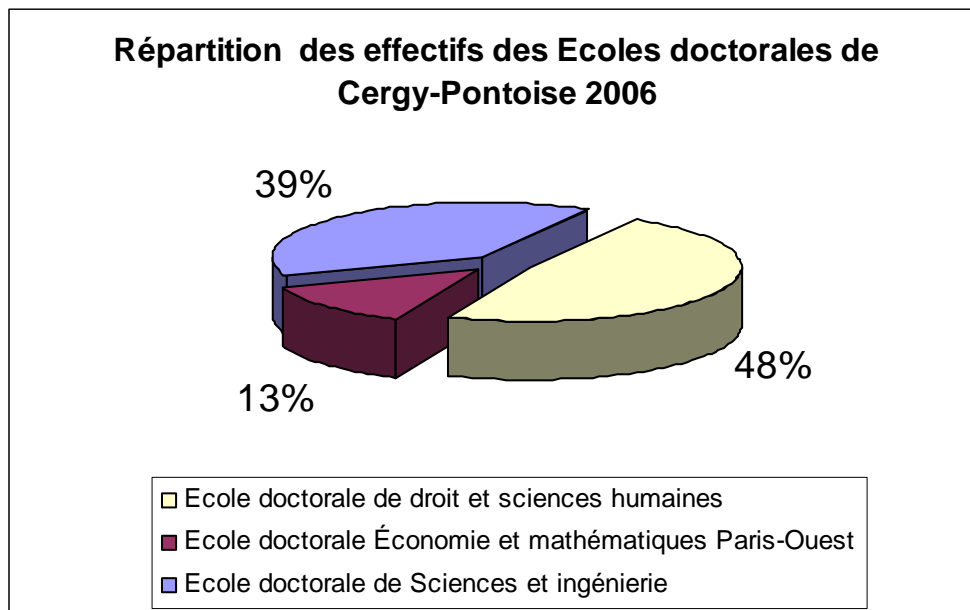
Workflow : ou « gestion électronique des processus », désigne la **modélisation** et la **gestion informatique** de l'ensemble des tâches à accomplir et des différents acteurs impliqués dans la réalisation d'un processus. Concrètement, le workflow décrit le circuit de validation, les tâches à accomplir entre les différents acteurs d'un processus, les délais, les modes de validation, et fournit à chacun des acteurs les informations nécessaires pour la réalisation de sa tâche. Il est en cela un outil de pilotage opérationnel.

Annexe 1 : Les doctorants de l'UCP

Graphique 1 : Nombre de doctorants inscrits en janvier 2006 dans les Ecoles doctorales

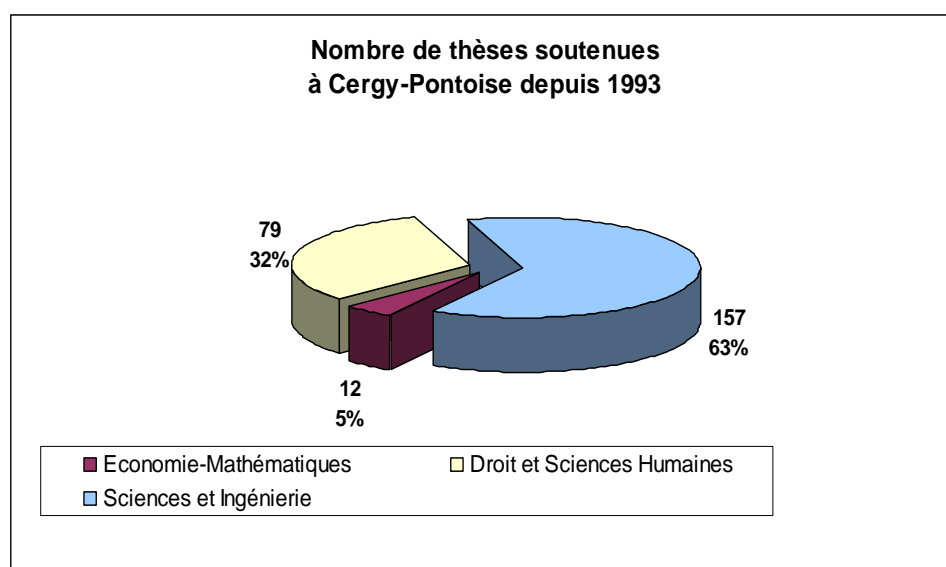


Graphique 2 : Poids respectifs des Ecoles Doctorales en fonction du nombre de doctorants

