

# Ar(abes)ques

FÉVRIER – MARS 2017

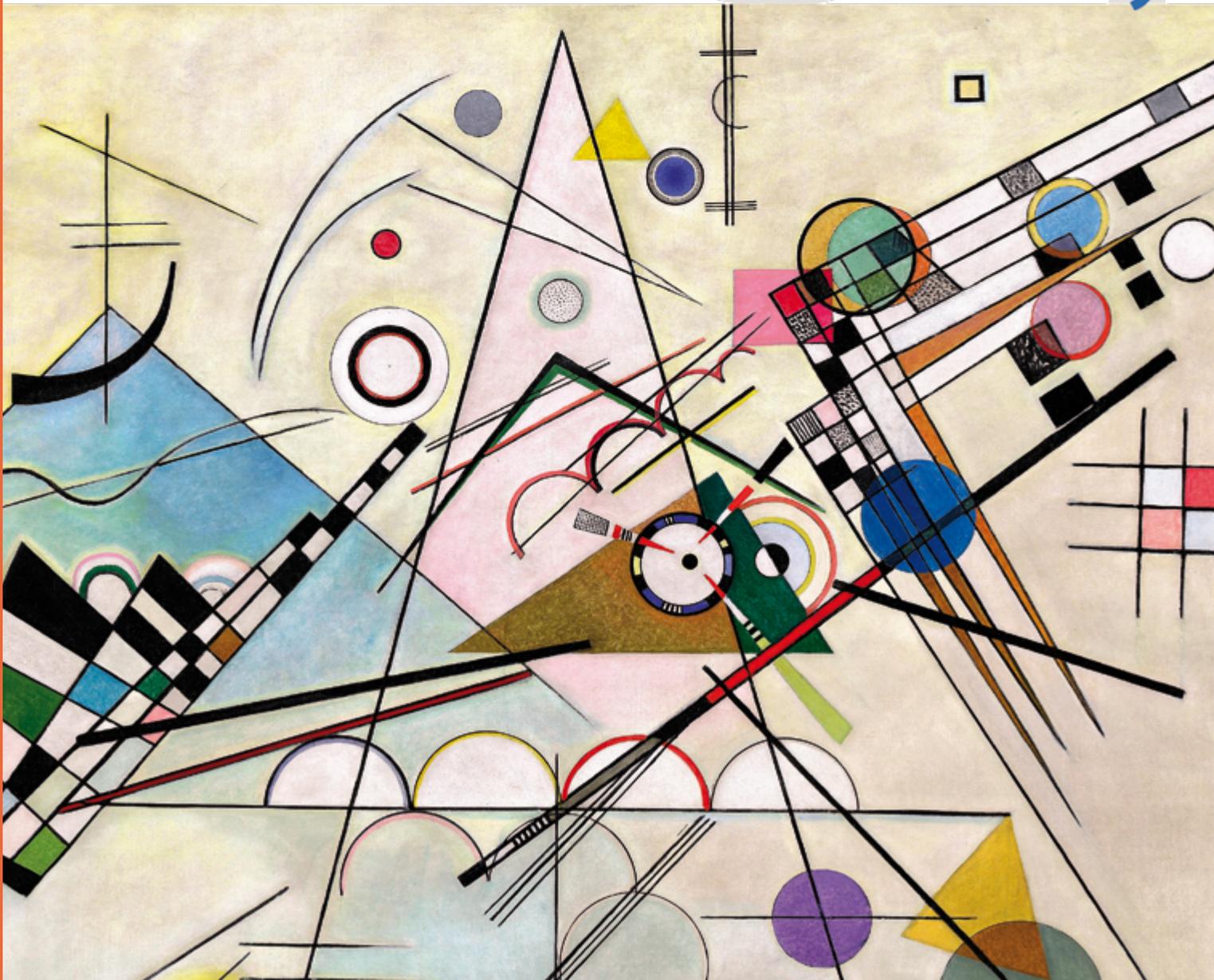
DOSSIER

## Politique nationale de l'IST *Des infrastructures en cohérence*

PLEINS FEUX SUR • Lilliad, Learning center Innovation

ACTUALITÉS • SGBm • EAD • Cercles-Bacon • Urfist de Bordeaux

84  
abes  
agence bibliographique  
de l'enseignement supérieur



## (Dossier) POLITIQUE NATIONALE DE L'IST *Des infrastructures en cohérence*

Elaborée en décembre 2008 par le ministère de l'Enseignement supérieur et de la recherche dans le cadre du Forum européen sur les infrastructures de recherche, la première feuille de route nationale pour les très grandes infrastructures de recherche (TGIR) donnait sa place à l'information scientifique et technique (IST) en posant les fondations de la Bibliothèque scientifique numérique (BSN). L'édition 2016 de la feuille de route « Stratégie nationale des infrastructures de recherche » offre à l'IST une visibilité renforcée et met en cohérence plusieurs infrastructures ayant déjà fait la preuve de leur capacité à répondre aux défis des changements de paradigme liés à la « transition numérique ». Afin de guider ses lecteurs dans ce maillage renouvelé, *Arabesques* donne la parole aux principaux acteurs qui vont structurer le paysage documentaire des années à venir.

## 20 (Pleins feux sur...)

### Lilliad, Learning center Innovation



© Atmosphère Photo

22

## (Actualités)

28

## (Portrait)

- 04 La politique nationale de l'IST au service de la recherche  
SOPHIE MAZENS
- 06 CollEx : promouvoir l'excellence documentaire  
VALÉRIE TESNIÈRE ET FRANÇOIS CAVALIER
- 08 Persée : de la numérisation à la production de connaissances  
NATHALIE FARGIER
- 09 Alliance Athéna : une réflexion à long terme sur les SHS  
JEAN-FRANÇOIS BALAUDÉ
- 10 HAL : une archive ouverte tournée vers ses utilisateurs  
CHRISTINE BERTHAUD ET AGNÈS MAGRON
- 12 OpenEdition : comprendre le monde en accès ouvert  
MARIN DACOS
- 14 Numédif : au service de l'édition numérique  
DOMINIQUE ROUX
- 15 Huma-Num : l'humain en majesté numérique  
STÉPHANE POUYLLAU
- 17 L'API Istex : le sésame pour accéder aux ressources acquises  
CLAUDE NIEDERLENDER

## Ar(abes)ques

REVUE TRIMESTRIELLE DE L'AGENCE BIBLIOGRAPHIQUE DE L'ENSEIGNEMENT SUPÉRIEUR  
227, avenue du Professeur-Jean-Louis-Viala, CS 84308, 34193 Montpellier cedex 5  
Tél. 04 67 54 84 10 / Fax 04 67 54 84 14 / [www.abes.fr](http://www.abes.fr)  
Directeur de la publication : David Aymonin  
Coordination éditoriale et secrétariat de rédaction :  
Danièle Stantcheva ([dstantcheva.abes@free.fr](mailto:dstantcheva.abes@free.fr))  
Comité de rédaction : Jean-Marie Feurtet, Christine Fleury, Philippe Le Pape, Cyril Leroy, Maryse Picard, Marie-Pierre Roux, Bertrand Thomas.  
Conception graphique : Anne Ladevie ([anneladevie.com](http://anneladevie.com))  
Impression : Pure Impression  
Couverture : Vassily Kandinsky, *Composition VIII* (détail), huile sur toile, 1923, Solomon R. Guggenheim Museum, New York.

Revue publiée sous licence Creative Commons CC BY-ND 2.0 (Paternité - Pas de modifications) sauf pour les images qui peuvent être soumises à des licences différentes ou à des copyrights.

Les opinions exprimées dans *Arabesques* n'engagent que la responsabilité de leurs auteurs.

ISSN (papier) 1269-0589 / ISSN (web) 2108-7016

# Bonne année 2017 !

(« où l'on va voir ce que l'on va voir »)

Cette année du Coq, selon l'horoscope chinois – une grande année donc pour la France, dont c'est en quelque sorte l'animal totem encore plus que l'emblème<sup>1</sup> – s'annonce d'ores et déjà comme une année charnière. Elle marque la fin d'un cycle commencé en 2010 – date symbolique que les amateurs de science-fiction apprécieront –, à mi-chemin entre 2009, quand fut créée la BSN, et 2011, quand démarra le projet SGBm. Une période d'étude, de conception, de mise au point et de tests qui a vu s'élaborer de nombreuses solutions et innovations en matière de gestion de l'information et de la documentation. Chacune se destine à faciliter le passage de la gestion de la bibliographie papier à la production et à l'exploitation des métadonnées de l'information scientifique et technique. En témoignent les gadgets géniaux (API et web services) et les réacteurs

commun leurs nouveaux savoir-faire en matière de Web sémantique pour bâtir le socle du futur Fichier national d'entités (FNE), qui permettra d'exposer les plus grands référentiels documentaires jamais créés en France et d'alimenter les bases de données et de connaissance de tous les opérateurs publics ou privés œuvrant dans le domaine de l'information scientifique. Pour couronner tout ceci, la législation elle-même changera, avec la publication des décrets d'application de la loi n° 2016-1321 du



7 octobre 2016 pour une République numérique, qui rendra quasi obligatoire la publication scientifique en libre accès, et fera du text and data mining une opération presque aussi courante que la recherche bibliographique.

## Une année décisive, où tout le paysage français de l'IST va se transformer, et où notre activité contribuera plus que jamais à la construction d'une société numérique ouverte, libre et évoluée.

