

Genèse d'une archive ouverte institutionnelle

Open Archive Toulouse Archive Ouverte :
une archive ouverte multiétablissements pilotée par l'INP de Toulouse

OATAO

Nous avons voulu offrir un service simple et performant pour la communauté scientifique toulousaine : visibilité de la production scientifique et indicateurs de gouvernance sont les deux objectifs qui nous ont guidés. OATAO permet à tout chercheur de déposer un article en quelques minutes seulement. La qualité des données et le respect du droit sont garantis par les professionnels de la documentation. À l'heure de la dématérialisation mondiale des contenus scientifiques et de la concurrence universitaire internationale, exister sur la Toile est une nécessité incontournable. Entre GRAAL et HAL, OATAO peut s'avérer un « go-between » discret et efficace.

G. Casamatta ✉ casamatta@inp-toulouse.fr

Gilbert Casamatta, président de l'Institut national polytechnique de Toulouse.

INP ✉ www.inp-toulouse.fr ☎ 05 34 32 30 00 📠 30 06 📮 6 allée Émile-Monso BP 34038 - 31029 TOULOUSE CEDEX 4

Quand elle a pris sa retraite en 2001, Thérèse Gros, ingénieur-documentaliste CNRS et avisée responsable d'une des bibliothèques de l'INP de Toulouse a dit : « La gestion des publications issues de la recherche est une mission qui doit être reprise par les SCD ».*

Les SCD avaient déjà un cœur de cible : la mise en ligne des thèses. C'est chose faite à l'INP de Toulouse depuis le 1er janvier 2005. Fort de cette expérience d'une archive ouverte reposant sur un logiciel libre (Eprints), le SCD a pris le taureau par les cornes pour proposer une archive ouverte des publications scientifiques, avec deux objectifs : visibilité et indicateurs.

First things first : d'abord le politique !

Les statistiques de consultation des thèses en ligne, en croissance exponentielle dès l'ouverture du site (<http://ethesis.inp-toulouse.fr/>), ont tout de suite marqué les esprits des décideurs. Dès l'automne 2005 et capitalisant sur ce succès, un travail de sensibilisation et d'information est entrepris en conseil scientifique, à la direction de la recherche, auprès des directeurs de laboratoires, auprès du président de l'université. Plusieurs présentations successives convainquent les plus réticents : **oui**, les éditeurs chez lesquels nos chercheurs publient autorisent la diffusion dans une archive ouverte, **oui** le taux de citation aug-

mente avec l'exposition dans une archive ouverte, **oui** les professionnels de la documentation sont les garants responsables au nom de l'établissement de la vérification des conditions juridiques de diffusion, **oui** on peut retirer de l'archive tout document qu'on ne veut plus exposer, **oui** on peut afficher automatiquement les publications d'un laboratoire sur le site de celui-ci à partir de l'archive ouverte, **oui** on peut extraire des bibliographies pour les CV des chercheurs, les enquêtes et les évaluations, **oui** la gouvernance de l'établissement pourra disposer d'indicateurs sur sa production scientifique, **oui** tous les moteurs de recherche du web « crawlent » dans l'archive en temps réel, **oui** l'établissement est garant de la pérennité de l'accès à l'archive, **oui** le dépôt ne prendra pas plus de cinq minutes au chercheur, **oui** l'archive pourra alimenter HAL et Prod'INRA, etc. Le projet est adopté à l'unanimité par le conseil scientifique de l'INP de Toulouse le 16 juin 2006. Un groupe de travail est constitué, un nom est trouvé : OATAO est née.

Un groupe et des grands principes

Une fois l'accord politique acquis est créé, dès octobre 2006, un groupe de travail mixte comprenant à la fois des représentants des bibliothèques et des correspondants de laboratoires. Un

point important : nous avons insisté pour que ces correspondants soient des chercheurs (et non des personnels administratifs) en phase avec les responsables des laboratoires.

La feuille de route est la suivante : définir les conditions de la création d'une organisation humaine et technique permettant de :

- diffuser les travaux des chercheurs sans critère d'appartenance, pour un meilleur affichage de l'activité scientifique ;
- mettre en place une base de données en *open access* sans remettre en cause le processus traditionnel de publication ;
- respecter les normes techniques et juridiques en vigueur.

Le groupe de travail (qui s'est réuni à huit reprises sur une période d'un an) s'est essentiellement attaché à :

- définir les besoins et les contraintes des laboratoires : disposer de données bibliographiques pour les rapports quadriennaux d'évaluation, les rapports d'activité, les CV personnels, les enquêtes ; effectuer un dépôt unique et alimenter automatiquement HAL et prod'INRA ; produire des indicateurs consolidés au niveau d'une équipe, d'un laboratoire, d'un établissement, de plusieurs établissements ;
- décrire les différents scénarios des dépôts possibles ;
- s'accorder sur la typologie des documents diffusés dans l'archive ouverte.



