RAPPORT SUR LA DIFFUSION ELECTRONIQUE DES THESES

établi par un groupe de travail

Rédaction de Claude JOLLY, chargé de la sous-direction des bibliothèques et de la documentation

MINISTERE DE L'EDUCATION NATIONALE

Direction de l'enseignement supérieur

sous-direction des bibliothèques et de la documentation

juillet 2000

PREAMBULE

Par note en date du 26 novembre 1999, Mme Jeanne-Marie PARLY, directrice du cabinet du ministre de l'Education nationale, de la Recherche et de la Technologie, demandait à la directrice de l'enseignement supérieur de constituer un groupe de travail sur la numérisation des thèses et leur diffusion par voie électronique.

Constitué en décembre 1999 et janvier 2000, ce groupe de travail était composé de représentants de la direction de l'enseignement supérieur (sous-direction des bibliothèques et de la documentation), de la direction de la recherche, de la direction de la technologie, ainsi que de représentants de la conférence des présidents d'université (C.P.U.) et de l'association des directeurs de la documentation et des bibliothèques universitaires (A.D.B.U.), auxquels ont été adjoints des experts ayant, pour la plupart d'entre eux, conduit des expérimentations en la matière. On trouvera en annexe 1 la composition détaillée du groupe.

Ce dernier a tenu trois séances de travail, les 10 février, 6 mars et 27 mars 2000. Chaque séance de travail était précédée d'un document introductif et suivie d'un relevé de conclusions. Ceux-ci sont reproduits en annexes 2 et 3.

Le présent rapport dresse la synthèse des conclusions du groupe de travail. Il est constitué de trois parties :

- le rapport lui-même, qui a été rédigé par Claude JOLLY, chargé de la sous-direction des bibliothèques et de la documentation ;

- des prescriptions techniques pour le dépôt des thèses en format électronique, rédigées par un sous-groupe animé par Christian LUPOVICI, directeur du service commun de la documentation de l'Université de Marne-la-Vallée ;

- des annexes, établies par Christine OKRET, conservateur au bureau de la modernisation des bibliothèques (sous-direction des bibliothèques et de la documentation).

PLAN

INTRODUCTION

Environ 11.000 thèses de doctorat sont soutenues chaque année. On sait qu'à ce terme sont attachées trois réalités complémentaires mais distinctes :

– la thèse est un *travail de recherche* ;

– elle est un *exercice académique* donnant accès à un grade universitaire ;

– elle est enfin un *document*.

C'est à la thèse en tant que *document* que le groupe de travail s'est principalement attaché. Depuis toujours en effet, l'Etat et les universités se sont préoccupés de la valorisation des thèses c'est-à-dire de leur conservation, de leur signalement et de leur accessibilité. Se rejoignent ici les intérêts du jeune docteur, ceux de l'établissement de soutenance et des chercheurs, et plus globalement, la nécessité d'assurer une visibilité internationale aux résultats des recherches conduites dans les universités françaises.

Il y a 15 ans, un arrêté du 25 septembre 1985 a défini les modalités de dépôt, signalement et reproduction des thèses. Il précise notamment que :
– deux exemplaires de la thèse doivent être déposés dans la bibliothèque de l'université de soutenance ;
– le catalogage de ces documents doit être effectué par trois pôles de signalement dûment identifiés ;
– deux ateliers de reproduction assurent la reproduction sur microfiches de l'ensemble des thèses ;
– ces microfiches sont largement diffusées dans les bibliothèques universitaires, de façon à favoriser la consultation de ces travaux.

Tout le monde s'accorde aujourd'hui à reconnaître que le dispositif mis en place en 1985 est désormais obsolète. Outre le fait que certaines dispositions du texte ne sont plus applicables ou plus appliquées, il est clair que les usagers ne sont pas satisfaits de la situation actuelle : cette conclusion est ressortie sans ambiguïté de *l'Enquête sur les pratiques des utilisateurs du signalement des thèses et des utilisateurs de thèses*, effectuée en novembre 1997 par la société SCP Communication, à la demande du ministère. Ceux-ci souhaitent manifestement avoir accès rapidement et facilement à l'information secondaire (bibliographique) correspondante, pouvoir apprécier dans les mêmes conditions si une thèse repérée est de nature à leur être utile dans leur travail, pouvoir consulter le texte intégral et, le cas échéant, pouvoir passer commande d'un substitut (photocopies, microformes, éditions à la carte).

Si l'on veut bien mettre en regard cette attente des usagers et trois phénomènes nouveaux mais déterminants, à savoir que
– la quasi-totalité des thèses sont désormais produites " nativement " sous forme numérique,
– les équipements et réseaux des établissements d'enseignement supérieur connaissent un grand développement,

– beaucoup d’universités se positionnent désormais en tant que producteurs et diffuseurs d’informations électroniques,

on peut considérer que le moment est venu de redéfinir radicalement les conditions de diffusion des thèses. C’est sur ces bases nouvelles que le groupe de travail a élaboré les propositions qui suivent.

I. LES OBJECTIFS

Avant de préconiser la mise en œuvre d'un nouveau dispositif, il est apparu nécessaire de clarifier les objectifs recherchés en répondant à quelques questions simples :

– toute thèse soutenue doit-elle être mise en ligne sur le réseau Internet ?
– faut-il privilégier l'accès au texte intégral ou s'en tenir à des extraits significatifs ?
– faut-il privilégier le mode image ou le mode texte ?
– faut-il fixer une durée de mise en ligne et en quels termes doit-on poser les questions de l'archivage ou de la conservation ?
– convient-il d'encourager ou de favoriser la fabrication de substituts ?

Sur ces différents points, un consensus s'est dégagé assez rapidement. Le parti suivant lequel il convient de bénéficier pleinement des possibilités ouvertes par les technologies actuelles sans s'imposer des limites ou restrictions *a priori* s'est assez logiquement imposé.

1. Une diffusion par voie électronique aussi large que possible

On sait que plusieurs dispositions législatives et réglementaires se rapportent à la diffusion et à l'accès aux thèses :

– en tant qu'œuvre de l'esprit, la thèse relève de la législation sur la propriété intellectuelle (loi de 1957 et textes dérivés) et donc des droits reconnus aux auteurs sur leurs productions ;
– les juridictions administratives assimilent les thèses à un document administratif, auquel s'appliquent les règles d'accès aux dits documents : sauf décision explicite du président ou directeur de l'établissement de soutenance, toute thèse est consultable dans la bibliothèque de l'université de soutenance où elle doit être déposée ;
– selon l'arrêté du 25 septembre 1985, il revient au président ou au directeur de l'établissement de soutenance d'autoriser ou non, sur avis du président du jury, la reproduction d'une thèse.

Si l'ensemble de ces règles juridiques qui correspondent à autant de protections ou de garanties (protection des découvertes ; maîtrise par l'auteur de l'exploitation de son travail ; etc.) doivent être respectées, il importe en revanche de favoriser la diffusion électronique la plus large possible des thèses, dès lors que le chef d'établissement (après avis du jury), d'une part, et l'auteur, d'autre part, auront donné leur accord. A cet égard, il conviendra de prévoir dans la nouvelle version du formulaire qui accompagne le dépôt de la thèse une rubrique correspondante afin de simplifier les procédures.

Cette recommandation concerne l'ensemble des thèses de doctorat, à l'exclusion des thèses d'exercice en médecine dont la valorisation ne paraît pas nécessaire pour la recherche. Par ailleurs, les travaux qui accompagnent les habilitations à diriger des recherches (HDR) ne rentrent pas non plus dans le cadre ici dessiné.

Les thèses constituées d'articles déjà publiés dans des revues scientifiques, phénomène qui tend à se développer dans certaines disciplines, par exemple la physique, constituent un cas

particulier appelant des solutions spécifiques. Au demeurant, on remarquera que les travaux correspondants ont déjà fait l'objet d'une diffusion et ont moins besoin que d'autres d'être valorisés.

2. L'accès au texte intégral …

La question de savoir s'il convient de donner accès à la totalité de la thèse ou seulement à des extraits significatifs (résumé, introduction, conclusion, bibliographie, table des matières) mérite d'être posée. En effet, l'on peut considérer que des extraits significatifs sont suffisants en première instance pour permettre au lecteur de vérifier si la thèse est susceptible de lui apporter des informations utiles et s'il lui est nécessaire de recourir au document primaire. Par ailleurs, l'expérimentation Webthèses effectuée par l'Atelier national de reproduction des thèses de Lille et le Centre informatique national de l'enseignement supérieur a montré que certains auteurs opposés à l'accès à l'intégralité de leur thèse pouvaient en revanche être favorables à la mise en ligne d'extraits.

En dépit de la valeur de ces arguments, le parti préconisé est clairement celui de l'accès au texte dans son intégralité : on voit mal en effet ce qui pourrait justifier des restrictions d'accès à l'information alors même que les questions de taille et de coûts de mémoire informatique ne se posent plus ; par ailleurs, un dispositif à géométrie variable par lequel on aurait accès, selon le cas, à des thèses en texte intégral et à des thèses constituées seulement d'extraits, souffrirait d'un défaut de lisibilité.

3. … en mode texte

Autrefois débattue, la question de l'alternative mode image/mode texte ne se pose plus. Dès lors que les thèses sont désormais dans leur quasi-totalité des documents électroniques " natifs ", l'option du mode texte doit à l'évidence être privilégiée. Cette option favorisera en outre une navigation enrichie et divers traitements secondaires. Rien ne justifierait de se priver à présent de ces possibilités.

4. Permanence de l'accès en ligne et archivage

Chacun s'accorde à reconnaître que les informations contenues dans les thèses n'ont pas la même pérennité scientifique selon les différentes disciplines. La dichotomie connue entre les sciences dites " accumulatives " (sciences humaines et sociales, mathématiques, astronomie) et les sciences " non accumulatives " est en l'espèce indiscutable. Dans ce contexte, il ne serait pas illégitime d'établir une correspondance entre la durée de disponibilité en ligne d'une thèse et sa discipline.

Après analyse, il n'est pas cependant apparu pertinent de fixer en la matière des normes *a priori*. C'est bien plutôt la fréquence des usages qui doit présider à la décision de mettre en ligne un document de façon permanente ou au contraire de l'archiver.

La question de la mise en ligne est donc directement liée à celle de l'archivage électronique dont le groupe de travail unanime a rappelé l'impérieuse nécessité pour deux raisons étroitement complémentaires :

– un impératif de conservation à long terme (on rappellera que les exemplaires papier conservés dans les bibliothèques ne présentent pas de ce point de vue de garanties

absolues)
– un impératif de mise en ligne qui doit être susceptible d'évoluer en fonction de la demande.

5. La fabrication de substituts

Si la communauté scientifique manifeste son attachement à un accès électronique aux thèses, elle souligne en même temps l'intérêt que peut représenter pour elle la fabrication de substituts de divers types (microformes, livres à la carte, etc.). L'activité en ce sens de l'A.N.R.T. de Lille en témoigne : il s'agit en outre d'une demande solvable émanant aussi bien de centres de documentation, de laboratoires, que de chercheurs ou de particuliers. Il conviendra en conséquence de préconiser des solutions qui facilitent la production de tels substituts à partir des fichiers informatiques qui auront été constitués.