Que ces longues phrases abstraites ne vous découragent pas de lire ce numéro d'*Arabesques*, destiné à disséminer auprès des cadres, responsables et décideurs

de données sophistiqués (Hub, IdRef V2, Bacon), développés avec passion dans les ateliers de l'Abes au bénéfice de sa communauté.

Et en cette année décisive, comme vous le révèle cette livraison épatante d'*Arabesques*, c'est tout le paysage français de l'IST, dans l'enseignement supérieur et dans le monde de la recherche en général, qui s'apprête à se transformer.

Des initiatives structurantes, pensées depuis des années dans le cadre de la BSN, aboutiront d'ici à quelques mois à des projets concrets – CollEx-Persée, Istex, Conditor... – dont les résultats transformeront encore davantage les conditions d'accès et d'exploitation de l'IST. Dans le monde des bibliothèques universitaires, avec le déploiement tant attendu des solutions SGBm sur les neuf sites pilotes, ces outils intégrés, qui apportent aux administrateurs et utilisateurs de systèmes de gestion de bibliothèques des fonctionnalités nouvelles et une meilleure visibilité pour les ressources documentaires électroniques, vont enfin fonctionner. Les grands partenaires de l'information bibliographique française mettront en

du monde des bibliothèques universitaires et des acteurs de l'information scientifique les nouvelles connaissances et les nouveaux concepts de base. L'intention du comité de rédaction est de les éclairer afin de les accompagner dans la conduite du changement dans leurs établissements, de les aider à inventer avec leurs équipes de nouveaux outils, de nouveaux fonctionnements, de nouveaux services. Nos données n'ont jamais été autant reliées au Web, comme notre activité n'a jamais autant contribué à la construction d'une société numérique, ouverte, libre et évoluée. Ne doutons pas que ces ambitions se réaliseront, grâce à l'engagement, l'énergie, la créativité, le sens du service de toute une communauté professionnelle... si bien sûr la politique, Donald et Vladimir restent raisonnables.

Belle année en perspective !

**DAVID AYMONIN**

Directeur de l'Abes  
david.aymonin@abes.fr

[1] Pour les anthropologues en herbe :  
[www.newworldencyclopedia.org/entry/Totemism](http://www.newworldencyclopedia.org/entry/Totemism)

# La politique nationale de l'IST au service de la recherche

En cohérence avec la volonté européenne de faciliter l'open access, la France se dote d'outils stratégiques : une feuille de route où sont inscrites les infrastructures de recherche d'envergure nationale, un comité de pilotage numérique et une loi, la bien nommée loi pour une République numérique. Une évolution très attendue par la communauté scientifique.

La politique nationale de l'information scientifique et technique est fondée sur la volonté d'ouvrir l'accès aux résultats de la recherche. Mais favoriser l'accès ouvert amène à traiter un sujet à multiples facettes, ce qui demande des compétences variées et spécialisées et une organisation articulée autour de structures fédératrices, représentatives d'enjeux de trois ordres, économique, juridique et technique.

Les premiers sont une préoccupation constante des établissements : le poids des budgets de la documentation numérique est tel que, pour maintenir un niveau d'offre ambitieux, l'accélération du libre accès devient incontournable.

Les enjeux juridiques imposent le respect du droit de la propriété intellectuelle et les modifications législatives nécessaires pour aller vers l'accès ouvert.

Les enjeux techniques sont quant à eux

fondamentaux pour promouvoir une large diffusion de l'IST et répondre aux besoins de la recherche.

La stratégie nationale de l'IST s'incarne dans des instances et des organisations qui fédèrent largement les acteurs de cette politique d'accès ouvert. Ainsi, les infrastructures de recherche suivent une feuille de route nationale et travaillent en lien étroit avec l'Europe. Chacune en déduit des programmes et des objectifs, évalués et remis à jour de façon régulière. Afin de piloter au bon niveau la transformation numérique de l'enseignement supérieur, le ministère de l'Éducation nationale, de l'enseignement supérieur et de la recherche (MENESR) a créé un comité d'orientation du numérique, dont une des composantes est la Bibliothèque scientifique numérique (BSN). Enfin, la promulgation de la loi pour une République numérique fait diminuer la

durée des barrières mobiles de la diffusion des textes scientifiques et facilite la fouille de données.

## TGIR, IR, PROJETS... SUR LA FEUILLE DE ROUTE

Rédigée dans le but de fournir au MENESR des outils nécessaires au pilotage des grands instruments de la recherche, la première feuille de route française des infrastructures de recherche a été publiée en 2008, puis mise à jour en 2012 et en 2016. Elle s'adresse aux Alliances et organismes ou établissements sous tutelle du ministère et les décline en quatre modèles : organisations internationales, très grandes infrastructures de recherche (TGIR), infrastructures de recherche (IR) et projets. La majorité des États membres de l'Union européenne se livrent au même exercice stratégique et contribuent à la feuille de route stratégique européenne.

Les infrastructures de recherche relèvent des choix des différents opérateurs de recherche et sont mises en œuvre par eux, qu'il s'agisse des Alliances ou de leurs membres, ou d'établissements publics en raison de leurs missions particulières. Quatre-vingt quinze d'entre elles ont été retenues dans le cadre de la feuille de route ESFRI 2016 (*European Strategic Forum for Research Infrastructures Roadmap 2016*). Parmi elles, quatre grandes infrastructures – **CollEx-Persée, HAL, Numédif et OpenEdition** – forment le chapitre « Information scientifique et technique ».

Les TGIR relèvent également d'une stratégie gouvernementale. Elles sont nationales ou font l'objet de partenariats européens ou internationaux, notamment par leur engagement dans la feuille de route du Forum stratégique européen. Instruments majeurs dans les réseaux de collaboration industrielle et d'innovation, elles se placent sous la responsabilité scientifique des opérateurs de recherche. Dans le domaine des sciences



STRATÉGIE NATIONALE  
DES INFRASTRUCTURES DE RECHERCHE  
ÉDITION 2016

Un « tableau de bord » qui fournit à l'État une base solide pour des plans d'action pluriannuels.

humaines et sociales, **Huma-Num** s'inscrit dans les TGIR. Elle vise à faciliter le tournant numérique de la recherche en sciences humaines et sociales, et est soutenue par l'**Alliance Athena**.

La feuille de route 2016 des infrastructures de recherche présente clairement les enjeux de la thématique « Information scientifique et technique » : « La production, le stockage et la mise à disposition de données sont des paramètres essentiels de la recherche d'aujourd'hui ; c'est particulièrement vrai pour les infrastructures de recherche. Certaines sont focalisées sur le numérique, en tant qu'objet des recherches elles-mêmes ou pour en développer les outils de calcul intensif, de transmission ou de stockage des données. D'autres ont pour but de mettre à disposition des bases de données, qu'elles soient brutes ou enrichies. Dans tous les cas, elles tendent à devenir découvrables, utilisables et interrogeables par une communauté toujours plus large. Dans certains domaines, cette mise à disposition est immédiate et entièrement publique, suivant une directive européenne. Dans d'autres, une période d'embargo est la pratique courante avant dissémination. »

## LES QUATRE PILIERS DU SCHÉMA NUMÉRIQUE

Le **Codornum** (Comité d'orientation numérique) est l'organe décisionnel, mis en place en 2014 par le MENESR afin de prendre des décisions d'orientation à un haut niveau stratégique et politique sur les sujets en relation avec la transformation numérique de l'ESR. Il s'assure, par un suivi régulier, que l'ensemble des dispositifs nationaux du numérique de l'ESR prenne les mesures adéquates en réponse aux orientations décidées.

Quatre comités de pilotage participent à cette instance : **InfraNum**, en charge de définir les orientations en matière d'infrastructures techniques nationales (calcul haute performance, datacenter, etc.) ; **SIESR**, le service interacadémique qui coordonne les évolutions et l'interopérabilité des systèmes d'information de l'enseignement supérieur et de la recherche ; le comité **Formation**, chargé en particulier de la coordination des portails ESUP et Fun-Mooc ; et enfin le comité **BSN**, chargé de la documentation et information scientifique et technique.

La BSN, créée en 2009 à l'initiative du MENESR, a pour mission de fédérer les acteurs des universités et organismes de recherche afin de permettre à l'ensemble des partenaires d'accéder, sous forme numé-

rique, à une offre de ressources scientifiques abondante et à des outils de qualité. Elle met en réseau un certain nombre d'initiatives, avec une volonté de structuration destinée à dépasser les traditionnels obstacles liés à l'organisation française, très éclatée, de l'information scientifique et technique. Elle pilote des programmes de mutualisation des ressources électroniques organisés en segments, tel le programme Istex, qui concerne l'achat de ressources dans le cadre de licences nationales et la réalisation d'une plateforme autonome accessible à la communauté de la recherche. La BSN soutient les plateformes de libre accès comme HAL, Persée ou OpenEdition et des projets autour de la bibliométrie et de la mesure des usages des ressources numériques.

## L'OUVERTURE INSCRITE DANS LA LOI

La loi pour une République numérique vise à favoriser l'ouverture et la circulation des données et du savoir, à garantir un environnement numérique ouvert et respectueux de la vie privée des internautes et à faciliter l'accès des citoyens au numérique. Promulguée le 7 octobre 2016 par le président de la République, elle paraît au

les contenus ont été acquis de façon licite ; il ne peut être pratiqué que pour « les besoins de la recherche publique » ; en est exclue « toute finalité commerciale ». Associés à la loi, deux décrets relatifs au TDM sont prévus. Leur objectif est de sécuriser juridiquement, techniquement et commercialement l'utilisation du TDM dans un équilibre respectant les intérêts et les besoins respectifs des chercheurs (larges extractions des ressources) et des éditeurs (contrôle des copies). La promulgation de cette loi est concomitante avec le travail sur la révision de la directive européenne sur le droit d'auteur dans le marché unique européen. En effet, le projet de directive présenté par la Commission européenne – et applicable à tous les États membres – conforte les formulations retenues par le législateur français en prévoyant une exception obligatoire au droit d'auteur, favorable au TDM.