OATAO à l'été 2008 : plus de 600 documents déposés

16 juin 2006 : adoption par le conseil d'administration de l'Institut national polytechnique de Toulouse

Octobre 2006 : groupe de travail mixte SCD & LABORATOIRES

Septembre 2007 : groupe de travail « pros doc »

Janvier 2008 : ouverture de l'archive OATAO

Mai 2008 : **plus de 250 documents en ligne, près de 1 000 visites en 1 mois**

Juin 2008 : **plus de 400 documents validés et plus de 120 en cours de validation**



Fin 2010 : 1 500 dépôts ; c'est l'objectif indiqué dans le contrat quadriennal...

Ce travail a aussi été l'occasion d'expliquer les buts d'une archive ouverte, de rassurer les chercheurs (sur le droit et les éditeurs), en bref, de commencer à changer les mentalités et de faire « mûrir » les esprits.

Au final, une série de grands principes sont arrêtés.

L'archive ouverte doit avoir un contenu

Il ne s'agit pas de recréer une nouvelle base de données bibliographiques qui ne comporterait que des références et non des documents. Même si les éditeurs n'autorisent pas la diffusion des documents au moment du dépôt, le texte intégral doit obligatoirement être déposé (sauf sur raisons motivées : documents confidentiels), d'une part, parce que la politique des éditeurs évolue dans le temps, d'autre part parce que notre rôle est de préserver les documents dans le temps.

Un format de diffusion unique

Les documents sont diffusés au format PDF, format de diffusion largement utilisé et qui n'est pas dépendant d'un éditeur de logiciel. Il n'est en revanche pas exigé des chercheurs qu'ils postent leurs documents dans ce format : dans ce cas, les gestionnaires de l'archive se chargent de la conversion.

Une interface web en langue anglaise

C'est un choix assumé par le groupe de travail qui souhaite toucher le plus largement possible la communauté scientifique internationale. Seule la FAQ est en français.

Des responsabilités bien réparties

Aux chercheurs, le dépôt des documents et la **validation scientifique**, l'accord des coauteurs pour la diffusion et l'assurance que les documents ne contiennent que des éléments aux droits acquis.

Aux gestionnaires de l'archive, la validation technique des documents et des métadonnées, le contrôle des droits de diffusion auprès des éditeurs à l'aide de ROME/SHERPA (<http://www.sherpa.ac.uk/romeo.php>) ou par contact direct auprès des édi-

teurs, la gestion technique et pérenne de l'archive.

Ces choix, différents de ceux de HAL, où le contrôle des autorisations juridiques de diffusion par les éditeurs est laissé à la responsabilité des chercheurs, sont délibérés : nous avons voulu que les chercheurs puissent déposer le plus rapidement possible leurs documents en évitant au maximum les contraintes.

Un dépôt unique

Pour le chercheur, lui assurer qu'il n'aura qu'un seul dépôt de ses publications à effectuer est une condition indispensable à l'adhésion au projet. Le transfert de ses publications vers HAL, par exemple, fait partie du cahier des charges.

Un logiciel et des « pros doc »

Au départ, notre choix s'est porté sur le logiciel ORI-OAI (<http://www.ori-oai.org/>). L'INP de Toulouse est très impliqué dans le développement de ce logiciel (deux personnes membres du comité de pilotage, un responsable du groupe fonctionnel pour le module « Thèses », un développeur du moteur de recherche et d'indexation). ORI-OAI répond parfaitement aux objectifs de l'archive institutionnelle souhaitée. Mais les délais de livraison du logiciel n'étaient pas compatibles avec le calendrier fixé : une ouverture du site en janvier 2008. Le choix s'est donc orienté vers le logiciel Eprints (<http://www.eprints.org/software/>) dans un premier temps, en attendant qu'ORI-OAI soit totalement opérationnel. Les données seront facilement récupérables lors de la migration vers ORI-OAI.

Une équipe resserrée de professionnels de la documentation est constituée, en septembre 2007. Ses objectifs sont les suivants :

- travail sur les métadonnées : assurer la compatibilité avec HAL et les recommandations du GTO, le groupe de travail sur les archives ouvertes (<http://gtao.wikidot.com/>) ;

- amélioration de l'interface web du site ;

- rédaction de la FAQ ;

- test de la maquette.

Des réunions hebdomadaires ont permis de tenir les délais. OATAO a ouvert le 17 janvier 2008.

Vérification et valeur ajoutée

En amont de l'ouverture de l'archive a été défini, sur un wiki dédié au projet, un ensemble de règles pour assurer un traitement efficace des dépôts sans jamais perdre de vue la pérennité et la cohérence de la base. Les principaux points ont porté sur la définition :

- des métadonnées obligatoires, visibles ou non au déposant, pour chaque type de document ;

- des règles de saisie des champs principaux, largement reprises des normes de saisie de catalogage ;

- des règles pour la création des comptes utilisateurs.