II. LES OPTIONS TECHNIQUES

Dès lors que la quasi-totalité des thèses sont, ainsi qu'on l'a rappelé, produites dans un format électronique natif, la question de leur numérisation ne se pose pas. En revanche, il convient de s'interroger sur les modalités de production du document original ainsi que sur sa conversion ultérieure dans un format d'archivage et de diffusion, les deux opérations étant au demeurant étroitement liées.

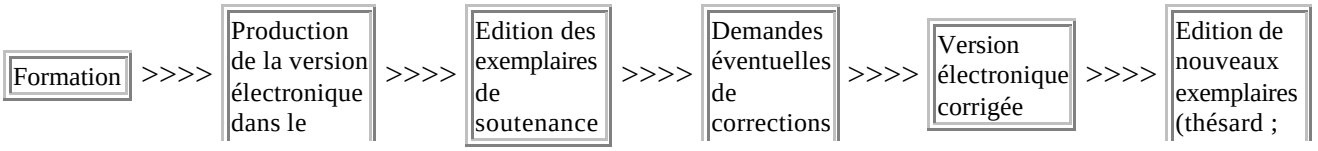
1. Les formats natifs

Suivant quels principes les thèses doivent-elles être confectionnées par les doctorants ? En la matière, il est essentiel de ménager un juste équilibre entre une absence de toute prescription qui déboucherait sur un trop grand désordre et rendrait problématique la mise en œuvre de chaînes de traitement propres à convertir chaque thèse dans un format approprié et, à l'inverse, des prescriptions trop exigeantes qui ne seraient dans les faits respectées que par un faible pourcentage de thésards. C'est la raison pour laquelle les préconisations formulées ci-après (*Prescriptions techniques pour le dépôt des thèses en format électronique*) constituent un ensemble de règles réalistes, compatibles avec les pratiques et le niveau de compétence des candidats, tout en représentant le plus petit commun dénominateur exigible pour constituer un corpus suffisamment homogène. Elles portent notamment sur :
– les formats de production (compatibles RTF et, pour certaines disciplines scientifiques, LaTeX)
– les feuilles de style qui président à la structuration et au balisage du document.

Il ne fait aucun doute que la mise en pratique de ces prescriptions, même si elles restent limitées, appelle au bénéfice de chaque doctorant un minimum de formation. Celle-ci, qui peut être évaluée à quelques heures, doit être engagée dès le début de l'élaboration de la thèse et relève manifestement de la compétence de l'école doctorale, qui peut – si elle le souhaite – enrichir les prescriptions techniques minimales évoquées plus haut.

Par ailleurs, il est impératif pour des raisons évidentes qu'il y ait une parfaite adéquation entre la version électronique native de la thèse et la version papier qui constitue le support de la soutenance. La meilleure façon de s'en assurer, tout en vérifiant que les prescriptions relatives à la production de la thèse ont bien été respectées, consisterait à faire éditer par l'établissement la version papier de la soutenance à partir des fichiers électroniques remis par l'impétrant.

Le schéma général de production du document originel peut donc être présenté comme suit :



respect des prescriptions	par le jury	dépôt)

2. Les formats d'archivage et de diffusion

Dans la mesure où chaque thèse sera produite dans le respect des prescriptions requises, il sera possible de lui appliquer une chaîne de traitement automatique propre à la convertir dans un format " structurant " et normalisé permettant son archivage et sa diffusion électroniques. On observera que c'est le reformatage en vue de l'archivage qui précède et conditionne celui prévu pour la diffusion et non l'inverse.

Les *Prescriptions techniques …* présentées ci-après préconisent une conversion dans le format SGML (demain XML) en vue de l'archivage puis de la diffusion, selon le schéma suivant :

Il est clair qu'il convient de mettre en œuvre autant de chaînes de traitement que de formats natifs retenus et de formats cibles. Ces chaînes de traitement ont vocation à être mutualisées.

III. L'ORGANISATION

1. L'organisation cible

Compte tenu . des objectifs assignés,
. des opérations requises,
. des outils techniques retenus qui ont vocation à être répartis,
. de la volonté manifestée par de nombreux établissements,

il a été défini une organisation cible. Résumée dans le tableau suivant, celle-ci est fondée sur deux principes simples :

– il revient à l'Etat ou à un opérateur national de fixer le cadre général, d'établir les prescriptions indispensables à la cohérence du dispositif, de fournir les outils techniques nécessaires et de jouer un rôle d'assistance et de conseil ;
– il revient en revanche aux établissements de soutenance d'assurer le signalement, la valorisation et la diffusion de leurs thèses.

FONCTIONS	**DOCTORANT / DOCTEUR**	**ETABLISSEMENT**	**ETAT / OPERATEUR NATIONAL**
Définition du cadre réglementaire			L'Etat fixe par arrêté les modalités de production matérielle, de dépôt, de signalement, de conservation et de diffusion des thèses.
Production de la thèse électronique native			Un opérateur national élabore, diffuse et tient à jour les prescriptions techniques de production (formats, feuilles de style).
			Il élabore et diffuse des supports de formation.
		L'école doctorale ou un autre service assure la formation des doctorants…	

	Le doctorant saisit ou fait saisir sa thèse sous forme électronique dans le respect des prescriptions	… et fournit une assistance technique pendant la phase de production.	
Edition de la thèse papier en vue de la soutenance	Le doctorant dépose sa thèse sous forme électronique		
		Le service compétent vérifie la qualité des fichiers et le respect des prescriptions.	
		Il procède à l'édition sur papier	
SOUTENANCE		Le chef d'établissement, sur la base du rapport du jury, délivre le grade de docteur	

FONCTIONS	**DOCTORANT / DOCTEUR**	**ETABLISSEMENT**	**ETAT / OPERATEUR NATIONAL**
Eventuellement, traitement des corrections demandées par le jury		*Le jury demande éventuellement des corrections*	
	Le docteur procède ou fait procéder aux corrections et dépose la version électronique de référence		
		Le service compétent procède à une nouvelle édition sur papier	
Autorisation de diffusion de la thèse		Le chef d'établissement, sur avis du jury, autorise la diffusion de la thèse (ou impose des restrictions)	

	Le docteur autorise (ou interdit) la diffusion électronique de sa thèse		
Conservation physique et Archivage électronique		Le service commun de documentation de l'établissement reçoit en dépôt un (ou deux) exemplaire(s) papier à des fins de conservation et de consultation	
			Un opérateur national labellise ou fournit les chaînes de traitement aux établissements
		Le service compétent • met en œuvre les chaînes de traitement • convertit la thèse en format d'archivage • assure un archivage local	

FONCTIONS	**DOCTORANT / DOCTEUR**	**ETABLISSEMENT**	**ETAT / OPERATEUR NATIONAL**

Signalement		Le service commun de documentation procède au signalement, soit par catalogage, soit par extraction des métadonnées : • dans le catalogue local • dans le catalogue collectif • dans des bases spécialisées	
		Il y ajoute l'adresse électronique de la thèse et en assure la mise à jour si nécessaire	Un opérateur national veille à l'homogénéité des adresses électroniques
Diffusion électronique		Le service compétent convertit la thèse en format de diffusion et assure la mise en ligne	
Production/commercialisation de substituts	Le docteur autorise (ou interdit) la production / commercialisation de substituts	L'établissement a vocation à produire / commercialiser des substituts	Un opérateur national peut, par délégation de l'établissement, procéder à la production / commercialisation de substituts

Ce tableau appelle plusieurs commentaires :

a. Sous la rubrique " *Etablissement* " sont impliquées plusieurs instances. A côté des responsabilités essentielles mais traditionnelles du chef d'établissement et du jury, on voit émerger les missions

. de l'école doctorale (formation et assistance techniques)

. du service commun de documentation (dépôt et signalement)

. et d'un opérateur technique, chargé de gérer les fichiers informatiques correspondants, de les convertir dans les formats adéquats et d'éditer la thèse papier. S'il ne convient pas de débattre ici de l'organisation interne des établissements, il est clair que la capacité de cet opérateur à mener ces travaux à bonne fin est essentielle.

b. Même si le parti d'une large décentralisation des responsabilités a été retenu, le rôle de l'opérateur national est déterminant pour garantir la cohérence de l'ensemble, à travers l'élaboration des prescriptions techniques et des supports de formation, la diffusion des chaînes de traitement et l'archivage de sécurité.

c. Le groupe de travail a souhaité souligner l'importance de la formation dispensée à chaque doctorant. Cette étape – évaluée en première approximation entre 3 et 8 heures d'enseignement – est décisive pour la réussite du dispositif tout entier et si l'on ne veut pas allonger les délais – déjà trop longs – entre la fin de la rédaction de la thèse, sa mise en forme matérielle et la soutenance.

d. Le bordereau qui accompagne traditionnellement le dépôt de la thèse par le doctorant devra, bien entendu, être sensiblement réaménagé de façon à faciliter le travail de signalement, d'une part, et à simplifier les procédures de délivrance des autorisations de diffusion du document, d'autre part.

e. Il a paru raisonnable de maintenir le principe du dépôt de la thèse papier au service commun de documentation, à la fois pour permettre une consultation traditionnelle et pour des raisons de conservation. Par ailleurs, si le principe de l'archivage électronique local est affirmé, une duplication de sécurité auprès de l'opérateur national est apparue nécessaire.

f. Par delà un signalement dans le catalogue collectif de l'enseignement supérieur et dans le catalogue local, plusieurs membres du groupe de travail ont appelé l'attention sur l'intérêt que représente un signalement dans des bases spécialisées, très consultées par les différentes communautés scientifiques. C'est l'une des conditions de la valorisation de la science française à l'étranger.

2. La trajectoire ou l'organisation intermédiaire

Le dispositif cible dessiné ci-dessus ne pourra en toute hypothèse être mis en place simultanément par tous les établissements, pour des raisons qui tiennent à la volonté politique de chacun d'entre eux, aux moyens et compétences dont ils disposent ainsi qu'à des questions d'organisation. Pour autant, il est tout aussi clair que le moment est venu d'une mutation et de la mise en œuvre d'un nouveau système qui soit en conformité avec la technologie actuelle, l'attente des usagers et la volonté de la plupart des acteurs.

A ce titre, il est préconisé la démarche suivante :

a. *vérifier l'adhésion de la communauté des établissements au schéma proposé*. A cet égard, il serait souhaitable que la CPU et au premier chef sa commission recherche se prononce sur ce point.

b. *afficher clairement l'objectif cible et mettre en œuvre un environnement favorable*. Cela passe par la rédaction d'un nouvel arrêté relatif aux modalités de production, dépôt, signalement, conservation et diffusion des thèses. Il serait nécessaire par ailleurs que l'opérateur national dont les principales missions ont été détaillées soit en mesure de jouer son rôle.

c. *assurer une information des établissements afin qu'ils prennent la mesure exacte des charges correspondantes*.

d. *mettre en œuvre le nouveau dispositif au fur et à mesure de la volonté manifestée par les établissements.* On peut en effet considérer que les perspectives ouvertes par la diffusion des thèses sur Internet créera un mouvement qui conduira la plupart des établissements à se rallier à cette architecture. Inversement, cela signifie qu'il convient de conserver le dispositif traditionnel pour les établissements qui n'auront pas encore basculé dans le nouveau : constituer une " couche intermédiaire " avec les thèses des établissements qui ne seraient ni dans l'ancien ni dans le nouveau dispositif n'apporterait aucune plus-value mais nuirait fortement à la lisibilité de ce qu'il importe de construire.

e. *accompagner la mise en place de ce dispositif par un programme de formation adéquat*.