La politique nationale de l'*open access* prend corps dans de nombreuses structures et projets – *a priori* foisonnants, mais qui répondent à la volonté de faciliter la diffusion de la culture scientifique dans tous ses aspects. Une stratégie nationale de développement s'impose, qui s'efforce de tenir compte des nouveaux apports technologiques et de

## La loi pour une République numérique ouvre de nouveaux droits et de nouvelles facilités dans les accès à l'information issue de la recherche scientifique sans remise en cause du droit d'auteur.

*Journal officiel* du 8 octobre 2016.

Elle comprend des avancées significatives autour du TDM (*text and data mining*, ou fouille de données) et ouvre de nouveaux droits et de nouvelles facilités dans les accès à l'information issue de la recherche scientifique. Son article 30 entérine ainsi un arbitrage favorable à la recherche sur l'accès ouvert, permettant un droit de valorisation secondaire, sans remise en cause du droit d'auteur, ouvrant aux chercheurs la possibilité de mettre à disposition gratuitement la version de l'auteur « acceptée pour publication » (mais non travaillée par l'éditeur) d'un écrit scientifique, au terme d'un délai de six mois pour les STM et de douze mois pour les SHS.

En outre, l'article 38 introduit une nouvelle exception TDM, inscrite dans le code de la propriété intellectuelle et encadrée juridiquement par trois limitations : le TDM ne peut s'effectuer qu'à partir de corpus dont

l'évolution des pratiques scientifiques. Elle doit s'articuler à une recherche devenue mondiale et se traduire par une participation active aux réseaux européens et internationaux. Elle nécessite une organisation qui permette l'émergence de collaborations et d'un réseau de partage de la connaissance.

**SOPHIE MAZENS**

*Chef du département de l'information scientifique et technique et réseau documentaire du MENESR*  
sophie.mazens@enseignementsup.gouv.fr



**POUR EN SAVOIR PLUS**

Consulter la feuille de route : [cache.media.enseignementsuprecherche.gouv.fr/file/Infrastructures\\_de\\_recherche/74/5/feuille\\_route\\_infrastructures\\_recherche\\_2016\\_555745.pdf](http://cache.media.enseignementsuprecherche.gouv.fr/file/Infrastructures_de_recherche/74/5/feuille_route_infrastructures_recherche_2016_555745.pdf).  
La section consacrée à l'IST commence en page 147.

L'infrastructure CollEx, appuyée sur la plateforme de numérisation Persée, s'attelle à la valorisation des fonds scientifiques et gisements documentaires inexploités. Au sein du GIS en formation, elle prépare un saut qualitatif par rapport aux actions passées des Cadist, dont elle signe la fin.

# CollEx : promouvoir l'excellence documentaire

Le dispositif CollEx (collections d'excellence pour la recherche)<sup>1</sup>, que le ministère de l'Éducation nationale, l'enseignement supérieur et la recherche (MENESR) met actuellement en place, est parti de plusieurs constats. Créé il y a trente ans, le dispositif Cadist (centres d'acquisition et de diffusion de l'IST) était destiné à assurer une couverture documentaire pour la recherche focalisée sur la production étrangère académique, dans un contexte où l'imprimé était dominant. Assorti d'une obligation de signalement et de fourniture de documents (les prêts entre bibliothèques, ou PEB), il offrait une gamme de services limitée. La révolution numérique et l'essor des licences nationales pour la documentation électronique l'ont rendu obsolète.

Plus largement, c'est le constat d'un bouleversement profond des pratiques des chercheurs, souligné dans l'enquête du Bureau VanDijk<sup>2</sup>, qui a poussé à revoir de fond en comble la politique du ministère. En effet, les ressources offertes sur le Web constituent le premier horizon de la plupart des chercheurs, non seulement en sciences et techniques mais aussi en lettres, sciences humaines et sociales, alors qu'il existe dans les bibliothèques et les centres de documentation s'adressant à la recherche bien des gisements documentaires insuffisamment exploités. Pour partie signalés dans les outils nationaux, ils demeurent insuffisamment référencés sur le Web. Dernier constat, les Cadist font la part belle aux bibliothèques universitaires

et laissent de côté les fonds de recherche des laboratoires ou encore de grands établissements tels que la BNFOU les Eprist. Autrement dit, les Cadist n'ont pas réussi à constituer un véritable réseau documentaire national de la recherche.

## UN SOUTIEN PLURIANNUEL

Il aurait été possible de supprimer le dispositif sans autre forme de procès, en misant sur une forme d'autorégulation du Web et en investissant exclusivement dans les licences nationales. Ayant investi dans ce dernier secteur avec détermination, l'Allemagne a cependant choisi de ne pas abandonner et de faire évoluer son propre modèle, dont les Cadist se sont d'ailleurs inspirés, en le reliant beaucoup plus étroitement aux modes de fonctionnement des programmes de recherche (appels à projets pluriannuels mettant l'accent autant sur la documentation que sur une gamme élargie de services aux chercheurs). En France, le programme des CollEx ne s'inscrit pas dans le même contexte politique, le CNRS et la DFG<sup>3</sup> n'ayant pas les mêmes missions : il a été décidé d'en faire un levier supplémentaire pour le rapprochement entre les universités et les structures de recherche, en accord avec la mission sur la documentation scientifique et technique confiée à l'Alliance Athena. L'idée d'un nouveau réseau documentaire national pour la recherche est ainsi née.

Du nouveau modèle allemand a été repris le principe d'un dispositif de soutien pluriannuel, en phase avec les grands programmes nationaux et internationaux de la recherche. Il ne s'agit plus désormais de financements récurrents pour la documentation courante. Le second objectif consiste à développer la gamme de services aux chercheurs, non seulement pour les inciter à utiliser des ressources méconnues, mais aussi pour s'inscrire au plus près de leurs attentes. La dimension numérique est également à renforcer, notamment *via* Persée.

Enfin, pour que le futur réseau se prenne en main, le ministère a choisi de donner à la future infrastructure de recherche la forme d'un groupement d'intérêt scientifique (GIS), composé de bibliothèques délégataires et des opérateurs nationaux que sont l'Abes, le Centre technique du livre de l'enseignement supérieur (CTLes), la BNF..., et

↳ Développer des services au plus près des besoins des chercheurs, un enjeu de taille pour CollEx.

Alain Bachelier/Flickr (CC BY-NC-ND 2.0)



avec les infrastructures du monde de la recherche, comme Huma-Num ou Persée. En articulation avec ces opérateurs, le GIS a pour mission d'organiser, avec l'appui du MENESR, la coopération entre les grandes bibliothèques patrimoniales et scientifiques et de développer des services répondant aux besoins des chercheurs.

## CARTOGRAPHIER LES COLLECTIONS

Pour rapprocher les chercheurs des gisements documentaires scientifiques et leur en faciliter l'usage, CollEx-Persée aura comme objectif la réalisation d'une cartographie nationale des fonds intéressant la recherche, alimentée par le réseau et soutenue par les opérateurs nationaux. Il a ainsi pour mission d'identifier les collections d'intérêt national en leur attribuant un label, en fonction de leur intérêt, leur originalité ou leur rareté et de contribuer à leur développement concerté et leur visibilité. CollEx-Persée s'intéresse de ce fait à la conservation partagée des fonds et à la numérisation. Pourquoi des bibliothèques délégataires ? Sélectionnées à l'issue d'un appel à manifestation d'intérêt, clos en décembre 2016, en fonction de la richesse et de l'originalité de leurs collections, de leur implication dans le monde de la recherche, de leurs initiatives en matière d'élargissement des collections à la collecte de matériaux documentaires pour la recherche et de leur rayonnement, elles animeront des réseaux spécialisés au sein du GIS et conduiront des programmes associant bibliothèques labellisées et chercheurs, en lien avec les opérateurs. La gouvernance du GIS, à laquelle elles participent, assure la cohérence des interventions.

Les domaines d'intervention de CollEx associent collections et services, y compris dans leur dimension numérique. CollEx vise à valoriser l'hybridation des matériaux de la recherche, quel que soit leur support, et à redonner aux collections des bibliothèques la visibilité qu'elles méritent et un accès facilité. À cet égard, CollEx se comportera comme un véritable instrument de politique documentaire nationale, à l'interface de la Bibliothèque scientifique numérique et de l'action des grands opérateurs nationaux.

Les initiatives de CollEx portent sur l'organisation et l'urbanisation des collections imprimées dans le périmètre documentaire national. Ainsi la cartographie évoquée plus haut fournira-t-elle l'occasion d'inscrire les collections – dans leur acception plus générale de ressources et matériaux pour la recherche – dans le panorama de l'activité de la recherche en croisant un annuaire thématique de la recherche et des informations documentaires. Par ailleurs, le développement des plans de conservation partagée permettra aussi de dessiner une carte de la conservation et de fournir une aide au fonctionnement du PEB. Ils présentent à la fois l'intérêt de débloquent des mètres carrés précieux

pour l'accueil des publics et de services et celui de dessiner un paysage de ressources signalées, gérées et accessibles à la demande.