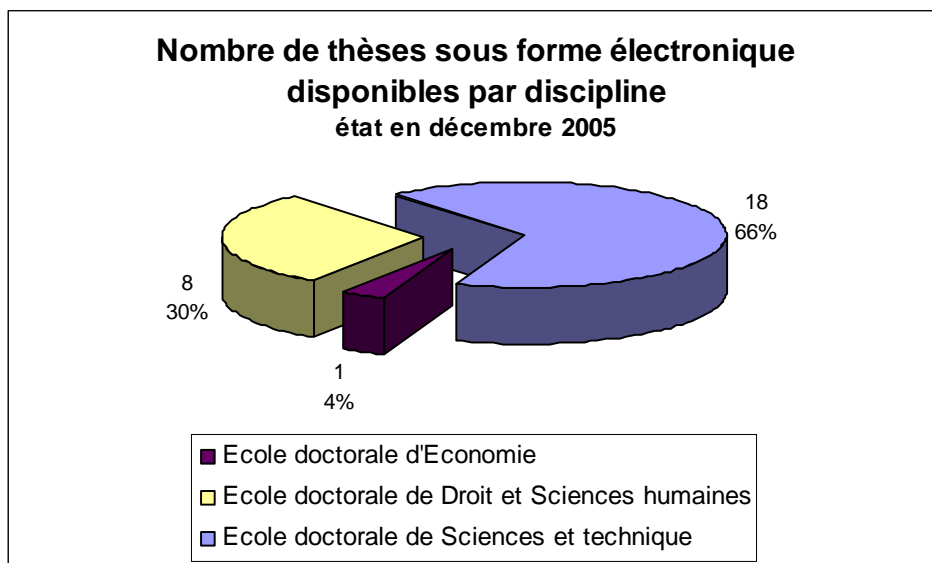


Annexe 2 : Volumétrie des thèses et ventilation par domaines

Graphique 4 : Répartition par domaine de l'ensemble des thèses soutenues à l'UCP depuis 1993



Graphique 5 : Répartition par domaine des thèses disponibles sur CD-ROM depuis 1993



Annexe 3 : Avantages et inconvénients des deux principaux formats

	XML	PDF
Avantages	Opère une séparation entre contenu, structure et présentation des documents ; se présente donc comme un format pivot	Forme fac-similé de la thèse : respect du choix de présentation de l'auteur, aucune modification de forme par rapport à la version canonique papier validée par le jury
	Format normalisé et codage de l'information en UNICODE : portabilité et relecture aisées	Très largement diffusé, compatibilité avec la quasi-totalité des plates-formes informatiques et logiciel de lecture gratuitement déchargeable
	Durabilité et réutilisation : l' archivage pérenne et la portabilité sont garantis	Protection des fichiers possible (contre la modification, le copier/coller, l'impression).
	Possibilités de recherches démultipliées : indexation du texte intégral et via les métadonnées	Fonctionnalités d'aide à la recherche incluses dans Acrobat Reader (fonction recherche , visualisation possible de la structure des documents...).
	Génération de texte possible en multiples présentations (diffusion en PDF ou HTML possible par exemple), tris, sélections, réorganisations de contenus...	Mode d'affichage ergonomique , graphiques affichés sans difficulté et format conçu pour permettre l'impression ;
	XML	PDF

	Langage nécessitant de la rigueur : nécessité d'utiliser une feuille de style afin d'obtenir un document structuré de manière conforme à la grammaire de la DTD.	Format propriétaire : pérennité moins garantie que l'archivage en XML
	Plate-forme de conversion nécessaire , et compétences techniques afférentes éventuelles	Poids du fichier qui peut être important
	Investissement en personnel	Fonctionnalités de recherche plus réduites

Annexe 4 : Panorama de l'offre actuelle d'archives ouvertes pour les thèses

Sur la base d'une répartition en trois catégories, on peut distinguer en France :

- les archives dédiées aux thèses, parmi lesquelles, les répertoires développés au sein des universités et contenant des thèses
- les archives pluridisciplinaires contenant des thèses
- les archives thématiques contenant des thèses.