3. La question des moyens

Le groupe de travail n'a pas été en situation de chiffrer de façon fine les moyens induits par la mise en place du nouveau dispositif. Il a néanmoins procédé à un premier inventaire :

a. *pour les établissements.* Dans la mesure où la formation des doctorants (évaluée entre 3 et 8 heures) et où une assistance technique seraient assurées de façon solide, la charge moyenne pour un établissement dans lequel 100 à 150 thèses sont soutenues par an est évaluée à un mi-temps d'assistant ingénieur, assurant la vérification de la thèse native, son édition sur papier et sa conversion dans un format d'archivage puis de diffusion. Rien n'interdit, bien entendu, à plusieurs universités de se regrouper pour mutualiser les fonctions d'opérateur technique du processus. Cela ne peut relever toutefois que de leur seule volonté et le groupe de travail n'a pas préconisé la constitution *a priori* d'agences régionales ou académiques qui prendraient en charge ce type de tâches.

b. *pour l'opérateur national*. A ce stade, on observera que plusieurs établissements ou services peuvent avoir vocation à jouer un rôle dans le dispositif à constituer :
– les ateliers nationaux de reproduction des thèses de Lille et de Grenoble qui devront être au moins partiellement reconvertis, leurs moyens étant pour partie redéployés. Même s'il devra être fait appel à d'autres profils et compétences qu'actuellement, ils disposent d'un savoir-faire et d'une familiarité avec les questions relatives aux thèses qui leur confèrent une capacité d'intervention incontestable dans ce domaine. Cette question devra faire l'objet d'un examen d'autant plus attentif que des problèmes de personnel sont en jeu.– l'agence bibliographique de l'enseignement supérieur (ABES), qui est concernée par tout ce qui touche au signalement des documents, en particulier dans les catalogues collectifs.– le centre informatique national de l'enseignement

supérieur (CINES), qui est susceptible d'assurer des fonctions dévolues à un serveur national.

PRESCRIPTIONS TECHNIQUES POUR LE DEPOT DES THESES EN FORMAT ELECTRONIQUE

Ce document résulte du travail effectué par un sous-groupe d'experts animé par Christian Lupovici, directeur du service commun de la documentation de l'université de Marne-la-Vallée. Il s'est appuyé sur les travaux menés par les organismes de normalisation au niveau international et sur les expériences en cours aux Etats Unis, au Canada, en Europe et en France, dans le domaine de la production et de la gestion des thèses en format électronique.

Principes généraux

Comme de nombreuses universités dans le monde l'ont déjà compris, la diffusion électronique des thèses est un formidable outil de valorisation : valorisation des établissements de soutenance bien sûr, mais également valorisation des thèses elles-mêmes que l'ajout de métadonnées appropriées rendra plus facilement repérables sur la Toile.

Pour l'auteur comme pour l'université de soutenance, cette publication électronique rehausse son image dans le paysage de la recherche publique ; elle est également un véhicule efficace de promotion de la recherche publique, notamment dans son désir de partenariat avec le tissu économique international.

Dans ce contexte, la chaîne de traitement des thèses doit être soigneusement définie, car il ne s'agit pas de diffuser uniquement la version électronique d'un document papier, mais d'exploiter au mieux toutes les possibilités qu'offrent les nouvelles technologies (navigation, insertion d'objets 3D, multimédia etc.), pour optimiser les résultats de la recherche. Il faut donc partir de la production électronique des thèses par les auteurs – thèses qui d'ailleurs sont aujourd'hui écrites au moyen des logiciels de traitement de texte, éditeurs structurés, tableurs etc. – pour en faire, non seulement le véhicule de diffusion, mais également de conservation pour un accès à long terme.

Pour ce faire, il est nécessaire de :

- respecter un certain nombre de prescriptions quant au choix et à la structure des éléments de données, des formats de production, de conservation et de diffusion.

- redéfinir la procédure de dépôt et de gestion des thèses dans les universités, en fonction des nouvelles caractéristiques techniques du document électronique.

2. Une exigence de structuration formalisée

Les thèses sont des documents qui respectent une structure dont les éléments sont bien définis (à travers la présentation, dans l'environnement papier). Ces éléments sont repris et formalisés dans la structure du document électronique, qui se décompose comme suit[1] :

- les préliminaires : page de titre, dédicace, remerciements, table des matières, table des illustrations, table des annexes, résumés et mots clés français/anglais et autre langue, éléments d'indexation spécialisée.

- le corps du texte avec ses différentes parties : introduction, chapitres, sections (correspondant à des niveaux de titres), paragraphes, conclusion.

- les postliminaires : bibliographie, glossaire, index, annexes.

Ces différentes parties du document et leurs éléments d'identification doivent être reconnus et typés dès la production du document par le doctorant, au moyen d'un "modèle" de saisie, grâce auquel on pourra :

- passer du document de saisie à un document électronique canonique (version officielle) dans un format qui permettra son stockage et sa conservation à long terme.

- et récupérer automatiquement les métadonnées nécessaires à l'identification, la gestion, la diffusion, la valorisation et la conservation de la thèse.

Cas particulier des thèses qui sont une collection d'articles : le document présentera outre les parties préliminaires et postliminaires, une introduction, un lien vers chaque article pour les intégrer comme autant de chapitres ou pour renvoyer à l'URN (Uniform Resource Name) du document, et enfin, une conclusion.

NB : La norme " Présentation des thèses " de l'ISO 7144, publiée en 1986, est inadaptée pour le document électronique. Les normes relatives aux références bibliographiques (ISO 690 : 1987 et ISO 690-2 : 1997) restent en vigueur.

3. Le modèle de saisie

Pour l'auteur, l'utilisation d'un modèle de saisie ne doit pas être vue comme une contrainte ; les expériences qui sont actuellement menées prouvent au contraire que le modèle est apprécié comme une aide à la rédaction et à la mise en forme du document : les éléments de présentation obligatoires sont prédéfinis et l'auteur n'a plus à s'en soucier. De plus, la formation au modèle lui permet d'utiliser au mieux les fonctionnalités avancées du traitement de texte, telles que l'actualisation automatique de son plan, des listes de tableaux, de figures.., la génération des liens hypertextuels pour exploiter toutes les possibilités de navigation à l'intérieur du document**.**

Les éléments de données essentiels qui sont définis ci-dessous doivent faire partie de la structuration minimum imposée par le Ministère au niveau national. D'autres éléments peuvent éventuellement être ajoutés. Mais seuls les éléments de structure qui correspondent à une DTD normalisée cible pourront être reconnus et récupérés dans le passage du format de traitement de texte au format de conservation en XML.

Le modèle est évolutif. Il tiendra compte, dans ses évolutions, des besoins des auteurs et de l'administration comme de l'évolution des logiciels de saisie. C'est pourquoi il est important que l'on puisse, à tout moment, faire référence à une version de ce modèle datée et "labellisée" par le Ministère de tutelle.

Eléments de données du modèle

Les préliminaires

a. **Les éléments d'identification et de présentation de la page de titre**

Elément de donnée	**O/F**[2]	**R**[3]	**Responsabilité**	**Commentaire**[4]
Etablissement de soutenance	o		Etablissement	[4] Identification / Recherche *Nom de l'université ou de l'établissement*
Composante	f		Etablissement	Identification / Recherche *UFR, Faculté, Institut, Département, Ecole doctorale*
Sous composante	f		Etablissement	Identification / Recherche *Département ou Laboratoire*
Etablissement en co-tutelle	f	ER	Etablissement	Identification / Recherche
Composante	f		Etablissement	Identification / Recherche
Sous composante	f		Etablissement	Identification / Recherche
Nom, prénom de l'auteur	o	R	Auteur	[4] Identification / Recherche *(Nom, prénom de jeune fille, suivi du Nom marital)*
Titre français de la thèse	o		Auteur	[4] Identification / Recherche
Sous-titre français	f		Auteur	Identification / Recherche
Titre Anglais de la thèse	o		Auteur	[4] Identification / Recherche
Sous-titre Anglais	f		Auteur	Identification / Recherche
Titre en autre langue	f		Auteur	Identification / Recherche
Sous-titre en autre langue	f		Auteur	Identification / Recherche
Type de doctorat	o		Etablissement	[4] Identification
Discipline	o		Etablissement	Identification
Date de la soutenance	o		Auteur	[4] Identification
Nom, prénom du directeur de thèse	o		Auteur	[4] Identification
Membres du jury	f	R	Auteur	Information / Recherche

Numéro officiel	o		Etablissement	[4] Identification / Gestion *numéro sur 10 caractères*
Mention de copyright	f		Etablissement	Réglementaire

b.

c. ***Les préliminaires***
d. **Les éléments d'aide à la recherche**

Elément de donnée	**O/F**	**R**	**Responsabilité**	**Commentaire**
Résumé en français	o		Auteur	[4] Recherche textuelle
Résumé en anglais	o		Auteur	[4] Recherche textuelle
Résumé en autre langue	f		Auteur	Recherche textuelle
Mots-clés en français	o		Auteur	[4]Recherche
Mots-clés en anglais	o		Auteur	[4]Recherche
Mots-clés en autre langue	f		Auteur	[4]Recherche
Classification	f	R	Auteur	Recherche *Eléments d'indexation spécialisée :* *seules les classifications reconnues au niveau international pourront être utilisées* *La référence de la ou des classification(s) utilisées devront être indiquées.*

e. ***Les préliminaires***
f. **les autres éléments d'aide à la lecture et personnels**

Elément de donnée	**O/F**	**R**	**Responsabilité**	**Commentaire**
Dédicace	f		Auteur	
Epigraphe	f		Auteur	

Remerciements	f		Auteur	
Table des matières	o		Auteur	*génération automatique*
Liste des figures	f		Auteur	*génération automatique*
Liste des tableaux	f		Auteur	*génération automatique*

Corps du texte : exemple d'éléments de structure hiérarchique

Elément de donnée	**O/F**	**R**	**Responsabilité**	**Commentaire**
Introduction	o		Auteur	
Parties ou Chapitres	o	R	Auteur	
1ère subdivision	o	R	Auteur	
2ème subdivision	f	R	Auteur	
3ème subdivision	f	R	Auteur	
4ème subdivision	f	R	Auteur	
jusqu' à la 9ème subdivision	f	R	Auteur	
paragraphe	f	R	Auteur	
conclusion	o		Auteur	

Les postliminaires

Elément de donnée	**O/F**	**R**	**Responsabilité**	**Commentaire**
Bibliographie	o		Auteur	
Entrée bibliographique	o	R	Auteur	

Annexes	f	R	Auteur	
Titre de l'annexe	f		Auteur	

Eléments administratifs et de gestion du document

Elément de donnée	**O/F**	**R**	**Responsabilité**	**Commentaire**
Code bibliothèque			Etablissement	[4] Utile, au moins pendant la phase de transition, pour identifier le lieu de conservation du document canonique
Autorisation de diffusion par le jury	o		Etablissement	[4]
Autorisation de diffusion par l'auteur	o		Auteur	[4]
Mention de correction	c		Etablissement	[4]Lié à mention diffus. jury
Mention de confidentialité	c		Etablissement	Raisons liées à la confidentialité
Date de fin de confidentialité	c		Etablissement	[4]date

4. Le format de production

Les auteurs utilisent des logiciels de saisie et de mise en page. Ceux-ci ont l'avantage de générer un fichier numérique selon une ergonomie à laquelle l'auteur est habitué, avec des fonctions spécifiques de génération de structure ou de liens. Mais ils ont, en général, l'inconvénient de produire un fichier dans un format qui n'est pas pérenne car lié à une technologie spécifique. De plus, ce format n'est pas obligatoirement le meilleur format de diffusion sous forme électronique, car il est fait pour une impression papier. C'est pourquoi il est nécessaire de reformater ce fichier pour le faire passer du format de traitement de texte à un format de conservation et de diffusion électronique.