## INITIATIVES ORIENTÉES RECHERCHE

La numérisation constitue un autre axe à développer, sur lequel Persée apportera une expertise et un concours précieux pour la réalisation de programmes de numérisation centrés sur les besoins de la recherche. Ces ressources, rendues ainsi plus accessibles et donc susceptibles d'être largement popularisées auprès des équipes de recherche, faciliteront l'exploitation, au moyen, notamment, d'outils de fouille de texte. L'autre manière, complémentaire, de donner plus de visibilité à ces ressources, consiste à améliorer leur référencement sur le Web ; cette action est un élément critique depuis que le Web est devenu la porte d'entrée naturelle à l'information, y compris pour les chercheurs. Les établissements CollEx auront une responsabilité particulière à cet égard et le GIS aura à encourager les initiatives en ce domaine. Enfin, les dossiers visant à créer des services nouveaux et innovants pour la recherche pourront recevoir le soutien de CollEx et servir d'expérimentations qui, en cas d'évaluation favorable, seraient reproduites à plus grande échelle.

CollEx doit à la fois redonner force et vitalité au réseau documentaire et à ses ressources et permettre de conduire des initiatives nouvelles, clairement orientées recherche, pour accompagner les projets des chercheurs et leur en faciliter la réalisation. La présence, au sein du futur GIS CollEx, de l'Alliance Athena, qui prendra la responsabilité de suivre les développements documentaires pour le compte de toutes les Alliances, est la garantie d'une meilleure articulation des services, ressources et matériaux collectés avec le monde de la recherche pour réaliser un saut qualitatif par rapport à l'action passée des Cadist.

Incontestablement, CollEx-Persée est riche de potentialités structurantes pour le monde documentaire au service de la recherche et peut s'avérer un puissant outil de coordination de la politique documentaire au travers d'une alliance renouvelée avec la recherche. Mais il faudra, pour réussir, à la fois bien sélectionner les opérations prioritaires et novatrices et élargir le socle des moyens mobilisés pour donner une véritable dimension nationale et internationale à ce programme.

### VALÉRIE TESNIÈRE

Directrice de la BDIC  
Coordinatrice du groupe de travail CollEx  
Cartographie et critères  
valerie.tesniere@bdic.fr

### FRANÇOIS CAVALIER

Directeur de la bibliothèque de Sciences Po  
Président du Comité opérationnel CollEx  
francois.cavalier@sciencespo.fr

[1] [www.collex.eu](http://www.collex.eu)

[2] Bureau van Dijk Information Management, MENESR, *Étude portant sur la politique nationale de soutien aux acquisitions de documentation recherche destinée à accompagner la mise en place du dispositif documentaire national CollEx, 2016.*

[3] L'agence allemande de moyens pour la recherche.

## LES PROCHAINES ÉTAPES

- **Constitution du GIS**, qui sera effective en début d'année 2017 après sélection des candidatures déposées fin décembre 2016
- **Choix d'un établissement** porteur du GIS, ayant une capacité opérationnelle à animer un réseau transdisciplinaire
- **Définition de sa feuille de route** par le GIS (neuf mois), à court et moyen terme, ainsi que des premières actions à conduire au niveau du réseau, à la fois sur les services transversaux (PEB...) et à partir des initiatives des bibliothèques et regroupements de bibliothèques via la procédure d'appels à projet
- **Élaboration des critères d'évaluation** de son action par le GIS pour produire un bilan qui permettra de décider des évolutions futures
- **Coordination et articulation** de l'activité du GIS avec les initiatives de la BSN, à définir de concert avec celle-ci sous l'égide du ministère

# Persée : de la numérisation à la production de connaissances

Labellisé infrastructure de recherche en lien avec CollEx, Persée, fort de son portail de référence, assurera le pilotage de la politique de numérisation, en associant les acteurs de l'ESR impliqués dans l'enrichissement, la diffusion et la préservation de contenus numérisés.

La numérisation des gisements documentaires, d'une grande richesse et complexité, rend accessibles et exploitables des corpus indispensables aux travaux de la communauté scientifique. Les chercheurs ont besoin d'un accès sans contraintes à des documents décrits, structurés, reliés entre eux par des index, des référentiels et des citations, avec la possibilité, actuellement plébiscitée, de manier et d'exploiter les données qui en sont issues. Persée, qui assume une mission générale de valorisation du patrimoine, est acteur de cette dynamique<sup>1</sup>. Si l'éventail de ses activités s'est ouvert, Persée demeure guidé par les principes définis lors de son lancement : favoriser les usages scientifiques du patrimoine documentaire, apporter une réelle plus-value aux documents numérisés et privilégier le libre accès.

du Web sémantique. Non seulement il contribue à leur visibilité mais aussi, et plus fondamentalement, il marque une évolution conceptuelle majeure et il offre un espace de liberté aux chercheurs : possibilité de récupérer les jeux de données qu'ils souhaitent, autonomie pour en créer, disparition des contraintes liées aux sites web traditionnels, dotés d'un design et d'une organisation figés en amont.

## NOUVELLES LOGIQUES

En 2016, Persée a été labellisé infrastructure de recherche en lien avec CollEx. L'ambition du dispositif CollEx-Persée est de faciliter l'accès et l'usage des ressources documentaires et de développer de nouveaux services articulés aux collections en lien direct avec les utilisateurs-chercheurs. Dans son domaine d'expertise, Persée assurera le pilotage de la politique de numérisation à mettre en œuvre en associant les acteurs de l'enseignement supérieur, de la recherche et de la culture qui sont impliqués dans la numérisation, l'enrichissement, la diffusion et la préservation de contenus numérisés (la BNF et Huma-Num en premier lieu). Cartographie des réalisations existantes, appels à projets, manifestations d'intérêt, contribution au développement des collections du portail Persée sont autant de pistes actuellement à l'étude pour fédérer un réseau de partenaires.

Que retenir de l'expérience de Persée ainsi dessinée à grands traits ? Initié pour créer une chaîne éditoriale et alimenter un portail web, Persée évolue en plateforme de recherche fédérant données, utilisateurs, outils et méthodes et bénéficiant d'une reconnaissance institutionnelle. L'évolution à l'œuvre témoigne du nouveau visage des collections numériques et des documents qui sont désormais des données reliées et éditorialisées suivant des logiques inconnues du monde de l'imprimé.

NATHALIE FARGIER

Directrice de l'UMS Persée  
nathalie.fargier@persee.fr

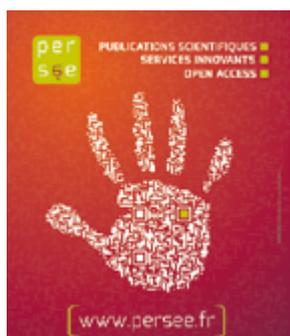


POUR EN SAVOIR PLUS

Le portail : [www.persee.fr](http://www.persee.fr)

Le triple store : [data.persee.fr](http://data.persee.fr)

Le serveur OAI : [oai.persee.fr/oai?verb=ListSets](http://oai.persee.fr/oai?verb=ListSets)



## PERSÉE EN CHIFFRES

- 22 disciplines représentées
- 246 collections diffusées
- 650 000 documents disponibles librement en texte intégral
- 6 millions de documents PDF téléchargés en 2016
- 37 millions de documents consultés en 2016

[1] Constitué en unité mixte de services depuis 2013, Persée a pour tutelles l'Université de Lyon, le CNRS et l'ENS de Lyon et bénéficie du soutien du MENESR.

[2] Ouverture, en 2017, d'athar.persee.fr (histoire de l'art) et d'education.persee.fr (histoire de l'éducation).

## LE PORTAIL ET LES PERSÉIDES

La réalisation la plus visible de Persée est le portail du même nom, qui réunit des collections de publications scientifiques de la première parution à la plus récente. Avec 650 000 documents en libre accès et plus de 30 millions de visites annuelles, ce portail de référence propose à un public croissant un panel étendu de publications scientifiques (revues, actes de colloque, livre, publications en série...), dans différentes disciplines, ainsi qu'une gamme d'outils de recherche et d'exploitation équivalents à ceux proposés par les portails d'édition courante. Cette expérience fondatrice a essaimé des bibliothèques scientifiques numériques présentant un haut niveau de spécialisation. Le label « Perséides » regroupe les projets réalisés en partenariat avec des équipes de recherche, des bibliothèques et des archives<sup>2</sup>. Des corpus adossés à des programmes de recherche sont numérisés, documentés, alignés sur des référentiels, diffusés par le biais de sites web dédiés et largement instrumentés pour en permettre une exploitation scientifique. L'enjeu actuel pour Persée est d'explorer d'autres modes d'exposition des données que les sites web et l'OAI et par là-même de favoriser de nouveaux usages.

Suivant cette tendance de fond, Persée a déployé [data.persee.fr](http://data.persee.fr). Ce triple store fédère les données décrivant l'ensemble des ressources produites par Persée de manière structurée selon les principes

# Alliance Athéna : une réflexion à long terme sur les SHS

Espace de concertation entre les grands acteurs institutionnels des sciences humaines et sociales, l'Alliance Athéna mène une réflexion prospective sur l'identification des atouts et ressources dans ce domaine scientifique, afin de faciliter et d'orienter les prises de décision.

