

Les thèses électroniques *et moi*

Le passage au dépôt électronique des thèses de l'Université Polytechnique Hauts-de-France en 2012 a nécessité la mise en œuvre d'un processus complexe mais indispensable à leur visibilité dans les grands réseaux nationaux.



C'est en 2012 que l'Université Polytechnique Hauts-de-France, alors Université de Valenciennes et du Hainaut-Cambrésis, a décidé du passage au dépôt électronique de la version de soutenance des thèses. Pour moi, qui venait d'intégrer le service des thèses, comme pour mes collègues, ce fut un gros défi à relever. En effet, le passage à l'électronique introduisait de nouveaux concepts tels que l'embargo de l'auteur limité ou illimité, la version d'archivage, la version de diffusion en fonction du droit d'auteur, le choix de signaler tous les membres du jury alors que jusque-là on ne signalait que les directeurs et l'université, ainsi que l'école doctorale et les laboratoires. Le traitement mis en place à l'époque se passe en deux temps : la collecte et la saisie des métadonnées par un agent (rôle de bibliothécaire), puis leur validation ou leur mise à jour par un autre agent (rôle de validateur, dont je suis chargé).

UN CIRCUIT DE TRAITEMENT COMPLEXE ET FRAGILE

Une des grandes nouveautés pour moi fut la prise en compte de l'évolution du statut d'une thèse. Même si chez nous 64% des thèses sont dès l'origine en diffusion libre, 17% sont sous embargo limité, 9% en embargo illimité et 10% sous confidentialité, soit 27% des thèses dont le périmètre de diffusion change. Il a fallu les identifier facilement et maîtriser les procédures pour les mettre à jour.

Comme nous avons fait le choix de diffuser ces documents dans une archive ouverte déployée en ORI OAI¹ par le service informatique de l'université, notre circuit de traitement est plus complexe que celui présenté par l'Abes. En plus des outils Step et Star, nous utilisons actuellement Nuxeo pour déposer le document et générer une URL, ORI-OAI pour générer les métadonnées et le scénario de diffusion, et un fichier sous Excel pour le suivi.

Si ce traitement est une amélioration indéniable par rapport au signalement papier et à la conservation imprimée, je constate avec le recul que le circuit est complexe et assez fragile car il fait intervenir, en plus de

nous, beaucoup d'acteurs différents : la scolarité (pour Apogée), la DSI (pour Nuxeo et ORI-OAI), l'Abes, le Cines² et le CCSD³ pour Step, Star, WinIBW et HAL. Les sources de panne sont nombreuses et les mises à jour, suite par exemple à une évolution normative ou administrative, longues et compliquées.

OBJECTIF : INTÉGRER THESES.FR

Après le courant, il a fallu penser au rétrospectif. Jusqu'alors, nos thèses étaient signalées dans le Sudoc mais avec des notices d'une qualité normative hétérogène. Pour qu'elles apparaissent dans theses.fr, l'Abes demandait un certain niveau de qualité bibliographique. Nous avons donc mis à jour les quelque 950 thèses soutenues entre 1974 et 2011, travail qui nous a pris, à mes trois collègues et moi-même, 6 ans, de 2011 à 2016. En tant que correspondant catalogage et pour nous faciliter le travail, j'avais créé un script dans WinIBW et un procédurier avec la liste des champs à supprimer ou corriger. Ce travail, bien que long et fastidieux, est pour moi une source de fierté car aujourd'hui toutes les notices de nos thèses soutenues entre 1985 et nos jours sont consultables sur Theses.fr.

UN PLAN DE BATAILLE POUR LA NUMÉRISATION RÉTROSPECTIVE

Le SCD a ensuite décidé de se lancer dans la numérisation de ce fonds. Dans ce gros et long chantier, qui nécessitait de recontacter les auteurs afin d'avoir leur autorisation pour la diffusion de leur thèse, j'étais chargé de réfléchir au catalogage des notices et de proposer « un plan de bataille ». Après discussion avec les collègues du service des thèses de l'Abes (que je remercie encore), confrontés au même besoin que nous, nous avons adopté la moulinette qu'ils avaient développée pour transformer en masse les notices « thèses papier », ce qui fut pour nous un gain de temps énorme.



Réseau de Mitochondria

© Odra Noel

Ce chantier de numérisation, qui implique en plus du SCD, notre DSI pour l'aspect stockage et diffusion, l'ANRT (Association nationale de la recherche et de la technologie) de Lille pour la numérisation, l'Abes pour le catalogage et le signalement, a commencé fin 2016. Il concerne toutes les thèses soutenues entre 1990 et 2011, environ 820 documents, interrogeables et trouvables dans le Sudoc, Theses.fr, notre catalogue, partiellement dans HAL mais pas dans notre archive ouverte institutionnelle, Theoreme. Notre DSI n'a malheureusement pas pu créer un flux opérationnel pour les y intégrer. Notre dernier chantier en date concerne le signalement dans HAL de ces thèses numérisées dont 150 restent encore à traiter.

FRÉDÉRIC PRUVOST

Bibliothécaire assistant spécialisé au SCD de l'Université Polytechnique Hauts-de-France
frederic.pruvost@uphf.fr

[1] ORI-OAI : Outil de Référencement et d'Indexation, réseau de portails OAI

[2] Cines : Centre informatique national de l'enseignement supérieur

[3] CCSD : Centre pour la communication scientifique directe