Le logiciel utilisé pour la création de la thèse n'est pas imposé au niveau national. Néanmoins il est de la responsabilité des établissements et des écoles doctorales de limiter le nombre de logiciels de saisie pour limiter les chaînes de traitement en aval, au moment du dépôt de la thèse. Les universités qui ont commencé à utiliser une chaîne de production de thèses en format électronique recommandent actuellement, parce que ce sont les plus utilisés, les formats compatibles RTF et LaTeX, à défaut d'un éditeur SGML ou XML.

5. La procédure de dépôt

Le dépôt de la thèse s'effectue sous forme électronique au secrétariat de l'école doctorale ou (en accord avec elle) au Service commun de la documentation ou auprès de toute structure mise en place à cet effet.

L’université met en place un processus d’édition à la demande, pour répondre aux besoins du jury et de l’auteur, en vue de la soutenance.

L’intérêt de cette procédure est de garantir une meilleure qualité des fichiers déposés par les doctorants. L’édition de la thèse sous un format papier à partir d’une structure centrale garantit la conformité des documents de soutenance avec le fichier électronique déposé.

Le document que le jury aura en main lors de la soutenance devra être conforme au fichier électronique déposé.

La procédure de correction éventuellement demandée par le jury, fait l’objet d’un nouveau dépôt et d’une procédure de vérification identique à celle de la phase de soutenance. La thèse est réputée diffusable après l’accord définitif du président du jury (et de l'auteur).

Le document qui sera archivé et signalé doit être conforme au dernier fichier électronique déposé par le doctorant (après modifications visées par le président du jury lorsque cela sera nécessaire).

Le groupe de travail recommande à chaque établissement de dresser une liste limitative des formats de dépôt des thèses qu’il accepte. Plus ce nombre sera élevé, plus il sera fastidieux (voire très difficile) de reformater les documents vers le format XML. La mise à disposition de logiciels standards dans les établissements ainsi que la mise en œuvre d’un " modèle " accrédité doit limiter de fait l’hétérogénéité des formats utilisés par les étudiants.

Ce reformatage vers XML est l’occasion de vérifier la structure du document en fonction d’une structure de référence. Il est également l’occasion d’ajouter les métadonnées nécessaires à la gestion administrative et à la conservation du document et qui ne sont pas contenues dans le fichier d’origine.

De ce point de vue, on peut en profiter pour remplacer les documents papiers qui accompagnent la gestion de la thèse par un " bordereau électronique ".

La procédure de dépôt des thèses sous forme électronique nécessite de modifier l’organisation du circuit actuel des thèses. Si le dépôt de la thèse en format électronique est plus exigeant sur la procédure et la qualité formelle, il ne doit pas ralentir la procédure de soutenance. Le groupe de travail conseille en particulier aux établissements de procéder à l’impression de la thèse pour la soutenance avant le reformatage en XML. Ce reformatage s’effectuera donc après le dépôt et la validation des corrections, et dans tous les cas, après la soutenance.

Les établissements, et particulièrement les écoles doctorales, devront veiller à ce que les auteurs pensent à obtenir l’accord des titulaires des droits pour les documents récupérés et incorporés à leur thèse. Pour les mêmes raisons, ils devront inciter les thésards à réserver leurs droits avant la publication de leurs articles.

Une thèse soutenue et qui a obtenu l'autorisation de diffusion (ou soumise à des restrictions de diffusion pour cause de confidentialité ou de non-autorisation par le doctorant) **doit être *ne varietur*.**

6. La diffusion

Les formats de diffusion électronique (typiquement PDF et HTML) sont adaptés aux plates-formes de lecture et d'impression largement distribués au moment de l'édition du document.

Ces formats sont souvent " propriétaires " et sont, en tout état de cause, soumis aux évolutions de la technique des plates-formes. Ils doivent donc être utilisés uniquement pour la diffusion, mais en aucun cas pour la conservation des documents.

La diffusion de la thèse s'effectue à partir de la forme canonique de conservation.

Les formats électroniques de diffusion peuvent être dans une présentation fixée, en PDF, ou en TIFF, ou dans une présentation dynamique qui fait référence à une feuille de style de présentation (en XSL[5] pour un fichier XML). Le format XML étant prévu pour être utilisé pour le Web, il est recommandé de diffuser la thèse (autant que faire se peut) dans le même format que celui de conservation.

L'université doit mettre en place un processus de diffusion des thèses sur le Web, en Intranet ou sur Internet (selon l'accord de l'auteur et du jury). L'université peut également effectuer une impression sur papier ou sur microforme.

Il faut noter que le grand avantage du format électronique sur le papier, c'est qu'il permet d'élaborer des thèses qui intègrent du texte, des graphiques, des photos, du son, de la vidéo et des images en trois dimensions.

C'est pourquoi il est nécessaire de considérer la forme électronique comme la forme canonique du document et de s'assurer dans la procédure de dépôt qu'elle le reste bien.

7. L'archivage

Les formats de conservation sont des formats indépendants des plates-formes de lecture (typiquement TIFF, SGML, XML) qui pourront permettre de lire le document dans le long terme, sans perte d'information. Au niveau international, que ce soit dans le cadre de l'édition scientifique, ou dans le cadre des bibliothèques, un consensus s'est établi pour adopter XML (eXtensible Mark-up Language) avec une référence explicite aux trois DTDs (Définition du Type de Document) normalisées dans le domaine de la documentation :

- ISO 12083 (pour les livres et articles scientifiques)

- TEI (pour les textes en lettres et sciences humaines et sociales)

- EAD (pour les documents de type archive)

C'est donc sur cette conception du format XML que le groupe de travail fonde sa recommandation d'adopter XML comme format de stockage et de conservation de la thèse électronique. Ce format permet aussi la gestion de documents composites puisqu'il peut référencer d'autres éléments de contenu que le texte tels quedu son, des images fixes ou animées. De même, XML pourra être utilisé pour " encapsuler " des documents d'autres formats (formats images résultant de la numérisation, par exemple) lorsqu'il n'est pas nécessaire de gérer un format en mode caractères ou balisé. Cette " encapsulation " permet d'indexer le document au niveau des parties qui le composent ou d'éléments spécifiques en ajoutant les métadonnées correspondantes.

Le passage du " modèle " de thèse normalisé (en format compatible RTF ou en LaTeX) à la DTD cible en XML est effectué grâce à des " scripts " de conversion, diffusés gratuitement à la communauté universitaire.

L'Université chargée de la conservation de la thèse électronique sera amenée à ajouter des métadonnées de gestion de la conservation (notamment pour une ultérieure émulation). Elle sera également amenée à reporter le document sur des supports de stockage successifs[6]. Elle ne peut donc s'engager sur la pérennité de la présentation d'origine, sauf quand celle-ci est signifiante au point d'avoir été balisée comme telle par l'auteur.

8. Métadonnées et signalement des thèses

Les métadonnées sont les éléments qui servent à identifier et à décrire le document, en l'occurrence, la thèse (ou une de ses parties), pour en faciliter la gestion, la recherche et l'accès.

Dans la mesure où les éléments d'identification et de description figurent dans le fichier de production (cf. les données préliminaires), ils pourront être extraits du document par des procédures automatiques, puis enrichies si nécessaire pour alimenter les catalogues signalétiques, que ce soient des catalogues collectifs (indexation Rameau pour le SU par exemple), des catalogues thématiques ou le catalogue propre à chaque établissement.

Les métadonnées pour les thèses proposées dans le tableau ci-dessous ont pour référence le "Dublin Core" ; elles correspondent au dernier résultat (janvier 2000) d'un travail de normalisation internationale.

Tableau des métadonnées préconisées pour les thèses

Elément	*Qualificatif*[7]	*Vocabulaire*	*"Scheme"*	*Langue*	*Commentaire*
DC.Contributor	contributorType contributorName contributorRole	DCAT1[8]	FNF[9]		"person" Nom, prénom du directeur de thèse "Directeur"
DC.Contributor	contributorType contributorName contributorRole	DCAT1	FNF		"person" Nom, prénom des membres du jury et rapporteurs selon leur rôle *zone à répéter autant de fois que de co-tutelles*
DC.Contributor	contributorType contributorName agentRole	DCAT1			"org" Nom de l'établissement, composante, sous composante "Université de soutenance"
DC.Contributor	contributorType contributorName	DCAT1			"org" Nom de l'établissement, composante, sous composante "co-tutelle"

	contributorRole				*zone à répéter autant de fois que de membres de jury*
DC.Coverage					
DC.Creator	creatorType creatorName	DCAT1	FNF		"person" Nom, prénom de l'auteur *zone à répéter si plusieurs auteurs*
DC.Date	issued		W3CDT[10]		date de soutenance
DC.Date	available		W3CDT		date d'autorisation de diffusion de la thèse
DC.Description	abstract			fre	Résumé français
DC.Description	abstract			eng	Résumé anglais
DC.Description	abstract			*selon la langue*	Résumé en une autre langue
DC.Description	table of contents				Table des matières de la thèse
DC.Format	medium	IMT[11]			ex "text/xml"
DC.Format	extent				ex."3419 bytes"
DC.Identifier			URI[12]		URN de la thèse en texte intégral
DC.Identifier			No officiel		No de la thèse attribué par l'université
DC.language			RFC1766		langue de la thèse, par défaut "fre"
DC.Publisher	publisherType publisherName	DCAT1			"org" Université responsable de l'édition électronique de la thèse
DC.Relation					
DC.Rights					indique les modalités de diffusion de la thèse
DC.Rights					mention de copyright
DC.Source					Mention d'origine du document
DC.Subject		auteur	mots clés	fre	Mots clés français de l'auteur *(utiliser le **;** comme séparateur de mots clés)*
DC.Subject		auteur	mots clés	eng	Mots clés anglais de l'auteur *(utiliser le **;** comme séparateur de mots clés)*
DC.Subject		auteur	mots clés	*selon la langue*	Mots clés de l'auteur dans une autre langue

					*(utiliser le **;** comme séparateur de mots clés)*
DC.Subject		Rameau FmeSH *etc.* *selon le vocabulairede référence*		fre	Mots clés français conformes au thésaurus Rameau ou au MeSH en français *(utiliser le **;** comme séparateur de mots clés pour un même vocabulaire de référence.* *répéter la zone si le vocabulaire de référence est différent)*
DC.Subject	classification	MSC[13] PACS[14] *etc.* *selon le vocabulairede référence*			équivalent du code de classification sur le bordereau thèse ou pour un autre type de classification référencée *(utiliser le **;** comme séparateur de mots clés pour une même classification.* *répéter la zone si la classification de référence est différente)*
DC.Title				fre	Titre et sous titre de la thèse en français
DC.Title				eng	Titre et sous titre de la thèse en anglais
DC.Title				*selon la langue*	Titre et sous titre de la thèse en une autre langue que le français et l'anglais
DC.Type		DCT1[15]			"text"
DC.Type		TéléThèse			type de diplôme

Il est clair que la diffusion électronique doit favoriser une plus grande publicité et un plus grand accès aux thèses. Pour cela, le signalement d'une thèse soutenue doit pouvoir s'effectuer très rapidement. Chaque thèse soutenue est signalée dans le catalogue collectif national du S.U., soit par une extraction automatique des métadonnées reformatées en une notice Unimarc téléchargée sur le S.U., soit par un catalogage original.

