

Le dispositif national d'archivage et de signalement des thèses électroniques

Depuis la publication de l'arrêté du 7 août 2006 relatif aux modalités de dépôt, de signalement, de reproduction, de diffusion et de conservation des thèses ou des travaux présentés en soutenance en vue du doctorat, chaque établissement habilité à délivrer des diplômes de doctorat fait face à une alternative : pour le dépôt de la thèse, il doit choisir entre le support papier et la forme électronique.

Marianne Giloux

giloux@abes.fr

Isabelle Mauger Perez

mauger@abes.fr

Agence bibliographique
de l'Enseignement supérieur

Sans ambiguïté, l'alternative conditionne le traitement du document thèse. En effet, toutes les thèses nativement électroniques doivent transiter par Star, l'application nationale pour le Signalement des Thèses, Archivage et Recherche, développée par l'Agence bibliographique de l'Enseignement supérieur à la demande de la Sous-direction des bibliothèques et de l'information scientifique.

D'aucuns regretteront l'impossibilité de recourir à un dépôt mixte¹ ; d'autres argueront qu'obliger les établissements à se prononcer est la seule façon d'identifier sans ambiguïté la thèse « officielle », *i.e.* le document dont le dépôt et l'examen par le jury débouchent sur la délivrance du diplôme². Tous s'accorderont sur le fait qu'une circulaire d'application pourrait aider les établissements à sortir du statu quo et à opter pour le dépôt électronique.

1. Voir « Dépôt mixte » dans l'encadré p. 48.

2. Ce que les catalogueurs traduisent par le code « m » document originel dans la sous-zone Unimarc 105\$b.

Un nouveau circuit à mettre en place

Le passage au dépôt électronique permet de rationaliser le circuit des thèses dans l'établissement. En effet, Star s'intéresse à l'ensemble des tâches à accomplir, aux différents acteurs impliqués et au mécanisme de validation dans la réalisation du processus de signalement et d'archivage de la thèse électronique. Application informatique de gestion d'un flux d'informations (*workflow*), Star identifie trois acteurs principaux qui correspondent (peu ou prou) aux écoles doctorales et/ou aux services de scolarité, à la bibliothèque, et au responsable de la collation du grade de docteur.

Lors de son inscription en thèse, le service de scolarité et/ou l'école doctorale saisit le nom du doctorant, son directeur de thèse, le sujet envisagé. Trois ans plus tard, une fois la thèse soutenue, le bibliothécaire saisit les mêmes informations dans la notice bibliographique de la thèse. Pourquoi ne pas tirer parti du travail déjà fait ?

Marianne Giloux est ingénieur de recherche à l'Abes, où elle est responsable du service Produits et services et responsable du projet Star.

Isabelle Mauger Perez est conservateur des bibliothèques à l'Abes, où elle est chargée de la gestion et de l'animation du réseau Star.

Chacun gagne à partager les données saisies avec les autres! Éviter la double saisie est donc un des buts de Star et c'est notamment pour cette raison que cette application pourra bientôt être alimentée par les données que les établissements possèdent déjà: données purement administratives issues des applications de gestion (Apogee, Graal...) ou données plus complètes issues des systèmes locaux de gestion des thèses électroniques (Dspace, ORI...). Mais sous quelle forme se présentent ces données?

TEF, métadonnées des thèses électroniques françaises

Un groupe d'experts³ rassemblant des documentalistes, des bibliothécaires et des informaticiens issus de divers établissements dépendant du ministère de l'Enseignement supérieur et de la Recherche s'est penché sur ce qui fait la spécificité des thèses. En effet, au carrefour de plusieurs mondes, la thèse est à la fois un objet documentaire, un document sanctionnant des études et une œuvre intellectuelle protégée par le droit d'auteur. Sous forme électronique, elle obéit à certaines caractéristiques techniques. La recommandation TEF (thèses électroniques françaises)⁴, jeu de métadonnées XML, vise à prendre en compte ces différentes facettes:

- les métadonnées descriptives (résumé de la thèse, mots clés...) sont basées sur Dublin Core;

3. Réunis dans le cadre de l'Afnor, groupe CG46/CN357/GE5.

4. TEF Métadonnées des thèses françaises: www.abes.fr/abes/documents/tef/index.html
La seconde édition de la recommandation TEF a été publiée en mai 2006.