Parmi les décisions importantes : seuls les *postprints* d'articles sont diffusés dans OATAO pour des questions de droits de diffusion et pour éviter la gestion des versions.

Un important travail a été consacré à l'interface :

- écriture des aides pour chaque champ ;

- création d'une FAQ illustrée de démonstrations animées.

Enfin, des réunions d'information (en nombre encore insuffisant !) ont été organisées dans les laboratoires.

Les premiers dépôts sont apparus assez rapidement dès l'ouverture de l'archive. Une partie du traitement diffère peu du catalogage traditionnel : titre, résumé, auteur(s), ISSN, volume, pages, etc. sont à renseigner.

La grande différence réside dans la **vérification des droits de diffusion** de l'article déposé, point essentiel de la **valeur ajoutée** d'OATAO, qui engage notre responsabilité, et donc celle de l'établissement en tant que gestionnaire de l'archive.

.../...



Cf. www.isae.fr



Non seulement les archives ouvertes assurent un support essentiel à l'activité scientifique, mais elles en constituent également un véritable outil. Il s'agit en premier lieu d'un vecteur efficace de communication des résultats de la recherche, assurant une excellente diffusion de la production scientifique d'un établissement. C'est également un outil précieux de suivi interne de cette production, particulièrement utile pour réaliser des bilans afin de répondre aux enquêtes et surtout de fournir des données fiables pour les évaluations. Dans un contexte où les outils automatiques choisis par les instances d'évaluation ne permettent pas toujours d'identifier avec précision la production d'un établissement ou d'un laboratoire, un établissement peut s'appuyer sur ses archives ouvertes pour réaliser ou compléter cette identification. Les **facteurs clés** de la réussite de la mise en place d'archives ouvertes au niveau d'un établissement sont au nombre de deux : **un support efficace** du service de documentation pour vérifier les aspects liés aux droits de publication et **une adhésion enthousiaste** des chercheurs au système. Pour susciter cette adhésion, il est capital de coupler la mise en place d'un tel système avec une exploitation pour différents usages : génération automatique de bilans de publication intégrables dans des rapports, génération automatique de pages web pour des chercheurs, des équipes, des laboratoires... C'est l'importance des services rendus qui convaincra le chercheur de l'intérêt de passer les quelques dizaines de secondes nécessaires à soumettre chacune de ses publications. Les archives ouvertes assurent la visibilité d'un établissement sur la scène électronique internationale de la production scientifique.

F. Thivet frederic.thivet@isae.fr

Frédéric Thivet, directeur de la recherche et des ressources pédagogiques de l'Institut supérieur de l'aéronautique et de l'espace

ISAE www.isae.fr ☎ 05 61 33 80 50

✉ 10 avenue Édouard-Belin BP 54032 - 31055 TOULOUSE CEDEX 4

L'Institut supérieur de l'aéronautique et de l'espace est issu du rapprochement de SUPAERO et ENSICA.

Le premier niveau de vérification se fait via le site de référence SHERPA/ROMEO (<http://www.sherpa.ac.uk/romeo.ph>), qui recense les politiques juridiques des éditeurs et, en cas de silence de ce site, par contact direct avec l'éditeur. Si un article n'est pas diffusable, nous bloquons sa visibilité sur Internet mais le fichier est tout de même conservé dans l'archive. Dans ce cas, OATAO propose la demande de copie directement à l'auteur par mail. De plus, le logiciel Eprints permet de diffuser la diffusion s'il existe un embargo sur l'article.

Le **staff OATAO** est constitué de dix gestionnaires, de profil essentiellement catalogueurs, répartis sur les trois établissements qui alimentent aujourd'hui OATAO : l'INP de Toulouse, l'École nationale vétérinaire de Toulouse et l'Institut supérieur de l'aéronautique et de l'espace (ex SUPAERO+ENSICA).

Les échanges et la coordination se font par messagerie, sur le wiki du projet ou lors de réunions en présentiel. Notre réactivité à traiter les posts est essentielle pour la crédibilité d'OATAO auprès des chercheurs : le délai moyen entre le dépôt et la mise en ligne est actuellement de 48 heures.

Trois administrateurs constituent une cellule de veille, de contrôle qualité et de sécurité de l'archive. Ils créent les nouvelles institutions, traquent la présence d'éventuels doublons ou anomalies.

Les chantiers actuels portent sur l'amélioration permanente d'OATAO pour répondre aux suggestions et besoins des chercheurs et des tutelles :
- export sous certains formats demandés par les chercheurs en fonction des outils de gestion bibliographique qu'ils utilisent déjà ;

- mise au point de statistiques de consultations (une première instance de Google Analytics a été installée le 7 avril 2008) ;

- cohérence dans l'écriture des noms de laboratoires et des institutions.