La thèse faisant partie du fonds " patrimonial " de l'université de soutenance, son signalement figurera obligatoirement dans le système d'information local et sera accessible sur le Web de l'université où le document lui-même pourra éventuellement (selon la décision du jury et l'autorisation de l'auteur) être accessible aux moteurs de recherche internationaux grâce aux métadonnées de signalement appropriées.

En outre, il est recommandé aux établissements de participer à des catalogues thématiques internationaux pour augmenter la visibilité et la valorisation de leurs thèses.

9. La nécessaire réorganisation du circuit de traitement des thèses

Le groupe de travail attire l'attention du Ministère sur la nécessaire concentration de moyens informatiques (logiciels et matériels) et en ressources humaines (avec les problèmes de

requalification) pour effectuer la vérification de la structure des thèses, leur édition pour la soutenance, la conversion en format de conservation et publication sur le Web (en Intranet ou sur l'Internet).

Dans la perspective d'une recommandation nationale pour la mise en place du dépôt électronique des thèses, il est préconisé de faire une étude théorique du " workflow " pour que chaque établissement puisse positionner son organisation spécifique par rapport au schéma théorique.

ANNEXES

Annexe n°1 : Composition du groupe de travail

Direction de l'Enseignement Supérieur – Sous-direction des bibliothèques et de la documentation :

Claude JOLLY	Chargé de la sous-direction des bibliothèques et de la documentation
Chantal FRESCHARD	Chef du bureau de la modernisation des bibliothèques
Charlette BURESI	Conservateur au bureau de la formation, de l'édition et des systèmes d'information
Christine OKRET	Conservateur au bureau de la modernisation des bibliothèques
Pierre-Yves RENARD	Conservateur au bureau de la coordination documentaire
Pascale VIGIER	Conservateur au bureau de la modernisation des bibliothèques

Direction de la Recherche :

Micheline BELIN	Chef du bureau des formations doctorales, des écoles normales supérieures et des écoles françaises à l'étranger (sous-direction de la recherche universitaire et des études doctorales)
Didier ARQUES	Directeur scientifique de l'informatique (mission scientifique universitaire)
Patrick BRASART	Coordinateur auprès de la directrice des sciences de l'homme et des humanités (mission scientifique universitaire)

Direction de la Technologie :

Jacques GUIDON	Expert enseignement supérieur et à distance au bureau de l'enseignement supérieur (sous-direction des technologies éducatives et des technologies de l'information et de la communication)

Association des Directeurs de la Documentation et des Bibliothèques Universitaires (ADBU) :

Jacqueline GAUDE	Directrice du service commun de la documentation de l'université de Nancy 1
Brigitte MULETTE	Directrice du service commun de la documentation de l'université de Lille 2

Conférence des Présidents d'Université (CPU) :

Gérard CHARBONNEAU	Vice-président de l'université de Paris Sud
Serge WOLIKOW	Vice-président de l'université de Bourgogne

Université Lumière Lyon 2 :

Jean-Paul DUCASSE	Chargé de mission auprès de la présidence de l'université Lumière Lyon 2

Institut National des Sciences Appliquées (INSA) de Lyon :

Monique JOLY	Directrice de Doc INSA

Atelier National de Reproduction des Thèses (ANRT) - Lille :

Elisabeth FICHEZ	Professeur d'université, directrice de l'ANRT

Université Joseph Fourier - Cellule Mathdoc :

Pierre BERARD	Professeur, directeur de la Cellule MathDoc

Université Marne-La Vallée :

Christian LUPOVICI	Directeur du service commun de la documentation de l'université de Marne La Vallée

Centre Informatique National de l'Enseignement Supérieur (CINES) :

Alain QUERE	Directeur du CINES
José SANCHEZ	Ingénieur au CINES

Agence Bibliographique de l'Enseignement Supérieur (ABES) :

Florence ROBERT	Conservateur

Laboratoire Lorrain de recherche en Informatique et ses Applications (LORIA) :

Jacques DUCLOY	Ingénieur

***

Ont également participé aux travaux de ce groupe en qualité d'experts :

Odile ARTUR	Ingénieur au service commun de la documentation de l'université de Marne La Vallée
Frédérique BLONDELLE	Ingénieur à l'ABES
Viviane BOULETREAU	Responsable édition électronique à l'université Lumière Lyon 2
Jean-Michel MERMET	Ingénieur d'études à Doc INSA (Lyon)
Brigitte PRUDHOMME	Assistant ingénieur à Doc INSA (Lyon)

Annexe n°2 : Documents préparatoires

ELEMENTS D'UN DISPOSITIF DE DIFFUSION

DES TRAVAUX DE DOCTORAT SUR LE RESEAU INTERNET

INTRODUCTION

Selon les statistiques du MENRT (direction de la Recherche)[16], environ 11 000 thèses ont été soutenues en 1997. Bien que leur nombre connaisse une légère décroissance, les thèses de sciences dures représentent environ la moitié du total des thèses soutenues, celles de sciences humaines et sociales un tiers (en augmentation) ; et celles des disciplines de santé environ un cinquième (stabilité). Toutes disciplines confondues, depuis 1992, autour de 10 000 thèses en moyenne sont soutenues par an. Il est estimé que le flux de docteurs pour les années à venir devrait rester stable.

Ces chiffres offrent une idée de l'importance du gisement scientifique produit dans les universités françaises qu'il convient de mettre en valeur. Il appartient au MENRT de veiller à la diffusion de ces travaux.

Le signalement des thèses soutenues se fait par l'intermédiaire de pôles de signalement, qui saisissent les notices de thèses dans la base de données Téléthèses.

Les procédés institutionnels de diffusion des thèses soutenues dans les universités et grands établissements français reposent sur deux circuits parallèles[17]. En premier lieu les thèses soutenues sont déposées sous forme papier dans la bibliothèque de leur université de soutenance, où elles peuvent être consultées sur place, et prêtées par le biais du réseau de prêt entre bibliothèques. En second lieu, un exemplaire papier est transmis aux ateliers nationaux de reproduction des thèses[18] chargés du microfichage des textes et de la diffusion systématique de ces microfiches auprès de toutes les bibliothèques universitaires. Ces deux formes de diffusion non commerciale présentent en regard des exigences des chercheurs trois inconvénients majeurs : la diffusion des thèses est géographiquement restreinte, puisque limitée aux bibliothèques universitaires ; le support microfiche est jugé peu convivial[19] ; le prêt entre établissements est parfois trop lent ou onéreux[20], ce qui peut décourager les utilisateurs qui désirent consulter sur papier les thèses non conservées dans leur bibliothèque universitaire.

Cette brève description témoigne de la complexité d'un circuit lent et peu efficace. Sa réorganisation apparaît indispensable.

La réalisation du nouveau catalogue collectif des bibliothèques de l'enseignement supérieur prévoyant un dispositif de fourniture de documents à distance (Système Universitaire de Documentation) va à court terme permettre une simplification et une plus grande rapidité du

signalement des thèses. Cette tâche sera désormais dévolue aux bibliothèques, qui cataloguerons elles-mêmes les thèses soutenues dans leur établissement, ce qui permettra un signalement plus rapide.

Face aux insuffisances de la procédure de diffusion, l'utilisation du réseau Internet offre la possibilité de multiplier les accès indépendamment des contingences physiques, de permettre une consultation rapide et souple grâce aux possibilités de navigation hypertexte. Une valorisation efficace des travaux scientifiques français doit donc s'appuyer sur les avantages du réseau, afin de leur assurer une visibilité accrue[21].

Le présent document a pour objectif de présenter les problématiques attachées à l'organisation de la diffusion numérique des thèses, et de présenter diverses solutions envisageables afin de définir les contours d'un circuit de diffusion efficace, fondé sur la mise en ligne des thèses courantes soutenues, produites soit sous forme papier, soit sous forme électronique.

1. Objectifs

a) Principe d'exhaustivité de la diffusion

Hormis les thèses d'exercice de médecine, dont la valeur scientifique est généralement considérée comme faible, toute thèse soutenue a vocation à être reproduite. Sa diffusion sur Internet est possible si deux conditions sont réunies :

- La décision prise lors de la soutenance par le jury d'accorder l'autorisation de reproduire la thèse (voir l'arrêté du 25 septembre 1985)
- L'accord de l'auteur

  Ce double accord reflète l'ambivalence du statut de la thèse, à la fois document administratif et œuvre soumise aux règles de la propriété intellectuelle.

- Toute thèse jugée reproductible pour laquelle l'auteur aurait donné son accord doit-elle être mise sur Internet par principe ?[22]

b) Durée de disponibilité des thèses sur le réseau

Selon les statistiques émises par le MENRT (direction de la Recherche), environ 10 000 thèses en moyenne sont soutenues chaque année (11 000 en 1997).

Les informations contenues dans une thèse ont une durée de validité scientifique variable selon les disciplines. Il est généralement admis que les thèses de science ont une durée de vie plus courte que les thèses de sciences humaines.

- Quelle durée maximale de disponibilité immédiate des thèses sur l'Internet peut-on recommander, compte tenu de ces spécificités disciplinaires ?

c) Texte intégral ou éléments significatifs d'une thèse

Les modèles d'accessibilité actuellement proposés aux auteurs dans les projets de diffusion de thèses électroniques existants illustrent deux approches. Chacun des éléments de l'alternative répond à un objectif différent.

- **Recommander le texte intégral**

Deux exemples illustrent ce modèle :

- La thèse est mise en ligne dans son intégralité, et les **restrictions d'accès** offertes au choix de l'auteur portent **sur l'étendue géographique de la diffusion**.

Dans ce cadre, le modèle américain (Virginia Tech) propose trois options : accès libre, accès au seul campus limité à un an renouvelable année par année, accès interdit pour une année (cas d'informations confidentielles)[23]. La West Virginia University offre des possibilités similaires[24].

L'INSA de Lyon offre également trois possibilités : intranet, extranet (réalisé avec des partenaires privilégiés, encore inexistant), Internet. L'auteur conserve la possibilité de retirer à l'établissement le droit de diffuser sa thèse sur Internet.