- les autres métadonnées, qualifiées de métadonnées de gestion, servent à décrire:

- le contexte administratif de la thèse (noms des membres du jury de soutenance, nom de l'école doctorale...),
- les conditions juridiques associées à la thèse (droits de propriété intellectuelle), et les impératifs de conservation (type d'encodage des caractères, format de fichiers...);
- le tout est « encapsulé » dans un vocabulaire METS⁵ qui présente l'avantage de proposer une structure modulaire tout en gérant le(s) fichier(s) électronique(s) de la thèse.

Ces métadonnées, certes spécifiques mais obéissant à des standards informatiques répandus sur internet, seront utiles à la diffusion de la thèse: dans les établissements eux-mêmes mais aussi dans des entrepôts (institutionnels ou non). Format pivot, TEF permet d'assurer les fonctionnalités obligatoires de Star décrites dans l'arrêté du 7 août 2006: le signalement automatique des thèses électroniques dans le Sudoc, Système universitaire de documentation (conversion TEF → Unimarc) et l'archivage pérenne au Cines, Centre informatique national de l'Enseignement supérieur (conversion TEF → PAC, plate-forme d'archivage du Cines).

L'archivage pérenne et les remontées d'archives

La nécessité de la mise en œuvre d'un processus d'archivage pérenne n'est pas difficile à démontrer; cette question est à l'étude dans tous les grands organismes nationaux et internationaux. Le développement phénoménal de la publication électronique a modifié considérablement la profession d'archiviste.

5. METS (Metadata Encoding and Transmission Standard) est un standard maintenu par la Bibliothèque du Congrès.

L'archivage pérenne consiste à conserver le document et l'information qu'il contient, à conserver son aspect physique comme son aspect intellectuel, et cela sur le très long terme (trente ans et au-delà), de manière à pouvoir le rendre accessible et surtout intelligible.

Sommes-nous capables de relire notre sauvegarde sur disquette effec-

**Au carrefour
de plusieurs mondes,
la thèse est à la fois
un objet documentaire,
un document sanctionnant
des études et une œuvre
intellectuelle protégée
par le droit d'auteur**

tuée il y a dix ans? Certains formats de fichier sont désormais inconnus, pour certains supports physiques le matériel ou le logiciel de lecture a disparu...

Ce sont ces contraintes-là qu'un projet d'archivage pérenne doit prendre en compte. Le Cines avec son système PAC met en place un système et un service d'archivage pérenne de documents numériques pour la communauté de l'enseignement supérieur et de la recherche.

Comment PAC procède-t-il afin de préserver sur le long terme l'accès à un document? Plusieurs types de solutions doivent être mis en œuvre, successivement et parallèlement:

- L'utilisation de métadonnées spécifiques en premier lieu, afin d'être capable rapidement et simplement de connaître le contenu du document et de l'identifier sans ambiguïté; il faut également attribuer au document un identifiant unique et pérenne.

- Ensuite, privilégier les formats durables lors de l'enregistrement des

fichiers. Mais comment savoir quels sont les formats durables? Le Cines émet des recommandations⁶, cette liste sera certainement amenée à évoluer au fil du temps.

- Enfin le système doit mettre en œuvre un processus de veille technologique et d'anticipation afin d'effectuer des migrations logiques et/ou physiques nécessaires à la pérennisation des documents.

Depuis le lancement de l'application Star, un certain nombre de thèses ont d'ores et déjà été archivées. Le processus mis en place effectue des contrôles et vérifications à plusieurs niveaux: le fichier déposé doit non seulement être dans l'un des formats acceptés mais il doit être également « bien formé » (conforme aux spécifications syntaxiques de son format) et « valide » (conforme à des spécifications sémantiques complémentaires).

Que faire lorsque le fichier d'une thèse est refusé? Le plus souvent, le docteur ayant déposé sa thèse n'est plus dans l'établissement et il est impossible de lui demander une nouvelle version « conforme ». Mais l'établissement est habilité à mettre en œuvre tous les moyens nécessaires afin de rendre le document compatible avec les contraintes d'archivage: il est donc possible de modifier le document (dans sa forme) afin qu'il y réponde.