**Le rôle de l'Alliance Athéna, créée en 2010,** est de conduire des missions de coordination et de réflexion stratégique et prospective touchant les sciences humaines et sociales. Il s'agit tout à la fois de rendre plus visibles les activités et productions de ce vaste champ scientifique, de mieux les valoriser, de contribuer à la définition des appels à projet nationaux ou européens et à l'élaboration de la stratégie nationale de recherche, incluant les infrastructures de recherche et les infrastructures en information scientifique et technique. Sur ce dernier plan, l'action de l'Alliance Athéna prolonge le mouvement amorcé dans les années 2000, qui visait à faire sortir la documentation, l'information et la communication scientifiques de leur situation en marge de l'économie globale de la recherche, et à structurer le secteur. Ce mouvement s'était traduit par l'élaboration en 2008 d'une feuille de route française sur les infrastructures de recherche qui avait reconnu l'IST comme un domaine à part entière doté d'infrastructures fondamentales pour la pratique scientifique.

## CONSOLIDER, RÉUNIR, STRUCTURER

La mise à jour en 2016 de cette feuille de route a été préparée à partir de 2014. Elle a été l'occasion d'un travail approfondi appuyé sur une consultation de la communauté scientifique à travers le réseau des Maisons des sciences de l'homme. À son terme, le directoire de l'Alliance Athéna a conclu qu'il convenait de ne plus retenir BSN comme infrastructure de recherche pivot, mais de consolider les deux TGIR existantes, Huma-Num et Progedo, de réunir en une infrastructure de recherche CollEx et

Persée, et d'en faire entrer trois autres dans la feuille de route : HAL, OpenEdition, et Numédiff. Ce sont ainsi quatre infrastructures de recherche en IST dont l'Alliance Athéna a soutenu l'émergence.

Elles forment un ensemble cohérent, chacune répondant aux critères d'infrastructure de recherche, et sont à même de constituer des ensembles complémentaires d'« usages-technologies » interagissant dans le processus de diffusion de la production scientifique. Ces plateformes permettent d'assurer une indépendance technologique, tout en se liant aux plateformes internationales (métadonnées partagées, accords de partenariat, dispositifs de réversion).

## ARTICULER LES COMPÉTENCES

La plateforme d'archive ouverte HAL est dédiée à la gestion des *preprints*, des multiples versions et à l'archivage de la production scientifique en libre accès, avec des règles d'articulation entre les systèmes d'archives locales institutionnelles et la plateforme HAL permettant aux établissements de bénéficier d'un portail personnalisé d'accès. Enrichie d'outils (Sciencesconf.org ; Episciences.org ; Héloïse), elle a rencontré depuis dix ans un succès croissant.

OpenEdition a, quant à elle, vocation à permettre la diffusion en libre accès de la production scientifique éditée, et l'expérimentation de nouvelles formes d'écriture scientifique (revues et ouvrages essentiellement SHS avec outils associés). Elle rassemble aujourd'hui quatre plateformes renvoyant à plus de 28 000 programmes scientifiques. Avec 50 millions de visites annuelles, OpenEdition est

une incontestable réussite. Son inscription dans la feuille de route nationale doit améliorer l'impact des projets de recherche pluridisciplinaires et leur capacité à transférer leurs résultats aux acteurs socio-économiques à un niveau européen.

Numédiff vient en complément : elle constitue une plateforme publique mutualisée de diffusion distribution multicanaux traitant des formes éditées multiples, selon divers périmètres (volumétriques, géographiques, de spécialité). Prescriptrice de normes et standards dans le domaine de l'édition et de la diffusion numérique, elle doit garantir la circulation des écrits scientifiques et leur mise à disposition optimale sous forme dématérialisée et matérielle en France et dans le monde entier. Une attention particulière est portée à la réalisation d'articulations avec HAL, OpenEdition et Persée, pour la circulation des contenus structurés et celle des métadonnées. Le modèle et les outils d'informatique éditoriale associés permettent déjà, à partir d'une source unique, de produire toutes les formes de diffusion.

Enfin, l'infrastructure CollEx-Persée, qui fait l'objet actuellement de deux appels à manifestation d'intérêt, pour constituer le GIS lui-même et sélectionner les bibliothèques délégataires, doit consolider les grandes collections scientifiques accessibles en ligne. Elle va, dans une stratégie à long terme, rassembler des bibliothèques qui font référence dans leurs disciplines (telles la Bulac, la BNU de Strasbourg, la BDIC, etc.). Appuyée sur la plateforme de numérisation Persée, elle gèrera et valorisera les grands fonds scientifiques thématiques. Elle va ainsi constituer un outil majeur des politiques de conservation partagée répondant à des logiques à la fois disciplinaires et de stratégies territoriales, en dialogue avec la BNF.

**JEAN-FRANÇOIS BALAUDÉ**

Président de l'Université Paris-Nanterre  
et de l'Alliance Athéna  
balaude@u-paris10.fr

## Processus de diffusion des résultats de la recherche



# HAL : une archive ouverte tournée vers ses utilisateurs

Partagée par l'ensemble de l'ESR et bénéficiant d'une forte visibilité à l'international, la plateforme collaborative HAL est basée sur l'auto-archivage. Elle développe des outils de création d'archives ouvertes – institutionnelles, thématiques... – et participe au projet Conditor, soutenu par la BSN.

Depuis maintenant seize ans, HAL (« hyper articles en ligne ») propose à la communauté scientifique un outil pour l'archivage et la diffusion en libre accès des résultats de la recherche. Début 2017, plus de 422 000 documents étaient disponibles en texte intégral : documents déjà publiés (articles de revue, ouvrages ou chapitres) ou non (preprint, working papers, rapports), communications et posters présentés à des congrès ou travaux universitaires (thèses et habilitations à diriger des recherches). Toutes les disciplines sont représentées : 380 champs disciplinaires, répartis en treize grands domaines scientifiques. L'objectif de transversalité conduit à archiver non seulement des données textuelles, mais aussi des images et des enregistrements vidéo et audio produits et utilisés dans le cadre de recherches. HAL est à la fois un entrepôt unique et une application, une plateforme mutualisée pour les archives institutionnelles, les archives ouvertes thématiques (HAL-SHS) et les thèses électroniques (TEL). Cet ensemble participe à constituer un historique patrimonial de l'enseignement supérieur et de la recherche. Les principes fondateurs de HAL sont conformes aux standards d'une archive ouverte : architecture OAI-PMH (Open Archive Initiative Protocol for Metadata Harvesting) qui garantit l'interopérabilité des données, la stabilité des identifiants et la pérennité des documents (assurée par le partenariat avec le CINES). Interconnecté avec les grandes archives

internationales thématiques comme arXiv, PubMed ou encore RePec, HAL offre un service apprécié des chercheurs : un dépôt effectué ici est automatiquement reversé dans ces outils de référence. Bien évidemment, HAL et ses portails répondent aux critères d'OpenAIRE et participent ainsi à la visibilité et l'accessibilité de la recherche française.

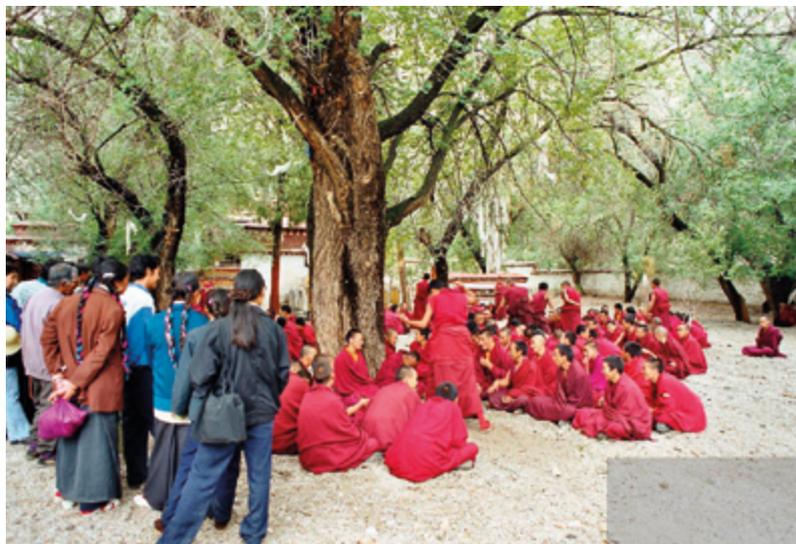
## TOUT AU LONG DE LA CARRIÈRE DU CHERCHEUR, DES RÉSULTATS RÉFÉRENCÉS

HAL est alimentée pour une grande partie par les auteurs des publications eux-mêmes. Leurs motivations sont diverses : volonté de diffuser largement leurs recherches, pratique bien ancrée dans leur discipline, incitation de leur organisme... Les fonctionnalités au service de leur présence numérique constituent aussi des incitations au dépôt : identifiant auteur (IdHAL) qu'ils créent eux-mêmes et qu'ils peuvent lier avec d'autres identifiants (IdRef, ORCID), leurs blogs et leurs profils sur les réseaux sociaux ; CV doté d'une URL pérenne, avec un affichage dynamique des publications déposées. De plus, ils y trouvent une base de données unique, dans laquelle ils peuvent référencer leurs productions tout au long de leur carrière, et qui leur est accessible quelles que soient leurs affectations. Mais de plus en plus souvent, les unités de recherche, les établissements et les institutions organisent aussi des dépôts ; des professionnels IST se consacrent aux saisies ou imports, à la formation au dépôt, etc. L'objectif est non seulement d'avoir une vision précise de la recherche produite, mais aussi de la valoriser.