- L'université Lyon 2 propose un **modèle** avec des restrictions d'accès **de même type, mais nuancées par des limitations apportées à l'accès au contenu de la thèse** : l'accès aux thèses numériques est soit libre sur Internet (mise en ligne rapide ou différée pour des contenus sensibles), soit limité à l'intranet de l'université. Des restrictions concernant l'usage des fichiers accessibles sur Internet sont possibles en activant les options de protection contre le copier/coller et l'impression disponibles pour le format PDF (visualisation seule permise).

Recommander le texte intégral sur le plan national ? :

Une solution contraignante paraît difficile à envisager, compte tenu de la législation sur la propriété intellectuelle. En revanche, une forte incitation en ce sens présenterait un double avantage : serait ménagée la meilleure accessibilité aux auteurs ; le dispositif gagnerait en visibilité. Cette solution est-elle acceptable pour la très grande majorité des chercheurs aujourd'hui ? Deux éléments indicatifs peuvent guider une réflexion sur cette question :

- Une enquête portant sur les pratiques des utilisateurs de thèses[25] conclut qu'il n'y a pas de réticence majeure des auteurs ou des futurs auteurs à la diffusion de leur thèse sur Internet (85 % de réponses favorables), en particulier chez les enseignants-chercheurs (à 90 %) et chez les utilisateurs en sciences (à 89 %).

- L'INSA de Lyon demande le dépôt de la thèse sous version électronique (en plus des exemplaires papier) et fait signer une autorisation de mise en ligne. Une seule thèse de 1999 répond actuellement aux critères de mise en ligne sur le site de l'INSA (dépôt de la version électronique *et* autorisation signée)[26]. L'obstacle juridique apparaît réel.

- **Laisser aux auteurs le choix entre une diffusion de l'intégralité ou des seuls éléments significatifs du texte**

  Le projet Webthèse offre ce choix. A titre purement indicatif, sur 2 500 demandes d'autorisation, 1 144 ont été renvoyées à l'ANRT, soit 46 % de l'échantillon.

- 4 % des auteurs ont refusé la mise en ligne de leur thèse. Ce pourcentage, très faible, confirme l'intérêt des docteurs pour ce mode de diffusion de leurs travaux.

- 74 % des réponses sont exploitables, soit un total de 842, réparties ainsi :

  - 61 % (517) des auteurs choisissent la mise en ligne de leur texte intégral, dont 53 % permettent l'impression

  - 39 % (325) des auteurs optent pour la mise en ligne des parties significatives de leur thèse. Parmi ceux-ci, 37 % excluent la bibliographie.

A la lumière de ces résultats, il semble que les auteurs soient tendanciellement favorables à une diffusion large de leurs travaux.

Laisser le choix sur le plan national :

Cette possibilité permettrait d'obtenir les autorisations d'auteurs qui souhaitent ne pas mettre en libre accès l'intégralité de leur thèse pour des raisons d'édition ou la crainte d'un pillage, et qui sont susceptibles de refuser une mise en ligne si celle-ci n'est pas partielle. L'adoption de cette possibilité traduirait une certaine prudence : dans un premier temps, cela permettrait de juger de la réaction des auteurs à une échelle plus significative (nationale), éventuellement en fixant un seuil de réponses favorables au-delà duquel la mise en ligne du texte intégral pourra peut-être à terme être imposée. Mais elle est contraignante pour une numérisation rapide de larges quantités de thèses.

d) Usages des fichiers numériques

Afin de faciliter la navigation à l'intérieur du texte de la thèse et de permettre des traitements de l'information contenue (recherche d'occurences par ex.), il convient de numériser les documents en mode texte.

e) Diversification des supports de diffusion

La version papier de la thèse est communicable dans les bibliothèques qui la conservent ou par prêt entre bibliothèques (PEB).

- Parallèlement aux supports papier et électronique, est-il opportun de fournir à la demande les thèses sur d'autres supports ?

Deux supports additionnels sont envisageables :

- **Microfiches :**

Actuellement, les thèses microfichées sont diffusées systématiquement auprès des bibliothèques. Ce mode de consultation n'est pas apprécié par les chercheurs[27].

- Est-il souhaitable de proposer une diffusion de microfiches sur commande ?

- **Ouvrages facs-similés :**
- Actuellement, l'ANRT de Lille offre l'exemple d'un établissement qui a passé un accord avec des presses universitaires (celles du Septentrion) pour assurer à partir des fichiers numérisés par l'ANRT une production de fac-similés des thèses pour lesquelles les auteurs ont donné leur autorisation. Cet accord est très profitable pour les PUS : le nombre de contrats signés par les docteurs est passé de 552 au 15 novembre 1997 à 773 au 20 novembre 1998, et à 1908 au 30 octobre 1999. Le nombre d'exemplaires vendus pour ces trois années est respectivement de 1877, 3357, 5916[28].

- Est-ce un modèle reproductible par les universités qui souhaitent numériser les thèses soutenues dans leur établissement ?

f) L'organisation de l'archivage

Quel choix effectuer afin d'assurer une conservation à (long) terme des thèses ?

Les différentes possibilités techniques sont les suivantes :

- Le papier : conservation d'un exemplaire par les BU, en plus de l'exemplaire dédié à la consultation et au PEB.

- La conservation de deux exemplaires en BU a-t-elle conservé sa pertinence ?

- La microforme : utilisé actuellement par les ANRT. Ce support est reconnu comme ayant une longue durée de conservation.

- Le format électronique. Voir la partie " Choix techniques ".

2. Choix techniques

Les choix techniques sont à opérer en fonction :

- Du support de dépôt des thèses : fichiers informatiques ou papier
- De l'objectif des opérations de numérisation : diffuser ou archiver

a) Thèses électroniques

Il est nécessaire de distinguer plusieurs usages : le format natif pour la rédaction de la thèse, la diffusion de la thèse, l'archivage des fichiers.

- ***Les formats natifs :***

Les thèses sont essentiellement rédigées et remises en format Word (PC, Mac), RTF et Wordperfect dans une moindre mesure, ou LaTeX (format spécifiquement dédié à la rédaction de travaux scientifiques comprenant des formules).

Les thèses comprenant des éléments multimédia (fichiers son ou images animées) sont encore rares. Les conséquences techniques d'un développement possible de ce type de travaux doit être évalué : quelle influence sur le format de diffusion choisi ? (HTML de préférence à PDF ?).

La structuration des thèses électroniques : les feuilles de style

Les universités de Lyon 2 et Marne-La-Vallée, ainsi que l'INSA de Lyon proposent une feuille de style à leurs étudiants.

Est-il pertinent de faire de telles préconisations au niveau national ?

- La création de feuilles de styles répond au souci de recueillir des travaux correctement structurés afin de permettre un traitement simplifié et rapide des fichiers pour la diffusion sur Internet.

Une recommandation nationale pourrait définir un certain nombre d'éléments obligatoires devant être intégrés à ces feuilles (ex. page de titre).

Plus généralement, l'expérience des établissements déjà confrontés à la remise de thèses électroniques offre un bon point de départ pour augmenter le guide de présentation des thèses du MENRT de préconisations spécifiques liées à ce type de document (format natif ...).

Les chaînes de traitement informatisé des fichiers de thèse

* Exemple de l'INSA de Lyon : conversion semi-automatique de fichiers Word, LaTeX et Postcript en fichiers PDF.

* Chaîne de conversion utilisée par le projet des Presses universitaires de Montréal et l'université de Lyon 2 pour transformer les fichiers de

traitement de texte bien structurés en SGML, puis HTML, XML, PDF (indépendamment de la mise en forme initiale).

Ces chaînes de traitement ont vocation à être mutualisées.

- **Les formats de diffusion**

Plusieurs formats sont proposés actuellement.

PDF : format le plus utilisé

Les avantages :

Forme fac-similé de la thèse : respect du choix de présentation de l'auteur, aucune modification de forme, respect de la version canonique papier validée par le jury, pas de difficulté pour faire apparaître les graphiques.

Très largement diffusé, avec un logiciel de lecture gratuitement déchargeable qui fonctionne sur la quasi-totalité des plates-formes informatiques.

Format compact.

Protection des fichiers possible (contre la modification, le copier/coller, l'impression).

Mode d'affichage ergonomique et fonctionnalités d'aide à la recherche incluses dans le reader (fonction recherche, visualisation possible de la structure des documents pour une consultation plus souple ...).

Bonne qualité d'impression.

Les inconvénients :

Format propriétaire

HTML : format natif du Web, utilisation répandue

Les avantages :

Format normalisé, indépendant des plates-formes informatiques

Les inconvénients :

Il existe plusieurs versions de HTML et des balises propriétaires.

Structuration par paragraphes et niveaux de titres, pas de structuration fine.

Format d'affichage ; qualité d'impression non contrôlée.

Postcript : format le plus ancien

Format proche de PDF (même origine : Adobe), mais très volumineux.

Utilisé à l'université Humboldt de Berlin.

XML : format le plus récent, quelques essais

Les avantages :

Structuration des documents réalisée de manière indépendante de la représentation de cette structure (ex. archivage en XML, diffusion en PDF ou HTML possible) – format pivot.

Format normalisé.

Codage de l'information en UNICODE : portabilité et relecture aisées.

Les navigateurs n'intègrent pas encore la possiblité de lire ce format très prometteur de façon conviviale (sans affichage des balises)[29].

Quelques thèses sont mises en ligne dans ce format sur le site de l'université Lyon 2.

SGML

Peu utilisé comme format de diffusion ; il est plus lourd à manier que XML. Les thèses numériques visibles sur les serveurs de l'université de Lyon 2 sont présentées en SGML. Mais il est actuellement un format structurant pivot efficace.

- **Les formats d'archivage**

D'une manière générale, les formats structurants, pivots – SGML, XML à terme -, sont les plus intéressants pour conserver les documents parce qu'ils offrent la possibilité de convertir les fichiers en format de diffusion, et sont normalisés.

SGML

Utilisé par Lyon 2, Marne-La-Vallée et les Presses universitaires de Montréal, avec archivage local et accès croisé pour les participants au projet.

PDF

Utilisé par Rhodes University, et par l'Australian Digital Theses Project.

Multiplicité des formats de conservation : INSA de Lyon

Conservation du format natif, du format source (format natif légèrement altéré en prévision de la conversion en PDF), des fichiers comprenant des éléments extérieurs au texte de la thèse (photographies, plans, diapositives ...), des fichiers PDF avant application des protections, des fichiers issus de la chaîne de conversion.

Moyen de conservation : le cédérom en 3 exemplaires, remplacé tous les 10 ans.

- Au vu de la diversité des solutions techniques adoptables pour la rédaction, la diffusion et l'archivage des thèses électroniques, est-il pertinent de prévoir des préconisations techniques au plan national ?

b) Thèses papier

Les thèses sont actuellement déposées sous forme papier. Même si le dépôt sous forme électronique est appelé à se généraliser, on ne peut pour l'heure exclure la persistance de thèses remises sous forme papier uniquement.

- Doit-on les numériser ?

Très peu de projets recensés, y compris au plan mondial, concernent la conversion des thèses papier. Il en existe trois : Webthèses (France), Melbourne (élément du Australian Digital Theses Project), université de Montréal (Canada).

- **Les formats de diffusion**

Tous les projets ont opté pour une diffusion en **PDF**.