Enfin, pourquoi archiver si ce n'est pour avoir un jour la possibilité de demander une restitution du document? Via Star, les établissements ont la possibilité d'effectuer auprès du système PAC une demande de communication de leurs thèses archivées. Grâce à une interface de recherche, il est possible d'interroger la base des thèses archivées puis, par l'intermédiaire du numéro d'identification unique et pérenne de l'archive PAC, de demander au système la remontée du fichier.

6. Voir « Formats de fichiers dans Star » dans l'encadré ci-contre.

Petit inventaire des faux amis et des confusions (trop) largement répandues... à propos des thèses électroniques

Archivage pérenne et archives ouvertes

La traduction d'« Open Archives » par « archives ouvertes » introduit une confusion malencontreuse dans le domaine des thèses électroniques.

Le mouvement international des archives ouvertes vise à favoriser la communication directe au sein de la communauté scientifique et à valoriser la littérature grise (dont les thèses sont un élément) en la diffusant largement sans intermédiaire.

L'archivage pérenne des documents électroniques vise quant à lui à la conservation du document à très long terme et à la préservation de son intelligibilité.

Dans le champ des thèses électroniques, Star permet l'archivage pérenne en envoyant au Cines le fichier de la thèse officielle et ses métadonnées de conservation; et peut, sous réserve de l'accord du docteur et des ayants droit, contribuer au développement des archives ouvertes en envoyant une version de diffusion de la thèse et ses métadonnées sur la (les) plate-forme(s) de diffusion choisie(s) par l'établissement de soutenance.

Circuit électronique des thèses versus circuit des thèses électroniques

Star gère les thèses électroniques (métadonnées et fichiers). C'est une application qui dépasse le cadre documentaire *stricto sensu* et concerne également les écoles doctorales ou les services de scolarité. À ce titre, on peut dire que Star s'intéresse au circuit des thèses.

Mais le circuit de la thèse n'est pas l'objet de Star qui vise à signaler et à archiver la thèse électronique une fois le diplôme délivré. Star n'est pas une application cherchant à dématérialiser les opérations préalables à la soutenance, comme il en existe dans certaines universités.

Dépôt mixte

Cette expression, largement répandue, vise à décrire la situation où l'établissement de soutenance demande au docteur de déposer deux éditions de sa thèse: une édition papier et une édition électronique. À noter que le dépôt mixte n'a pas d'existence légale: ni la circulaire du 29 mars 2005 sur le dépôt, signalement, diffusion et archivage des thèses sous forme électronique ni l'arrêté du 7 août 2006 n'y font référence.

En effet, la pratique du dépôt mixte introduit une confusion d'ordre juridique: laquelle des deux éditions est la thèse officielle? Si c'est l'édition papier, la thèse électronique n'est pas soumise à l'obligation de signalement et de conservation: elle ne passera pas par Star. Inversement, si c'est l'édition électronique, la bibliothèque n'est pas tenue de conserver indéfiniment la thèse papier.

Formats de fichiers dans Star

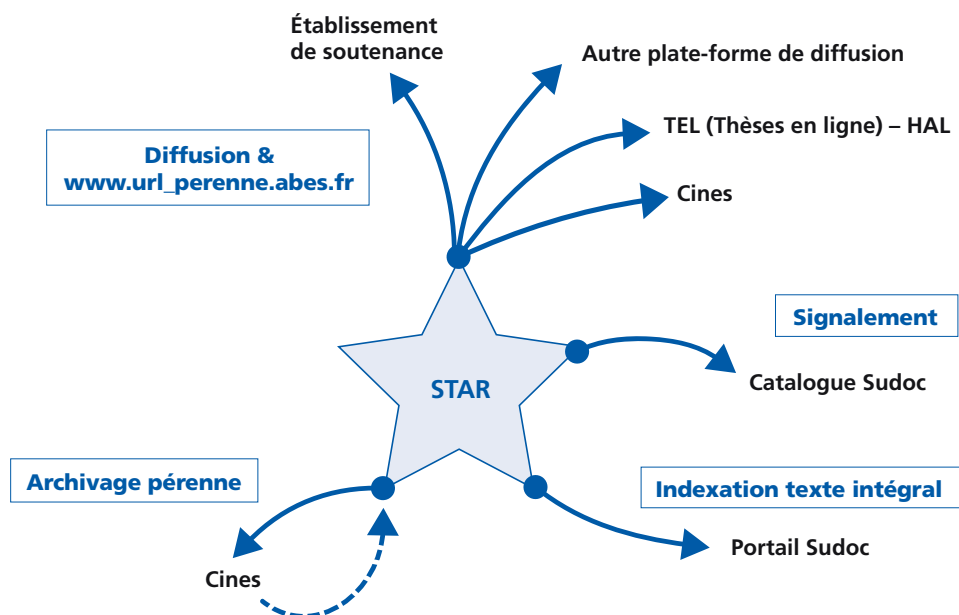
Pour l'archivage pérenne, Star se conforme aux recommandations du Cines: le document thèse doit être soit au format PDF/A (*i.e.* PDF 1.4) soit au format XML (comportant éventuellement des éléments textuels en TXT, ou des images en TIFF, GIF, JPEG ou PNG).