Sous ces catégories seront signalées sans distinction archives institutionnelles ou collaborative, ainsi que d'autres exemples significatifs à l'étranger.

1. ARCHIVES FRANÇAISES DÉDIÉES AU THÈSES

ARCHIVE LIBRE DE L'ULP (Université Louis Pasteur, Strasbourg) : Ce serveur fonctionne avec le logiciel d'archives ouvertes GNU EPrints (voir

paragraphe dans la troisième partie).

<http://eprints-scd-ulp.u-strasbg.fr:8080/>

BIBLIOTHEQUES UNIVERSITAIRES D'AIX-MARSEILLE 2 (Université de la Méditerranée) : serveur proposant 126 thèses de Doctorat (1999-2002) interrogeables en Intranet. Il s'agit d'un projet, mis en oeuvre par le SCD lors du précédent contrat quadriennal, devant permettre d'évaluer la faisabilité du dépôt électronique des thèses de l'université inscrit au prochain contrat. Parallèlement, un accès est ménagé vers TEL pour les thèses soutenues à Luminy.

<http://bu2.univ-mrs.fr/> (rubrique "thèses")

CITHER (= Consultation en texte Intégral des THèses en réseau) **INSA Lyon** : projet pionnier de diffusion des thèses électroniques, initié en 1999 par l'INSA de Lyon. Il a donné lieu à de nombreuses coopérations avec d'autres établissements universitaires : travaux de normalisation des métadonnées, circuit de traitement des documents. Dans le cadre de ce projet, les thèses sont converties à partir de documents structurés et diffusées sur le site web de Doc'INSA sous la forme d'un PDF enrichi (table des matières active, renvois aux notes de bas de page et bibliographie, liens hypertextes).

Remarque : les documents ne sont pas convertis en XML. Le module de conversion réalisé en interne s'appuie sur des logiciels propriétaires (payants).

http://csidoc.insa-lyon.fr/these/recherche_ti.html

CYBERTHESES : Issu d'un programme francophone (coopération entre les Presses de l'Université de Montréal et l'Université Louis Lumière, Lyon 2) qui s'est élargi à l'Amérique du sud. Cette plate-forme opérationnelle est portable, libre, et propose une chaîne de production allant de la rédaction du document à sa diffusion, en passant par le stockage.

Accès aux textes intégraux des thèses soutenues dans les différents établissements participant à Cyberthèses (Université Lyon2, Université de Montréal).

<http://cyberdocs.univ-lyon2.fr/>

Un didacticiel est proposé en collaboration avec le programme Cyberthèses pour produire et diffuser des documents électroniques structurés : <http://sophia.univ-lyon2.fr/didacticiel/>

ECOLES DOCTORALES DE L'UNIVERSITE BORDEAUX 1 : 350 thèses, spécifiquement sur les domaines des sciences et des techniques, accessibles sous

format PDF.

http://147.210.235.2/ecoles/docs/ligne_imp_the.html

UNIVERSITE CHARLES DE GAULLE-LILLE 3

<http://www.univ-lille3.fr/portail/index.php?page=Theses>

UNIVERSITE JOSEPH FOURIER – GRENOBLE : renvoie directement sur une page de TEL configurée après négociation avec le CCSD pour constituer une interface spécifique dans la section « collection » de cette application de HAL.

<http://tel-ujf.ujf-grenoble.fr/>

UNIVERSITE LUMIERE- LYON 2/ CYBERTHESES : Accès à 778 thèses en ligne par une recherche simple ou en mode expert. (voir paragraphe sur Cyberthèses dans la partie 3).