Note : existence d'une chaîne de conversion à l'université de Montréal avec utilisation du logiciel Ariel (même principe que le projet Webthèses : passage par le format TIFF avant la conversion définitive en PDF).

- **Les formats d'archivage**[30]

Essentiellement en TIFF / PDF.

Pour le projet Webthèses, les fichiers TIFF préalables à la conversion en PDF sont conservés à l'ANRT de Lille. Les fichiers PDF résultant de la conversion en PDF sont conservés à l'ANRT de Lille et au CINES.

3. L'organisation

De façon croissante, les universités souhaitent valoriser les ressources scientifiques produites dans leur établissement, accroître leur accessibilité pour élargir leur visibilité. Cette tendance est appelée à s'accentuer au cours des années à venir. Les missions jusqu'ici dévolues à l'Etat en matière de diffusion des travaux scientifiques sont appelées à se transformer en conséquence.

En récapitulant les différentes étapes d'une chaîne de numérisation des thèses, sont examinés les différents scénarios possibles de partage des responsabilités entre les établissements et l'Etat.

a) Le dépôt d'une thèse électronique

Les thèses doivent encore aujourd'hui être remises sous forme papier. Les doctorants rédigeant désormais dans leur grande majorité leur thèse sur un ordinateur, leur demander de **rendre une version électronique de leur travail** paraît légitime.

Afin d'assurer l'exacte adéquation entre la version papier soutenue et validée par le jury et la version électronique, il est alors indispensable que les services de l'université se chargent du tirage de cette version papier à partir de la version électronique déposée par l'étudiant.

La même procédure est à envisager si le doctorant doit apporter des corrections à son texte sur demande du jury.

- Le principe retenu est celui du dépôt de la version électronique en format natif (conformément aux recommandations) puis du déclenchement du processus de conversion de ces fichiers dans le format choisi pour l'archivage / la diffusion. Cependant, pour donner davantage d'efficacité à ce processus, il est prévu de former les doctorants à une présentation plus rigoureusement structurée de leur travail (usage de feuilles de style), et ce dès leur première année de doctorat.

b) La numérisation des thèses

Il apparaît nécessaire au préalable de recenser précisément les projets en cours de réalisation ou sur le point d'être réalisés. Selon l'enquête sur les projets de numérisation réalisée début 1999 par le MENRT (SDBD), sur huit projets cités, cinq étaient encore à l'étude. Trois seulement étaient en cours de réalisation. Une actualisation de cette enquête est en cours. A ce jour, si les thèses sont considérées comme un gisement scientifique mal exploité, les établissements d'enseignement supérieur ont produit peu de projets de numérisation et de diffusion électronique de ce type de documents.

Cette constatation induit deux conséquences :

- Si la numérisation est considérée comme du ressort de l'université, il sera nécessaire d'encourager la définition de projets précis et leur réalisation.

- Avant que tous les établissements se soient dotés du matériel nécessaire et aient affecté du personnel à un projet de numérisation, se trouve une période transitoire qu'il convient de gérer.

    - Deux solutions sont possibles :

    - Soit les universités progressent à leur rythme sans intervention de l'Etat,

    - soit l'Etat ou un opérateur national prend en charge la numérisation des thèses des universités qui le souhaitent en attendant qu'elles reprennent cette tâche elles-mêmes si elles le souhaitent, selon des modalités qui restent à définir.

      Pour les opérations de numérisation des thèses, quel que soit l'opérateur choisi, il sera nécessaire de passer un accord (" cahier des charges " ?) avec les établissements concernés. Cet accord pourrait aborder les points suivants :

- **La garantie d'un accès permanent aux textes.** Les BU (ou un éventuel opérateur) devraient gérer les éventuels changements d'URL de leurs thèses en les mettant à jour dans le SU.

- **Selon les voies choisies, le partage des responsabilités concernant l'archivage,** et en particulier, si l'archivage est réalisé par l'Etat, la compatibilité des choix techniques réalisés par l'université avec les nécessités de l'archivage.

- Il faudra prévoir dans chaque établissement une **validation de la forme**, afin de s'assurer que la thèse est complète (aucun fichier manquant).

  Il ressort des projets américains observés que la mise en place d'une chaîne de diffusion efficace dépend d'une coopération des instances de direction avec la bibliothèque (signalement) et le centre de ressources informatiques (compétences techniques).

  c) La signalement

  Le signalement des thèses soutenues dans les catalogues collectif et local doit être effectué par les BU, y compris la saisie de l'URL d'accès aux fichiers numériques.

  Si un opérateur national se charge de l'opération de numérisation, il serait plus simple qu'il puisse signaler dans le catalogue (national) l'URL de la thèse.

  d) La diffusion

  Si les établissements prennent en charge la numérisation des thèses, les fichiers pourraient être accessibles depuis leur site web.

Pour les établissements qui ne souhaitent pas se charger de la diffusion, pourrait être envisagé un site maintenu par un opérateur national, l'essentiel étant que l'accès aux thèses électroniques soit possible depuis le catalogue du Système Universitaire.

e) L'archivage des fichiers numériques

Quel partage des tâches est envisageable entre l'Etat et les établissements pour l'archivage numérique ?

* 1° solution : L'Etat peut jouer ce rôle d'archivage des thèses, dans le prolongement de sa mission actuelle de conservation des thèses sous forme de microformes.

La durée de disponibilité en ligne d'une thèse dépend de la fréquence de sa consultation. L'Etat peut éventuellement archiver les thèses soutenues pour lesquelles la mise en ligne excède un certain nombre d'années (à définir).

* 2° solution : L'archivage partagé : de façon systématique, les établissements transmettent à l'Etat leurs archives numériques de thèses, mais en détiennent une copie. La copie conservée par l'Etat serait une sorte de copie de secours.

Dans ce cas, les établissements doivent-ils transmettre à l'Etat les fichiers sous un format numérique approprié, ou sous format natif, à charge pour l'Etat d'effectuer la conversion dans le format d'archivage souhaité ?

* 3° solution : les établissements se chargent de l'archivage, ce qui nécessite une capacité importante de stockage de documents mis en ligne.

- Pour une mise à disposition des archives, il est nécessaire de définir des modalités d'accès : A la demande (procédure de commande à mettre en place) ? Délai de mise à disposition en ligne ? ...

f) Confection de produits de substitution

Lors de la première réunion du groupe de travail a été affirmée la complémentarité d'autres supports (papier, microfiche) avec l'électronique, et a été soulignée l'importance de pouvoir répondre à des demandes ponctuelles des utilisateurs.

- Qui doit prendre en charge la fabrication et la diffusion de produits de substitution confectionnés à partir des fichiers numériques ?

FONCTIONS	**ACTEURS**

	DOCTORANT (DOCTEUR)	ETABLISSEMENT	OPERATEUR NATIONAL
- *Mise en forme de la thèse*			• Etablit des prescriptions de format, feuille de style • Diffuse les supports de formation
		• Forme le doctorant	
	• Saisit la thèse • La dépose sous forme électronique		
- *Tirage de la thèse papier pour la soutenance*		• Prend cette opération en charge, avec les navettes des corrections	
- *Soutenance*	• Autorise la numérisation de sa thèse après accorddu jury		
- *Numérisation*			• Fournit la chaîne de traitement aux établissements
		• Met en œuvre la chaîne de traitement des fichiers • Archive en format pivot (SGML, XML) et diffuse (XML, PDF ...)	• Apporte une assistance technique

- *Signalement*		• La BU catalogue la thèse dans les outils national et local, et saisit l'URL de la thèse numérisée	
- *Conservation / Archivage*		• Archive sous forme électronique • Un exemplaire papier est déposé à la BU	• Fait un archivage de sécurité
- *Production de substituts et commercialisation*	A déterminer		

Annexe n°3 : Relevés de conclusions des réunions de travail

GROUPE DE TRAVAIL SUR

LA DIFFUSION ELECTRONIQUE DES THESES

Réunion du 10 février 2000 : OBJECTIFS

Paris, MENRT

Présents :

MENRT – DES : Charlette Buresi, Chantal Freschard, Claude Jolly, Christine Okret, Pierre-Yves Renard, Pascale Vigier
MENRT – DR : Micheline Belin, Didier Arques (représenté), Patrick Brasart
ADBU : Jacqueline Gaude, Brigitte Mulette
CPU : Gérard Charbonneau
Université Lumière Lyon 2 : Jean-Paul Ducasse, Viviane Bouletreau
INSA Lyon : Monique Joly
ANRT : Elisabeth Fichez
Université Joseph Fourier - Cellule Mathdoc : Pierre Bérard
Université Marne – La Vallée - SCD : Christian Lupovici
CINES : Alain Quéré, José Sanchez
LORIA : Jacques Ducloy

Excusés :

ABES : Florence Robert
MENRT – DT : Jacques Guidon
CPU : Serge Wolikow

La valorisation des thèses est une mission traditionnelle du MENRT, qui l'a conduit à favoriser la diffusion de ces travaux par impression et microfichage. Le circuit de diffusion des thèses fait l'objet d'un suivi régulier afin de mettre à profit les évolutions techniques pour améliorer son efficacité (enquête auprès des usagers des thèses en 1996, groupe de travail mis en place avec la direction de la recherche en 1997).

Le **groupe de travail** réuni ce jour a été mis en place à la demande du cabinet du Ministre. Son objectif est d'énoncer une série de recommandations relatives à la création d'un dispositif de diffusion des thèses par voie électronique. Il comprend, outre des agents du MENRT, des représentants académiques, des représentants des expérimentations en cours, et des experts.

Trois réunions sont prévues, consacrées successivement aux objectifs, aux options techniques, à l'organisation. Elles s'appuient sur un document de cadrage envoyé aux

participants avant chaque réunion.

1. ***Principe d'exhaustivité de la diffusion***

- Il apparaît souhaitable que toute thèse soutenue, dans sa version prenant en compte les modifications éventuellement demandées par le jury, puisse être diffusée par voie électronique, à la double condition que le jury n'ait pas émis un avis contraire et que l'auteur ait donné son accord. Il convient de privilégier la diffusion la plus large possible (Internet de préférence à l'extranet ou l'intranet).

- Dans certains domaines scientifiques, les thèses sont composées d'articles déjà publiés, dont les droits de publication ont été cédés aux éditeurs. La mise en ligne de ce type de thèse appelle une solution spécifique.

1. ***Durée de disponibilité des thèses sur le réseau***

   Il n'apparaît pas souhaitable de fixer *a priori*, même en fonction des disciplines, une durée de disponibilité des thèses sur le réseau, celle-ci devant être subordonnée à la demande des usagers et à la fréquence des consultations. La solution est à trouver dans une bonne articulation de l'archivage et de la mise en ligne.

2. ***Texte intégral ou éléments significatifs d'une thèse***

   Le groupe se prononce pour une mise en ligne du texte intégral des thèses.

3. ***Usage des fichiers numériques***

   Les thèses doivent être numérisées de préférence en mode texte.

4. ***Diversification des supports de diffusion***

   La complémentarité des autres supports (papier, microfiche) avec l'électronique est affirmée. Il importe en outre de pouvoir répondre à des demandes ponctuelles des utilisateurs. Il devra être examiné, notamment au cours de la troisième réunion, la question de la prise en charge (quel(s) opérateur(s) ?) de la fabrication et de la diffusion de produits de substitution confectionnés à partir des fichiers numériques.