## PORTAILS, COLLECTIONS... DES OUTILS ADAPTÉS AU PAYSAGE FRANÇAIS

L'affiliation des auteurs dans les dépôts a été un choix dès la création de HAL. Elle correspond à la façon spécifique dont la recherche française est généralement organisée : en unités de recherche relevant de plusieurs tutelles. L'ensemble de la communauté scientifique et universitaire a ainsi pu être représenté dans HAL et s'est approprié l'outil. Un laboratoire de recherche, par exemple, utilisera la fonctionnalité de collection : site web, dans lequel sont affichés les dépôts de ses membres, avec une charte graphique personnalisée, et qui peut être organisé par projet de recherche, par équipes...

MediHal est une archive ouverte, intégrée à HAL, qui permet de déposer des données visuelles et sonores. Ici, *Joute oratoire de moines tibétains*, photo prise au monastère de Sera, au Tibet, par Elisabeth de Pablo (CFI-Inalco).



CC BY-NC-ND 4.0

La fonctionnalité de portail, plus particulièrement destinée aux établissements et organismes, permet de créer et gérer en toute autonomie une archive institutionnelle. Il s'agit d'un sous-ensemble de HAL qui n'affiche que les dépôts des auteurs affiliés à l'établissement. Le bénéfice immédiat est que, dès sa création, l'archive institutionnelle peut déjà afficher des publications si des chercheurs ou enseignants-chercheurs ont déjà déposé. Toutes les thèses seront également disponibles : qu'elles soient archivées par les auteurs eux-mêmes sitôt la soutenance passée, ou bien si l'établissement a choisi HAL pour leur diffusion. La plateforme Star, en effet, les exporte automatiquement dans TEL, et donc dans HAL, grâce à un partenariat Abes-CCSD. Par ailleurs, un établissement qui gère déjà une archive institutionnelle en local s'engage à reverser dans HAL, pour bénéficier notamment de sa visibilité internationale. C'est le cas, par exemple, d'OATAO (pôle de recherche et d'enseignement supérieur de Toulouse), ou de Spire (Sciences Po). Ce mode de mutualisation est prévu par la convention de partenariat inter-établissements issue des travaux de la BSN. Cette convention a été signée en 2013 par la Conférence des présidents d'université, la Conférence des grandes écoles, l'Agence de mutualisation des universités et établissements et 23 autres établissements.

## UN RÉSERVOIR COMMUN, ALIMENTÉ AU FIL DE L'EAU

Cette philosophie collaborative est également au cœur du projet Conditor, soutenu par la BSN et dont HAL sera un pilier. Son objectif est de mieux identifier la production de l'enseignement supérieur et la recherche et de permettre la mutualisation des travaux de recensement de la production scientifique

effectués au sein des institutions. Sur un modèle de collaboration interétablissements, on prévoit un réservoir commun de métadonnées, articulé avec HAL, alimenté au fil de l'eau par les sources existantes, signalant l'ensemble de la production de la recherche publique française, en s'appuyant sur des référentiels communs. HAL est ainsi légitimée comme archive commune et partagée, au cœur de la coordination nationale en faveur des archives ouvertes. Son inscription dans les infrastructures de l'IST en 2016 est un signe fort de reconnaissance et lui apporte une visibilité à long terme. La feuille de route du CCSD pour 2016-2020 poursuit une stratégie renforcée dans le sens d'un projet collaboratif, avec des choix de développements visant à faciliter l'utilisation de HAL pour le chercheur et l'institution, à pérenniser son rôle d'infrastructure commune et partagée tout en l'inscrivant dans le paysage international.

**CHRISTINE BERTHAUD**

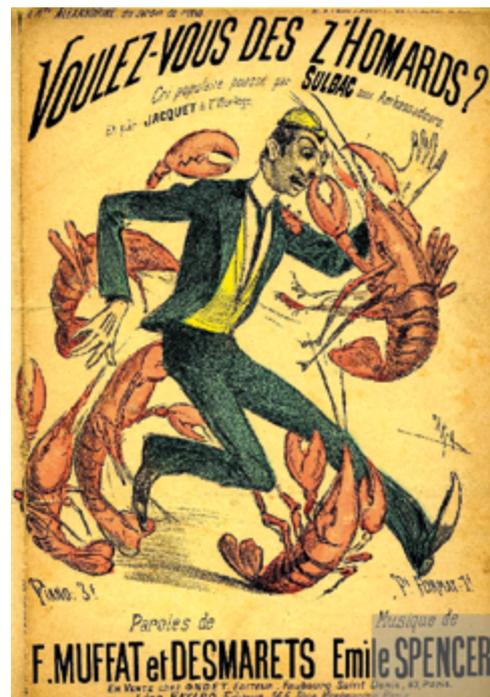
Directrice du CCSD, copilote de BSN4  
christine.berthaud@ccsd.cnrs.fr

**AGNÈS MAGRON**

Responsable communication, CCSD  
agnes.magron@ccsd.cnrs.fr



**POUR EN SAVOIR PLUS** [hal.archives-ouvertes.fr](http://hal.archives-ouvertes.fr)



↳ Ici, une illustration signée Faria (ca 1895), déposée sur MédiHal par Éliane Daphy, Institut interdisciplinaire d'anthropologie du contemporain.

## ● ● ● LA CONSTRUCTION COLLABORATIVE DE L'ARCHIVE OUVERTE

**Que ce soit dans la saisie ou la curation de données, une multiplicité d'acteurs interviennent à différents niveaux.**

- **Partage du dépôt** : plusieurs utilisateurs ont les droits d'édition sur un dépôt. Ainsi, les coauteurs, par exemple, peuvent mettre à jour les données. Cela permet aussi d'organiser le dépôt : saisie des références bibliographiques par un professionnel de la documentation et ajout par l'auteur lui-même du fichier contenant le texte intégral.
- **Des rôles avec des droits** : le référent structure peut modifier/compléter les données d'un dépôt lié à son unité de recherche ; le gestionnaire de collection valorise les publications qui correspondent à un critère (laboratoire, conférence, projet européen) ; le valideur technique vérifie les dépôts avant leur mise en ligne ; l'administrateur d'un portail valorise les publications de son établissement, met à jour les dépôts et les référentiels.
- **Des référentiels en accès ouvert<sup>1</sup>** : auteurs, structures de recherche, revues, disciplines, projets ANR et projets européens.
- **CasuHAL, le club d'utilisateurs** : en 2016, un groupe d'utilisateurs, pour l'essentiel des administrateurs de portails, ont formé une association, CasuHAL. Liste de discussion, réunions régulières en visioconférence, wiki<sup>2</sup> sont mis à leur disposition par le CCSD pour le partage d'expérience et l'échange de bonnes pratiques. La qualité des échanges et la confiance entre le CCSD et ses utilisateurs sont les garanties des meilleurs choix de développement.

[1] [aurehal.archives-ouvertes.fr](http://aurehal.archives-ouvertes.fr)

[2] [wiki.ccsd.cnrs.fr/wikis/hal/index.php/Accueil](http://wiki.ccsd.cnrs.fr/wikis/hal/index.php/Accueil)

Portail de publication multilingue en sciences humaines et sociales créé par le Centre pour l'édition électronique ouverte, OpenEdition se décline en quatre plateformes complémentaires et un laboratoire de R&D. Il mise sur la formation à de nouveaux usages éditoriaux, sur l'internationalisation, sur un modèle économique durable et sur des innovations technologiques régulières.

# OpenEdition : comprendre le monde en accès ouvert

Le 1<sup>er</sup> janvier 2016, sur proposition de l'Alliance Athena, le ministère de la Recherche a attribué à OpenEdition le statut d'infrastructure de recherche nationale. OpenEdition devient donc une infrastructure de recherche au même titre qu'Huma-Num, HAL, et que les accélérateurs synchrotron (Soleil, ESRF), les lasers (Luli), les centres de diffusion neutronique (LLB, ILL), les avions de recherche instrumentés pour la recherche (Safire), la flotte océanographique de recherche (FOF) et la station polaire internationale Concordia. Ainsi les services d'OpenEdition s'inscrivent-ils désormais durable-

- les revues sur Revues.org (450 revues) ;
- les carnets de recherche sur Hypothèses (1 700 carnets) ;
- les événements sur Calenda (34 000 événements).

Les quatre plateformes ont reçu près de 70 millions de visites en 2016 et représentent plus de 500 000 documents publiés. L'ensemble des plateformes connaît une croissance annuelle à deux chiffres.

En devenant une infrastructure de recherche, OpenEdition a stabilisé son rôle national et confirmé son ambition internationale. OpenEdition est déjà fortement implanté en Italie (OpenEdition Italia, à Turin) ainsi qu'au Portugal (LusOpenEdition, à Lisbonne), et a noué de fructueux partenariats en Allemagne (De. hypotheses.org avec la Max Weber Stiftung, Bonn), Espagne (Es.hypotheses.org avec l'Uned, Madrid) et aux Pays-Bas (forte implication dans le Directory of Open Access Books, La Haye). La diversification de la provenance géographique et linguistique des contenus se confirme d'année en année.

OpenEdition est par ailleurs porteur de deux projets européens (H2020) et d'une candidature à la reconnaissance en tant qu'infrastructure européenne (feuille de route ESFRI).

## L'accès ouvert : une opportunité de financement pour l'édition scientifique, et pas seulement un (excellent) modèle de diffusion.

ment dans le paysage national de la recherche. Il s'agit d'une reconnaissance du rôle joué par OpenEdition dans la diffusion des résultats de la recherche en sciences humaines et sociales. C'est le résultat du travail collectif réalisé depuis dix-huit ans au service de la diffusion des connaissances.