Pour la diffusion, aucune contrainte technique n'existe dans Star. Les deux formats de diffusion les plus répandus sont le PDF et le HTML.

Star, Sparte et Cyberdocs

Star n'est pas une chaîne de transformation du document thèse; Star n'intervient pas sur le document thèse, quel que soit le format reçu (PDF ou XML). Sparte, expérimentation menée à l'Abes entre 2004 et 2006, était un serveur de conversion de documents du format RTF (traitement de texte) vers le format XML (en incluant la DTD TEI). Sparte a été abandonnée car trop de scories étaient générées lors de la conversion. Il existe d'autres chaînes de conversion comme Cyberdocs par exemple (utilisée notamment par l'université Lyon-II).

Les fonctionnalités de Star



Star, un gage pour la diffusion des thèses

Grâce au développement des archives ouvertes, toute personne peut aujourd'hui aisément déposer un document sur une plate-forme d'auto-archivage thématique, nationale ou internationale. Ce type de dépôt « volontaire » concerne les articles scientifiques comme les thèses⁷.

Pour autant, il n'exonère pas l'établissement de l'obligation légale qui lui est faite de communiquer la thèse en son sein (physiquement ou virtuellement).

Trois points méritent une attention particulière : comment assurer la communication du document thèse tel que validé par le jury ? Quel canal choisir pour valoriser le patrimoine scientifique de l'établissement ? Comment garantir dans le temps l'accès à la thèse ? Star remplit ces trois impératifs :

- une thèse passée par Star est nécessairement la thèse qui a permis l'obtention du diplôme ;
- avec l'accord du docteur, Star irrigue la (les) plate-forme(s) de diffusion choisie(s) par l'établissement de soutenance en envoyant la version de la thèse destinée à la diffusion, et ses métadonnées. Par ailleurs, les métadonnées des thèses sont disséminées via l'exploitation d'un serveur OAI/PMH, multipliant ainsi les occasions de signalement de la thèse ;
- la pérennité de l'accès à la thèse est garantie par l'attribution d'une URL pérenne, à la fois adresse fiable et label d'authenticité.

L'utilisateur final accède alors au document en étant assuré d'être en présence de la thèse sanctionnée par l'établissement de soutenance (éventuellement après corrections demandées par le jury). La thèse peut être citée, de façon pérenne, dans n'importe quel environnement (article, notice descriptive...).

Rejoindre le réseau Star

Choisir de rejoindre le réseau Star est donc une question plus politique que technique. Tout d'abord parce que c'est une décision qui engage l'établissement et non la seule bibliothèque. Ensuite parce que l'outil est une chose, la refonte du circuit des thèses et la mise en place de procédures de travail collaboratif en sont une autre. Enfin parce que décider que les doctorants doivent rendre leur thèse sous forme électronique ne suffit pas : il faut leur en donner les moyens.

Cependant, entrer dans le dispositif Star permet à l'établissement de s'acquitter de ses obligations de signalement et de conservation des thèses électroniques de façon automatisée et transparente. Et Star apporte une plus-value pour la diffusion : la thèse électronique n'est plus un document nu mais elle est identifiée comme la thèse officielle, dotée d'un identifiant pérenne, et enrichie de métadonnées interopérables.

Novembre 2007

7. Voir « Archivage pérenne et archives ouvertes » dans l'encadré ci-contre.