<http://theses.univ-lyon2.fr/>

UNIVERSITE DE PARIS II (PANTHEON - ASSAS) - BIBLIOTHEQUE DOCTORALE

Thèses soutenues à l'Université de Paris II (Panthéon - Assas) en droit, science politique, sciences économiques, sciences de gestion, sciences de l'information et de la communication et sciences de l'éducation. Accès uniquement par thème (pas de module de recherche) et documents téléchargeables en PDF.

<http://www.u-paris2.fr/html/recherche/Theses%20en%20ligne/index.php3?>

UNIVERSITE PARIS IV-SORBONNE : utilise la plate-forme cyberthèses ; pour l'instant 2 thèses en ligne (juillet 2006).

http://www.paris4.sorbonne.fr/fr/article.php3?id_article=3368

UNIVERSITE DE PARIS X NANTERRE: PUBLICATIONS : Ce serveur Eprints fonctionne, tout comme l'archive de l'ULP avec le logiciel d'archives ouvertes GNU EPrints. Contrairement à l'archive de l'ULP, ne sont proposés ni de recherche simple, ni de avancée, et les résultats ne proposent pas un résumé en plus du document en PDF.

Les fonctions de recherche permettent un accès direct par sujets ou par année, une mise en valeur des derniers ajouts (dernière semaine). Les résultats sont présentés : avec une arborescence reproduisant le plan de classement pour une recherche par sujet; avec une présentation par type de document pour une recherche par année.

L'aspect collaboratif est mis en avant dans l'interface : un accès au dépôt par identifiant est proposé sur la page d'accueil.

<http://publications.u-paris10.fr/>

UNIVERSITE DE REIMS CHAMPAGNE-ARDENNE (URCA) : le serveur de thèses emploie un système de GED intitulé « projet Pelleas », développé par la société Jouve pour le compte de l'Université de Marne-la-Vallée, maître d'ouvrage du projet, et l'Université de Reims Champagne-Ardennes. Ce système, fondé sur une conversion au format XML, assure la codification, le stockage, la gestion des métadonnées et la diffusion de documents de structures et de formats hétérogènes. Il optimise l'exploitation et la consultation des données sur le Web (Internet/Intranet) via un système d'interrogation unique.

<http://scdurca.univ-reims.fr/exl->

[php/cadcgp.php?MODELE=vues/commun/charte/present.html&query=1](http://scdurca.univ-reims.fr/exl-php/cadcgp.php?MODELE=vues/commun/charte/present.html&query=1) (rubrique Thèses)

Archives ouvertes

INP (Institut National Polytechnique de Toulouse) : 168 thèses sont actuellement présentes dans cette archive ouverte; les types de recherches peuvent se faire par Sujet, par Année ou par Laboratoire. Un accès spécifique a été ajouté pour consulter les thèses primées.

<http://ethesis.inp-toulouse.fr/>

PASTEL (site de ParisTech) : diffuse au format PDF les thèses de ParisTech, grandes écoles d'ingénieurs de Paris sur un serveur commun. Le dépôt électronique est effectué par le lauréat lui-même après soutenance, en théorie sous tous formats (mais majorité effective de PDF). Pas de système de production propre. Cette plate-forme, déclarée auprès de l'[Open Archive Initiative](#), propose un accès au texte intégral et les thèses, diffusées au format PDF, et proposées également en Word ou Postcript, peuvent être retrouvées dans TEL (Thèses-En-Ligne).

<http://pastel.paristech.org/>

PROJET « THESES EN LIGNE A L'INRA »: utilise la chaîne de traitement cyberthèses ; accès au texte intégral.

<http://www.inra.fr/theses/projet-these.html>

TEL (Thèses en ligne) / CCSD (Centre de la Communication Scientifique Directe): serveur de thèses multidisciplinaires qui propose actuellement environ 4000 documents en texte intégral. Il dispose également d'un large réseau de participants, en particulier la cellule Mathdoc et l'IMAG en tant que premiers membres.