Le staff OATAO tisse peu à peu des liens étroits avec les chercheurs. Une bonne surprise : deux laboratoires n'ont pas hésité à embaucher un stagiaire en IUT Documentation, formé et encadré par le staff, pour procéder au dépôt rétrospectif des publications en vue du rapport quadriennal. Une campagne de communication est programmée, avec des réunions d'information et des formations pratiques dans les laboratoires intéressés, ainsi qu'un « flyer » sur support papier, qui résume l'essentiel des informations à connaître en tant que chercheur pour déposer.



Cf. www.envt.fr

À l'École nationale vétérinaire de Toulouse (ENVT), nous avons opté pour les archives ouvertes afin de valoriser l'activité recherche (augmentation du taux de citation des articles) et de contribuer à la diffusion libre et gratuite de l'information scientifique.

La participation à OATAO a aussi été l'occasion de rapprocher documentalistes et chercheurs. En effet, le personnel de la bibliothèque connaît désormais mieux les thématiques de travail des équipes de recherche et les chercheurs s'intéressent davantage aux services documentaires qui leur sont proposés.

Nous avons travaillé à partir d'une liste des publications de l'ENVT depuis 2007 (mise à jour par veille documentaire) et avons contacté les auteurs afin d'obtenir les postprints ou les preprints de leurs articles ou l'autorisation de les « appauvrir » de leur mise en forme (pour respecter les politiques des éditeurs).

M. Andro m.andro@envt.fr

Mathieu Andro est responsable de la bibliothèque de l'École nationale vétérinaire de Toulouse.

ENVT www.envt.fr ☎ 05 61 19 38 25 📠 38 27

📍 23 Chemin des Capelles BP 87614 - 31076 TOULOUSE CEDEX 3

Lancez-vous !

Créer une archive ouverte institutionnelle n'est pas si difficile ! Il est important de comprendre que l'essentiel réside dans la sensibilisation de la communauté : nombreux sont encore les chercheurs français qui ignorent tout du mouvement de l'*open access* et qui, une fois bien informés, sont rapidement convaincus. Le deuxième point important est de ne pas se laisser effrayer par les aspects techniques : un ingénieur d'études (IGE) à double compétence, informatique et documentaire, est capable d'adapter un logiciel libre en une dizaine de jours. Le troisième point essentiel est la définition du *workflow*, l'organisation humaine nécessaire à son accomplissement. Enfin, *last but not least*, il faut savoir rester PRAGMATIQUE ! OATAO a démarré avec un outil opérationnel à l'instant T pour répondre rapidement à des besoins exprimés. L'objectif reste de migrer vers ORI-OAI, outil plus sophistiqué et plus intégrable dans le système d'in-

formation des établissements. Il ne faut jamais perdre de vue la réactivité et la souplesse pour répondre aux demandes de ceux que nous servons :

les chercheurs, les laboratoires, les établissements.

Les équipes des professionnels de la documentation, avec leur grande expérience de la gestion des données structurées, du contrôle qualité, de la conservation des contenus et du travail en réseau, sont les meilleurs garants d'une archive ouverte institutionnelle de qualité. Le métier évolue, les compétences restent !

M. Titonel

magali.titonel@ensat.fr

J.-M. Le Behec

jean-marie.lebehec@inp-toulouse.fr

C. Forestier

catherine.forestier@inp-toulouse.fr

Magali Titonel, de l'ENSAT (www.ensat.fr), est coresponsable de l'équipe « pros doc ». ☎ 05 34 32 31 17

Ingénieur d'études, Jean-Marie Le Behec est chef du projet OATAO.

☎ 05 34 32 31 16

À l'INP de Toulouse, Catherine Forestier est directrice du SCD

☎ 05 34 32 31 15 📠 31 00

OATAO oatao@inp-toulouse.fr

oatao.univ-toulouse.fr

* Parallèlement, deux thèmes progressent dans le Landerneau de l'enseignement supérieur et de la recherche :

open access pour l'un, gouvernance et indicateurs pour l'autre. ArXiv a déjà plus de 10 ans, PubMedCentral est né en 1999, HAL en 2001, la première version du protocole OAI-PMH est rendue publique la même année.

La diffusion et la visibilité de l'information scientifique deviennent une exigence pour les bailleurs de fonds publics et une demande de la société civile. Par ailleurs, la pression sur la gouvernance et les indicateurs des acteurs de l'action publique augmente : la LOLF est votée en juillet 2001, les acteurs de l'enseignement supérieur et de la recherche sont concernés. La mesure de la production scientifique devient un des critères majeurs d'évaluation par les tutelles.