### QUATRE PLATEFORMES EN RAPIDE CROISSANCE

OpenEdition met en œuvre quatre plateformes dédiées à quatre objets éditoriaux distincts :

- les livres sont publiés sur OpenEdition Books (3 500 livres) ;

### UN MODÈLE ÉCONOMIQUE EN APPUI À L'ÉDITION SCIENTIFIQUE

La diffusion en accès ouvert est privilégiée par OpenEdition, qui laisse cependant le choix du modèle de diffusion aux revues et éditeurs qui lui confient des contenus à diffuser. Afin de conforter la diffusion en accès ouvert, OpenEdition propose depuis quelques années un modèle de diffusion innovant associant diffusion en *open access* et offre commerciale. Ce modèle de diffusion, le *freemium*, conjugue une diffusion en accès ouvert (la partie free du nom) et une commercialisation de services (la partie premium du nom). Les revues et les livres diffusés en *open access freemium* sont donc accessibles en accès ouvert dans leur version HTML, et des services complémentaires sont vendus aux bibliothèques : accès aux versions PDF et Epub, statistiques, webservices, formats de métadonnées « métiers », etc. sont accessibles seulement aux



OpenEdition-Lisa George/Flickr (CC BY-NC-ND 2.0)

établissements qui soutiennent le modèle. Les deux tiers des revenus ainsi collectés sont versés aux revues et aux éditeurs. Il s'agit de faire en sorte que l'accès ouvert devienne une opportunité de financement pour l'édition scientifique, et pas seulement un (excellent) modèle de diffusion. C'est là une condition nécessaire à l'inscription dans la durée d'une édition de qualité. Nous sommes loin, ici, du modèle auteur-payeur qui se propage dans de nombreuses disciplines, non sans poser des problèmes économiques et éthiques.

## R&D ET SOLUTIONS D'AVENIR

Enfin, le projet s'est doté d'une activité de recherche et développement, OpenEdition Lab. Il a démarré grâce à des financements de Google en 2010. Après avoir initié des projets autour de la fouille de texte (qui ont débouché sur le logiciel d'annotation bibliographique Bilbo), OpenEdition Lab a diversifié son activité. Désormais, des programmes de recherche portent sur l'innovation éditoriale, l'*open peer review*, la sociologie des usages de lecture et de citation numérique, la recommandation, la fouille de textes, etc. L'activité du Lab fonctionne toujours dans une logique partenariale.

Début 2018, Revues.org deviendra OpenEdition Journals. Cette mue consistera d'abord en un changement d'URL, qui permettra de regrouper sous un même domaine les livres et les revues. Après ce déplacement vers un nouveau nom, la plateforme fera annuellement l'objet de profondes améliorations, qui seront toutes tournées vers une plus grande accessibilité, une amélioration des services proposés aux éditeurs, aux bibliothèques, aux lecteurs. OpenEdition gravera ces lettres sur la Toile comme une ambition ultime : comprendre le monde en accès ouvert.

### MARIN DACOS

Directeur d'OpenEdition  
 marin.dacos@openedition.org  
 Twitter : @marindacos



**POUR EN SAVOIR PLUS**

Sur OpenEdition, [openedition.org](http://openedition.org)

Twitter [@openeditionsays](https://twitter.com/openeditionsays)

Sur OpenEdition Lab, [lab.hypotheses.org](http://lab.hypotheses.org)

Sur l'offre *freemium* et comment soutenir l'accès ouvert, [freemium@openedition.org](http://freemium@openedition.org)

Pour recevoir des supports de communication, [contact@openedition.org](mailto:contact@openedition.org)

Lire : Peter Suber. *Qu'est-ce que l'accès ouvert ?*  
 Marseille : OpenEdition Press.

<http://books.openedition.org/oe/1600>

ISBN: 9782821869806.

DOI: 10.4000/books.oe/1600



➤ OpenEdition au congrès de l'Ifla en 2014.



➤ Assemblée générale des revues d'OpenEdition (Revue.org), le 31 mars 2014.  
 De gauche à droite : Patrice Bellot, Marin Dacos.

OpenEdition est une initiative développée par le Centre pour l'édition électronique ouverte (Cléo), une unité mixte de services implantée à Marseille et à Paris, sous la tutelle du CNRS, de l'Université d'Aix-Marseille (AMU), de l'École des hautes études en sciences sociales (EHESS) et de l'Université d'Avignon. L'infrastructure bénéficie de soutiens du Ministère de la Recherche et est lauréate d'un équipement d'excellence (2012-2019).

Les établissements publics de l'ESR trouvent en Numédif un support pour la circulation de leurs écrits scientifiques, qui développe les outils permettant la production numérique de contenus structurés interopérables et un appareil spécifique de diffusion-distribution.

# Numédif : au service de l'édition numérique

Portée conjointement par la Fondation Maison des sciences de l'homme (FMSH) et la Maison de la recherche en sciences humaines (MRSH) de Caen,

l'infrastructure Numédif<sup>1</sup> est conçue à l'usage des éditeurs et au service de l'activité éditoriale des établissements publics d'enseignement supérieur et de recherche (ESR) dans le cadre de leur mission de diffusion de la connaissance. Elle a pour fonctions principales de fournir un ensemble d'outils et de méthodes permettant la constitution de fonds éditoriaux et de contenus scientifiques numériques structurés

et d'optimiser leur diffusion numérique ou imprimée, au service de la diffusion des résultats et des données de la recherche et en pleine connaissance des impératifs de l'idéal d'*open access* et des enjeux économiques qui les sous-tendent. Elle doit permettre d'améliorer la visibilité de l'édition scientifique publique française à l'international.

## EN AMONT ET EN AVAL DU TRAVAIL ÉDITORIAL

Autour de l'activité éditoriale, Numédif s'attache :

- en amont, au développement et à la diffusion d'outils et de méthodes permettant la constitution des fonds de contenus édités, normalisés, interopérables, bâtis sur le modèle du *single source publishing*, dont un des apports principaux est de rationaliser et de factoriser le travail éditorial tout en favorisant les stratégies de diffusion multisupport et en assurant une pérennisation des contenus et une haute qualité des métadonnées associées ;
- en aval, à la construction d'un appareil spécifique de diffusion-distribution, propre à l'édition savante, articulé avec les plateformes de diffusion de contenus en ligne et offrant toutes les échelles de diffusion : réseau des bibliothèques vers un public spécialisé, mises en place massives pour un public



Grant Hutchinson/Flickr (CC BY-NC-ND 2.0)

large, diffusion en ligne, diffusion de formes numériques détachables...

Le respect rigoureux de standards (XML-TEI, Onix) est la garantie de forts potentiels d'articulation avec les autres infrastructures ou avec les opérateurs du domaine, soit par l'alimentation directe en contenus, pour un coût résiduel et sans perte de qualité éditoriale, comme dans le cas d'OpenEdition, soit par le partage ou la fourniture de métadonnées.

## UNE APPROCHE RÉSOLUMENT « MÉTIERS »

Développée avec une logique « métiers », dans le droit fil des réflexions menées au sein de la BSN, la mise en œuvre de Numédif résulte d'un travail de collaboration entre deux structures préexistantes :

- **Métopes** (Méthodes et outils pour l'édition structurée), développé à la MRSH de Caen et porté par son pôle Document numérique, qui maintient et diffuse le standard de structuration des contenus sur la base des recommandations de la *Text encoding initiative*, définit les principes du flux éditorial, assure les développements des environnements de travail et des outils de transformation (vers OpenEdition-TEI, Cairn-Érudit, InDesign pour des éditions imprimées ou PDF...), ainsi que leur diffusion et leur déploiement auprès des structures éditoriales publiques, dont près de 50 ont été formées à ce jour ;

- **FMSH-Diffusion**, au travers de ses différents services : CID, pour la promotion et la vente d'ouvrages au sein du réseau de la bibliothèques et leur diffusion auprès des bibliothèques ; SLU pour les ouvrages à rotation lente ; LCDPU pour la vente au comptoir et en ligne aux particuliers sur la base d'un catalogue, formalisé en Onix (Online Information Exchange), riche de plus de 23 500 références.

L'activité éditoriale touche les champs technique, économique et, nécessairement, juridique.

## S'EXTRAIRE DES LOGIQUES DE RÉTROCONVERSION

Numédif intervient donc de plusieurs manières :

- la prescription de normes et standards dans le domaine de l'édition et de la diffusion numérique ;
  - le développement d'environnements de saisie et de traitement de flux structurés normés, le développement d'automates de composition de flux éditorialisés et la formation à leur usage ;
  - la définition et la mise en œuvre de dispositifs contractuels ;
  - la construction de partenariats modèles en coédition et en diffusion-distribution ;
  - la veille technique et informationnelle et l'animation de réseaux de professionnels.
- Plus immédiatement, Numédif se veut une boîte à outils pour la pleine intégration des logiques numériques à la construction et à l'exploitation des fonds éditoriaux – des outils qui permettent de s'extraire des logiques de rétroconversion, coûteuses en temps, en moyens et... en qualité éditoriale.

**DOMINIQUE ROUX**

Responsable du projet Numédif-Métopes  
dominique.roux@unicaen.fr

[1] Numédif est l'acronyme de Numérique pour l'édition et la diffusion de la production scientifique.

# Huma-Num : l'humain en majesté numérique

La TGIR Huma-Num développe des réponses aux besoins nouveaux, apparus avec l'irruption du numérique, des différents acteurs de la recherche en sciences humaines et sociales en France. Elle offre un appui aux dispositifs français et, s'appuyant sur l'expertise réunie dans les consortiums, coordonne la production de corpus de sources. Survol de sa stratégie et de son fonctionnement.