<http://tel.ccsd.cnrs.fr/>

2. ARCHIVES FRANÇAISES NON DÉDIÉES AUX THÈSES

Archives pluridisciplinaires contenant des thèses

Grisemine est une initiative du SCD de Lille qui propose en texte intégral des documents de littérature grise (hors circuit de publication et de diffusion commercial) : communications effectuées lors de colloques ou de séminaires, articles de périodiques, rapports de recherche, thèses, supports de cours ou de travaux dirigés, mémoires d'étudiants.

<http://crdoc.univ-lille1.fr/default.asp?bustl/grisemine>

CASTORE est également une bibliothèque numérique, destinée à la création d'archives de littérature grise pour les instituts de formation supérieure et de recherche. Pour l'instant, accès direct en ligne aux résultats des travaux uniquement par demande de test auprès des administrateurs. <http://www.emn.fr/x-info/bn/>

Archives thématiques contenant des thèses

a) Les serveurs du CCSD : HAL et ses dérivés

Hyper Article en Ligne (HAL) : HAL est un outil de communication scientifique directe entre chercheurs, permettant à ces derniers de déposer et de rendre publics des documents scientifiques de toutes les disciplines. Le contenu doit être comparable aux manuscrits que les chercheurs soumettent pour publication aux comités de lecture de revues scientifiques, mais les documents ne sont soumis à aucune évaluation scientifique détaillée. Les documents déposés dans HAL bénéficient de la pérennité de cette base, y restent de façon permanente et reçoivent une adresse web stable; ainsi, comme une publication dans un journal scientifique habituel, ils peuvent faire l'objet d'une

citation dans un autre travail.
<http://hal.ccsd.cnrs.fr/>

Les instances hébergées par HAL sont :

Thèses en ligne : [TEL](#) ; **Les archives des Sciences de l'Homme et de la Société**

– **HAL-SHS:** <http://halshs.ccsd.cnrs.fr/> ; **Les archives de l'IN2P3 :** [HAL-IN2P3](#) ; **Les archives de l'INRIA :** [HAL-INRIA](#) ; **Les archives de l'INSERM :** [HAL-INSERM](#) ; **Archive EduTice, éducation et Tic** (Maison des sciences de l'homme) : <http://archive-edutice.ccsd.cnrs.fr/index.php> ; **Archive de**

l'Institut Nicod : <http://jeannicod.ccsd.cnrs.fr/> ; **@rchiveSic**, Sciences de l'Information et de la Communication : <http://archivesic.ccsd.cnrs.fr/> ; **CIEL**, Codes Informatiques en ligne : <http://ciel.ccsd.cnrs.fr/> ; **Artxiker**, Archive pour la langue basque : <http://artxiker.ccsd.cnrs.fr/> ; **AOLab**, une archive ouverte expérimentale de laboratoire : http://ccsd.cnrs.fr/article.php3?id_article=20 ;

Projet Européen des thèses en ligne, ETOL (site expérimental) :

<http://etol.ccsd.cnrs.fr/> ; **Cours en ligne, CEL :** <http://cel.ccsd.cnrs.fr/> ; **Mémoires de 3° cycle en SIC, memSic :** <http://memsic.ccsd.cnrs.fr/>.

b) Archives thématiques hors CCSD

A L A D I N [Accès Libre aux Archives du Dépôt Institutionnel Numérique de la MSH-Alpes] : archive ouverte élaborée avec Dspace, qui propose des working papers, préprints et travaux de chercheurs de la Maison des Sciences de l'Homme Rhône-Alpes. Actuellement (juillet) en refonte car passage à la version 2.0 de Dspace.

<https://dspace.msh-alpes.prd.fr/index.jsp>

IFREMER : Accès aux thèses en texte intégral (format PDF) par une recherche simple, avancée (mot-clé, auteur...) ou feuilletage. Possibilité d'export de la notice (format .txt).

<http://www.ifremer.fr/docelec/>

IRCAM : Base de donnée d'articles, moissonnables en OAI qui comporte 62 thèses répertoriées, dont 17 consultables sur internet, format HTML ; utilise le CMS Typo 3.