5. ***L'organisation de l'archivage***

   En raison de la crise de l'édition en sciences humaines, de nombreuses thèses, même de qualité, ne font pas l'objet d'une publication. Le principe de conservation de ces travaux garde dès lors toute sa pertinence, sachant que l'exemplaire papier conservé dans la bibliothèque de l'université de soutenance ne présente pas de véritable garantie à long terme. L'archivage électronique (éventuellement doublé d'un archivage sur microfiche ?) est dans ces conditions fortement recommandé.

*

* *

Les représentants de la direction de la recherche indiqueront leur position définitive sur les différents points ci-dessus après consultation des directeurs scientifiques, du chef de la M.S.U. et du directeur de la recherche.

Les deux réunions suivantes sont ainsi fixées :

- **6 mars à 14h30 : Choix techniques**
- **27 mars à 14h30 : Organisation du circuit de diffusion**

GROUPE DE TRAVAIL SUR

LA DIFFUSION ELECTRONIQUE DES THESES

Réunion du 6 mars 2000 :

CHOIX TECHNIQUES

Paris, Agence de Modernisation des Universités

Présents :

MENRT – DES : Charlette Buresi, Chantal Freschard, Claude Jolly, Christine Okret, Pierre-Yves Renard, Pascale Vigier
MENRT – DR : Micheline Belin, Patrick Brasart
MENRT – DT : Jacques Guidon
ABES : Florence Robert, Frédérique Blondelle
ADBU : Jacqueline Gaude
CPU : Gérard Charbonneau, Serge Wolikow
Université Lumière Lyon 2 : Jean-Paul Ducasse, Viviane Bouletreau
INSA Lyon : Monique Joly, Brigitte Prudhomme, Jean-Michel Mermet
ANRT : Elisabeth Fichez
Université Joseph Fourier - Cellule Mathdoc : Pierre Bérard
Université Marne – La Vallée - SCD : Christian Lupovici
CINES : Alain Quéré, José Sanchez
LORIA : Jacques Ducloy

Excusés :

ADBU : Brigitte Mulette
MENRT – DR : Didier Arques

En préambule, les représentants du MENRT – DR font part de l'accord de Maurice Garden avec les conclusions de la réunion précédente. Toutefois, M. Garden souligne qu'il est

souhaitable que le délai s'écoulant entre la soutenance et la mise en ligne soit aussi bref que possible.

Puis sont abordés les points suivants :

1. ***La formation***

   Une formation serait souhaitable pour les doctorants dès la première année de thèse, et pour les directeurs de thèse. A titre d'exemple, une formation représente 2 x 4h à Lyon 2 (prise en charge par la cellule SENTIERS) ; et plusieurs sessions de 2h ½ à l'INSA de Lyon, avec une " hotline " (prise en charge par Doc'INSA).

   Plus généralement, il apparaît que les volumes de formation sont maîtrisables. Il serait utile de mutualiser les supports de cours.

   ***2*. Les thèses en format électronique natif**

   Pour les documents composites, des liens hypertexte assurent leur accessibilité depuis le texte de la thèse. Ces fichiers peuvent également être convertis par la suite en format standard.

   Il ressort des discussions que le format natif importe peu, sous réserve de quelques contraintes logicielles (compatibilité des formats avec RTF ou emploi de LaTeX ...). L'étudiant qui utilise des polices non standard ou qui transmet des fichiers nécessitant des logiciels de lecture peu usités devrait les fournir avec sa thèse, dans le respect des règles de droit.

   Un ensemble de recommandations seront rassemblées et formalisées pour la prochaine réunion par un sous-groupe confié à Christian Lupovici. Il est préconisé de les diffuser largement par la suite (web du MENRT, des établissements ...).

2. ***Les feuilles de style***

   Les feuilles de style ici évoquées sont celles qui structurent le document (récupération des niveaux de titres et sous-titres).

   Il est souhaitable d'établir au niveau national une feuille de base comportant les éléments obligatoires, en laissant aux établissements la possibilité d'y apporter des enrichissements (disciplinaires par exemple).

   Pour la prochaine réunion, une liste de recommandations doit également être établie sur ce point par le sous-groupe de C. Lupovici.

3. ***Les chaînes de traitement***

- L'université de Lyon 2 présente sa chaîne de conversion, qui peut traiter des documents d'origine logicielle différente en passant par RTF, format intermédiaire, afin d'archiver les fichiers en SGML puis de les transformer en format de diffusion.

LaTeX présente un cas à part de format très usité par les scientifiques et pour lequel une conversion en RTF est actuellement impossible.

- L'INSA de Lyon propose un circuit différent qui part des fichiers en Word, Postscript ou LaTeX pour les diffuser en PDF.

Selon les expériences de Lyon 2 et de l'INSA de Lyon, le délai de traitement d'une thèse correctement structurée (avec usage de la feuille de style) est d'environ une heure.

1. ***Les formats d'archivage et de diffusion***

Pour l'archivage, est préconisé un format structurant, tel SGML, avec en perspective l'usage de XML. Il importe en règle générale de s'appuyer sur des formats normalisés.

Pour la diffusion, l'INSA de Lyon s'en tient actuellement à PDF, et la chaîne de conversion de Lyon 2 permet une diffusion en PDF, SGML, HTML, XML.

La prochaine réunion - 27 mars à 14h30 – sera consacrée à l'organisation du circuit de diffusion.

GROUPE DE TRAVAIL SUR

LA DIFFUSION ELECTRONIQUE DES THESES

Réunion du 27 mars 2000 :

ORGANISATION DU CIRCUIT DE DIFFUSION

Paris, Agence de Modernisation des Universités

Présents :

MENRT – DES : Charlette Buresi, Chantal Freschard, Claude Jolly, Christine Okret, Pierre-Yves Renard, Pascale Vigier
MENRT – DR : Micheline Belin, Patrick Brasart
MENRT – DT : Jacques Guidon
ABES : Florence Robert
ADBU : Jacqueline Gaude, Brigitte Mulette
CPU : Gérard Charbonneau, Serge Wolikow
Université Lumière Lyon 2 : Jean-Paul Ducasse, Viviane Bouletreau
INSA Lyon : Monique Joly, Brigitte Prudhomme, Jean-Michel Mermet
ANRT : Elisabeth Fichez
Université Joseph Fourier - Cellule Mathdoc : Pierre Bérard
Université Marne – La Vallée - SCD : Christian Lupovici

CINES : Alain Quéré, José Sanchez
LORIA : Jacques Ducloy

Excusés :

MENRT – DR : Didier Arques

En préalable, Christian Lupovici présente l'état d'avancement du document sur les prescriptions techniques. Une première version soumise à commentaires par voie électronique est parvenue aux membres du sous-groupe. Ce texte doit être stabilisé avant sa transmission à la DES.

Puis sont traités les différents points de l'ordre du jour :

1. L'organisation du dispositif

a) Le dispositif cible

La discussion s'organise autour d'un tableau offrant une image synthétique du schéma proposé. La version enrichie par les commentaires du groupe est jointe à ce compte rendu. Il est précisé que ce schéma prend en considération les grandes entités concernées par le circuit, et non l'organisation des tâches entre les services de ces entités.

L'examen des différentes étapes du processus a donné lieu à un certain nombre de remarques :

- *Tirage de la thèse papier - Vérification des fichiers*

  Il importe de veiller à ne pas trop rallonger les délais avant soutenance avec cette procédure.

- *Diffusion numérique*

  La labellisation des chaînes de traitement suppose la rédaction d'un cahier des charges précisant les conditions de cette certification.

- *Signalement*

  La question du signalement enrichi des thèses par adjonction de mots clés propres à une discipline est évoquée. Il paraît souhaitable d'introduire ces mots clés au moment de la rédaction de la thèse, parmi les métadonnées.

- *Archivage*

  Une distinction est établie entre l'archivage local procédant du dépôt de la thèse sous forme électronique et l'archivage – duplication sur d'autres serveurs.

L'exemplaire papier doit être conservé au moins tant que le dispositif n'est pas stabilisé.

b) Le dispositif intermédiaire

Le groupe souhaite que l'Etat annonce une volonté politique claire en matière de diffusion électronique des thèses. Il est suggéré que les présidents d'université, *via* la CPU, soient sensibilisés à cette forme de valorisation des travaux scientifiques réalisés dans leur établissement. Des journées d'information seraient alors nécessaires.

Il faut toutefois prendre en compte les situations locales, très diverses : tous les établissements ne seront vraisemblablement pas en mesure d'opter pour la mise en œuvre rapide du dispositif cible. C'est pourquoi l'application de ce dispositif doit être volontaire, dans le cadre incitatif défini par le groupe de travail. Les établissements choisissant de ne pas s'y joindre auront toute latitude pour le faire ultérieurement.

2. L'évaluation sommaire des charges induites par les modèles préconisés

- A titre d'exemple, pour environ 130 thèses à traiter par an, l'INSA de Lyon et l'université Lyon 2 estiment la charge de personnel nécessaire à un ½ temps d'assistant ingénieur.

  Il est à noter que les universités pluridisciplinaires devront utiliser plusieurs chaînes de traitement.

- Pour l'archivage, le CINES considère les charges de personnel et de matériel nécessaires peu importantes (hors développements spécifiques).

3) La validation des conclusions du rapport

Le document final doit comporter une partie " politique ", une conclusion rassemblant des recommandations explicites, une annexe technique (sous-groupe de C. Lupovici), une partie " sources " (documents de cadrage, comptes rendus des réunions).

Il est décidé que le rapport sera transmis aux membres du groupe de travail par messagerie pour remarques etvalidation, une ultime réunion pouvant éventuellement être organisée si besoin est.

FONCTIONS	**ACTEURS**		
	DOCTORANT (DOCTEUR)	**ETABLISSEMENT**	**OPERATEUR NATIONAL**

- *Mise en forme de la thèse*			• Etablit des prescriptions de format, feuille de style • Diffuse les supports de formation
		• Forme le doctorant	
	• Saisit la thèse • La dépose sous forme électronique	• Apporte une assistance technique et logicielle	
- *Tirage de la thèse papier pour la soutenance*		• Vérifie les fichiers • Prend cette opération en charge, avec les navettes des corrections	
- *Soutenance*		• Le jury autorise la diffusion de la thèse	
	• Autorise la diffusion numérique de sa thèse après accord du jury		
- *Diffusion numérique*			• Labellise et / ou fournit les chaînes de traitement aux établissements

		• Met en œuvre la (les) chaîne(s) de traitement des fichiers • Archive en format pivot (SGML, XML) et diffuse (XML, PDF ...)	• Apporte une assistance technique
- *Signalement*			• Homogénéise les adresses électroniques des thèses
		• La BU catalogue la thèse dans les outils national et local, et saisit l'adresse électronique de la thèse numérisée	
- *Conservation / Archivage*		• Archive sous forme électronique • Un exemplaire papier est déposé à la BU jusqu'à stabilisation du dispositif	• Fait un archivage de sécurité
- *Production de substituts et commercialisation*		• A vocation à produire / commercialiser des substituts	• Confectionne des produits dérivés de façon contractuelle