<http://articles.ircam.fr/>

Le portail documentaire du SUDOC :

plate-forme de signalement bibliographique pluridisciplinaire, qui permet la consultation en texte intégral des thèses dans le catalogue SUDOC, le catalogue XTCat NDLTD (USA) et le catalogue Thèses.

<http://www.portail-sudoc.abes.fr/>

[A noter que [Die Deutsche Bibliothek](http://www.ddb.de/) en Allemagne signale 29 000 thèses électroniques, sur le même modèle que le portail du Sudoc en France.

<http://www.ddb.de/>]

3. BIBLIOTHÈQUES NUMÉRIQUES ET RÉPERTOIRES DE THÈSES À L'ÉTRANGER

AUSTRALIAN DIGITAL THESES PROGRAM : ce programme, développé sur une base collaborative à partir de 1998-1999 par sept grandes universités australiennes, consiste en une plate-forme unique de digitalisation et de diffusion de thèses. Sur un plan local, la plate-forme logicielle ADT permet aux établissements volontaires de convertir les thèses dans un format numérique, et de générer automatiquement des métadonnées en Dublin Core. Les thèses sont archivées sur les propres serveurs des établissements, lesquels ont aussi toute latitude pour donner un accès direct à ces thèses via une infrastructure locale de diffusion. L'archive comportait 9 782 thèses en ligne (format PDF) sur un total de 135 254 répertoriées début août 2006.

<http://adt.caul.edu.au/>

La BIBLIOTHEQUE DE L'UNIVERSITE DE LAVAL : diffuse les thèses et mémoires rédigés et déposés sur support électronique depuis 2002, la rédaction et le dépôt sur support électronique étant laissés au libre choix des étudiants. Plusieurs milliers de thèses québécoises sont disponibles, outre celles de l'Université Laval.

<http://www.theses.ulaval.ca/>

BICTEL/e : rassemble dans un répertoire commun la production de la recherche (38 thèses, et des e-prints) des universités de la Communauté française de Belgique, et la rend librement accessible en ligne, en texte intégral. La base est alimentée par les scientifiques, chercheurs, étudiants de l'institution selon le principe de l'auto-archivage. L'interface propose un accès sur tout le répertoire,

par type de document (thèses ou e-prints) ou par affiliation.

http://edoc.bib.ucl.ac.be/participant_bictel.html

CYBER DOCUMENTS / Université de Genève : également bâtie à partir de cyberdocs (version antérieure à cyberthèses) ; propose 357 thèses en ligne au format PDF

<http://www.unige.ch/cyberdocuments/>

DISSERTATION.com : Site permettant de rechercher et publier un travail universitaire (thèse ou mémoire) en anglais ; contient les publications du NDLTD.

<http://www.dissertation.com/> (accès document payant, cependant une partie du document est gratuite).

DISSONLINE : site allemand qui coordonne les développements autour des thèses électroniques.

http://www.dissonline.de/index_e.htm

DIVA : site fédérateur multidisciplinaire d'une douzaine d'universités nordiques.

La base comporte environ 3500 thèses et rapports de recherche.

<http://www.diva-portal.se/index.xsql?lang=en>

DOKUMENTENSERVEN DER HÜMBOLT-UNIVERSITÄT -BERLIN :

<http://edoc.hu-berlin.de/>

(pour une liste plus exhaustive des universités allemandes proposant ce type d'initiatives, voir sur le lien suivant :

http://www.dissonline.de/convert_engl/index_such_e.htm)

ELECTRONIC THESES AD DISSERTATIONS : répertoire issu d'un projet de collaboration entre 200 universités et institutions, qui propose environ 8600 thèses et mémoires universitaires sous format électronique classés par auteur et par discipline.

<http://scholar.lib.vt.edu/theses/>

ETOL : site expérimental de mise en ligne des thèses européennes (Allemagne, Suède, Espagne, Royaume Uni, France), dont les archives contiennent actuellement plus de 36600 thèses.

<http://etol.ccsd.cnrs.fr/>

INFOSCIENCES / EPFL : base de données de publications, rapports de recherche, thèses, travaux d'étudiants, cours, etc., provenant des facultés, laboratoires et chercheurs de l'Ecole Polytechnique Fédérale de Lausanne (EPFL). Plus de 3600 thèses de doctorat pour la plupart disponibles en ligne sous format PDF.

<http://infoscience.epfl.ch/>

Networked Digital Library of Theses Dissertations (NDLTD): Répertoire des bibliothèques électroniques sur les thèses des institutions participants.

<http://www.theses.org/> (la plupart des documents sont gratuits).

MIT (Massachusetts Institute of Technology) Theses Online : archive qui comprend l'intégralité de la production de recherche au MIT sous forme numérique (preprints, rapports techniques, working papers, thèses, conférences...). Elle utilise le logiciel DSpace, et offre une recherche par affiliation, titre, auteur, année, mais aussi un accès personnalisé en vue du travail collaboratif (auto-archivage).

<http://theses.mit.edu/>

PAPYRUS (Université de Montréal) : dépôt institutionnel structuré avec le logiciel Dspace, proposant différentes formes de travaux de recherche et d'enseignement (pré-publications, articles, rapports techniques, présentations, ensembles de données, thèses, etc.) produites par les professeurs et chercheurs de l'Université de Montréal. Interface identique à celle du MIT, et 133 documents archivés à ce jour.

<https://papyrus.bib.umontreal.ca/dspace/>

RERO (Réseau des Bibliothèques de Suisse Occidentale) : propose une bibliothèque numérique comportant plus de 4000 documents (livres, journaux, postprints et thèses). Cet outil dispose d'un mode de recherche riche, notamment des possibilités de navigation par domaines, institutions, collections et sites.

<http://doc.rero.ch/?ln=fr>

UMI (University Microfilms) Proquest : accès gratuit aux deux dernières années de la base de données, sinon accès aux documents payants.

<http://wwwlib.umi.com/dissertations/gateway/>

Annexe 5 : tableau synthétique des avantages et inconvénients des scénarii

	AVANTAGES	INCONVENIENTS
Scénario 1 : Implantation de Cyberthèses	<ul style="list-style-type: none"> - Proximité institutionnelle avec les différents acteurs et réactivité - Rapidité de mise en œuvre - Excellente visibilité/ identification institutionnelle - Production de documents structurés (XML): performances de recherches et pérennité de l'archivage - Archivage local : maîtrise de l'université - Intégration possible dans un autre programme 	<ul style="list-style-type: none"> - Forte implication humaine nécessaire - Corpus et publics cible limités dans un premier temps - Antagonisme / effet doublon avec STAR - Archivage local : suppose un serveur / de l'espace serveur, avec un usage finalement limité par rapport aux potentialités.
Scénario 2 : Partenariat avec TEL	<ul style="list-style-type: none"> - Coût humain faible (quasi externalisation) - Moindre technicité nécessaire pour des documents convertis en PDF - Extrême rapidité de la mise en ligne (24 heures) - Archive pluridisciplinaire avec vues institutionnelles, mais spécialisée dans les thèses - Insertion dans une logique nationale - L'usage d'un serveur d'archivage local n'est pas nécessaire, TEL permettant la récupération de collections entières 	<ul style="list-style-type: none"> - Moindre maîtrise de la gestion par l'université - Visibilité assurée, mais lien identitaire avec l'université également moindre - Fichiers en PDF : - recherches moins performante - question de la pérennité des données - Problèmes liés à l'autoarchivage
Scénario 3 : Une archive institutionnelle	<ul style="list-style-type: none"> - Proximité institutionnelle et réactivité - Vitrine prestigieuse pour l'établissement - Compatibilité avec tous les formats envisageables - Archivage local : maîtrise de l'université sur toute sa production - Un projet de forte ampleur à terme 	<ul style="list-style-type: none"> - Mise en œuvre nécessairement lente à l'échelle de l'établissement - Forte technicité nécessaire pour la mise en place; - Communication beaucoup plus lourde à prévoir au démarrage ; - Archivage local (serveur) : coût de maintenance