Dans le domaine des sciences humaines et sociales (SHS), le tournant numérique de la recherche a provoqué une évolution considérable des pratiques de recherche. Aujourd'hui, les chercheurs produisent et utilisent collectivement d'importants volumes de données, qui nécessitent des outils spécialisés pour y accéder, les manipuler et les visualiser et d'importants moyens technologiques. La maîtrise de ce nouvel environnement, composé de sources hétérogènes – des bibliothèques numériques aux données numériques récentes, exclusivement traitées et diffusées en ligne – appelle des compétences en constante évolution, aussi réflexives que les changements sont rapides. Un tel écosystème ne peut plus être réalisé avec les moyens dont disposent les chercheurs, ni avec ceux d'un laboratoire, ni même avec ceux d'un programme de recherche collectif. Une véritable infrastructure, au sens élargi du terme, peut fournir et mutualiser ces différents services. La très grande infrastructure de recherche nationale (TGIR) Huma-Num est tournée vers les services pour la production et la réutilisation de données numériques, mais aussi vers l'accompagnement d'outils émergents et de bonnes pratiques issues des communautés scientifiques des SHS elles-mêmes. Elle est associée étroitement à la construction d'infrastructures européennes et internationales.

## DES OUTILS POUR TOUT LE CYCLE DE VIE DES DONNÉES

Les services d'Huma-Num doivent fournir des outils tout au long du cycle de vie des données : de la production de données brutes jusqu'à leur préservation après traitement. Pour proposer des ressources bien dimensionnées et adaptées, Huma-Num n'a pas créé son propre centre de données, mais s'appuie sur les moyens de centres de calcul nationaux (centre de calcul de l'IN2P3-CNRS, CINES).

Un objectif sous-jacent consiste à amener les chercheurs à s'approprier les méthodes et services numériques au cœur des processus de la recherche. Les technologies d'interopérabilité, en particulier celles proposées par le Web sémantique, sont au centre de cette chaîne de traitement, permettant des échanges fluides entre les différentes « briques » qui la composent.



Huma-Num développe à la fois des services propres (**Nakala, Nakalona, Isidore**), tous interopérables, munis d'API et utilisant le Web sémantique, et met aussi à disposition des services forgés par les communautés SHS, principalement issus des consortiums et Maisons des sciences de l'homme. Ainsi, Nakala (qui veut dire « copie conforme » en langue swahili) permet de déposer des données, de les documenter à l'aide de métadonnées. Nakalona (qui veut dire « voir Nakala » en swahili) utilise le logiciel *open source* Omeka pour permettre à un chercheur d'éditorialiser les corpus de données sous la forme de bibliothèque numérique. L'ensemble de ces données, issues des corpus en SHS, qu'elles utilisent nos services ou qu'elles viennent d'ailleurs, peuvent être signalées dans notre moteur de recherche sémantique Isidore.

Un point important dans le fonctionnement de cet écosystème complet de services pour les données numériques en SHS est la sensibilisation à la pérenni-

↖ **Cologne, grand pont sur le Rhin** [fin du XIX<sup>e</sup> siècle]. Photographie extraite du fonds du lycée Colbert (Paris), mise en ligne sur le site du labex Ehne, site hébergé par Huma-Num.

[1] European Research Infrastructure Consortium.

[2] Common Language Resources and Technology Infrastructure, [www.clarin.eu](http://www.clarin.eu)

[3] Digital Research Infrastructure for the Arts and Humanities, [www.dariah.eu](http://www.dariah.eu)

sation, au partage des données et à la normalisation par une concertation (disciplinaire et interdisciplinaire) entre chercheurs. Cette sensibilisation devrait être systématique lors d'une demande d'utilisation.

## UNE ACTION EN LIEN AVEC LES CONSORTIUMS

Les consortiums réunissent des acteurs issus de différents projets de recherche autour de thématiques et d'objets communs coédités. Ils favorisent l'appropriation des dispositifs numériques par ces communautés et créent ainsi des synergies en lien avec les services d'Huma-Num. Leurs activités vont de la création d'applications web à la définition de schémas de métadonnées cogérés, en passant par la production de guides de bonnes pratiques et des formations spécifiques.

Actuellement, Huma-Num soutient les activités de dix consortiums, principalement disciplinaires. Il s'agit de : **Cahier** (Corpus d'auteurs pour les humanités : vise à fédérer les différentes initiatives françaises autour de corpus d'auteurs, qu'ils relèvent de la littérature, de la philosophie ou d'une thématique liée à une école ou à une pratique) ; **Archives des ethnologues** (conservation et pérennisation des données collectées sur le terrain ainsi que de tous documents pouvant aider à la contextualisation de ces données ou à une meilleure compréhension de l'élaboration des travaux anthropologiques) ; **Musica** (qui coordonne la réflexion épistémologique des chercheurs, ainsi que le développement de nouveaux outils de représentation et d'analyse de la musique) ; **Corli** (Corpus, langues et interaction, né du rapprochement de deux consortiums de linguistique, Corpus écrits et Corpus oraux et multimodaux) ; **3D** (qui se propose de fédérer les équipes de recherche travaillant sur l'usage des méthodologies et technologies 3D dans l'étude des sociétés humaines) ; **Masa** (Mémoires des archéologues et des sites archéologiques, qui réunit Musée d'archéologie nationale des acteurs nationaux déjà rassemblés dans le réseau des Maisons des sciences de l'homme traitant d'archéologie) ; **Cosme** (Consortium Sources médiévales, qui réunit des scientifiques de toutes les disciplines autour de

la recherche sur le passé médiéval) ; **Imageo** (qui numérise, archive et valorise les cartes et photographies par différents laboratoires du CNRS et la cartothèque de Bordeaux-Montaigne) ; **Archipolis** (Archives des sciences sociales du politique) ; **ArcMC** (Archives des mondes contemporains, se consacrant à la problématique de la fragilité et dispersion des archives des XIX<sup>e</sup> et XX<sup>e</sup> siècles – microsillons, photographies, disquettes...).

## VERS UN ENSEMBLE COHÉRENT, NATIONAL ET EUROPÉEN

Les activités internationales d'Huma-Num sont de natures différentes. Elles visent essentiellement à valoriser les dispositifs français d'infrastructures et de recherche et de participer aux différents réseaux des humanités numériques de par le monde. Concrètement et sur le long terme, Huma-Num porte la participation française à des Eric<sup>1</sup>, disciplinaires, comme **Clarin**<sup>2</sup>, ou incluant le domaine culturel, comme **Dariah**<sup>3</sup>. Les Eric ne financent pas les projets qui les composent ; ceux-ci et l'Eric sont financés, pour toute leur durée, par chacun des États membres qui y participent. L'originalité du budget de ces structures consiste dans le fait qu'il est en majeure partie constitué de contributions « en nature », afin de ne pas dupliquer à l'échelle européenne des services qui existeraient déjà aux niveaux nationaux. Le rôle d'Huma-Num est donc d'identifier les services et expertises nationaux – en particulier ceux en provenance des consortiums – et d'en faire un ensemble cohérent, qui intégrera en la complétant l'offre européenne. Ce rôle englobe la participation d'Huma-Num, de façon plus directe, à des projets à plus court terme, qui soutiennent la construction de ces grandes infrastructures du numérique et qui sont, eux, financés par la Commission européenne par le programme-cadre Horizon 2020.

Enfin, Huma-Num maintient des liens importants avec les mondes francophones (le Québec) et hispanophones (l'Amérique du Sud). Par exemple, depuis 2016, Huma-Num travaille en coopération avec l'Université de Montréal et le consortium Érudit autour du développement de la cyberinfrastructure COSHS, en partageant de l'information sur la façon de développer des services, d',k,animer des communautés en humanités numériques, et en réalisant des démonstrateurs web avec Isidore, le moteur de recherche.

À travers ces différentes composantes, Huma-Num s'emploie à développer des réponses aux besoins nouveaux, provoqués par l'utilisation du numérique, des différents acteurs de la recherche. Le fil rouge qui les relie est de rendre possibles de nouvelles recherches dans le domaine des sciences humaines et sociales.

**STÉPHANE POUYLLAU**

Directeur technique d'Huma-Num  
[stephane.pouyllau@cnrs.fr](mailto:stephane.pouyllau@cnrs.fr)

## LES 3 AXES D'HUMA-NUM

- **les services technologiques** pour l'outillage des données de la recherche (traitement, conservation pérenne, accès et l'interopérabilité)
- **la création et le soutien de consortiums**, disciplinaires ou non, qui nourrissent Huma-Num en guides de bonnes pratiques, formations, plateformes de données ancrées dans les Maisons des sciences de l'homme
- **l'ouverture à l'international** : participation à des projets d'infrastructures européennes (en particulier d'expertise), en complément d'échanges bilatéraux plus ciblés portant sur le déploiement des services d'Huma-Num.

Depuis 2013, la politique nationale d'acquisition de documentation confiée à Istex a mis à la disposition de l'ESR un ample gisement documentaire. Les outils pour y accéder et l'exploiter s'intègrent directement dans les systèmes numériques des établissements. Visite dans les arcanes de l'API Istex, sur laquelle reposent ces interfaces et fonctions adaptées aux utilisateurs.

# L'API Istex : le sésame pour accéder aux ressources acquises



Actualité/Flickr (CC BY 2.0)

← Espace commun de démonstration HAL-Istex-Persée lors du Salon du livre 2015.

À la question « Comment accède-t-on aux ressources Istex? », la réponse la plus courante est : « par la plateforme Istex », ou encore « par l'API Istex », cette dernière étant disponible depuis de nombreux mois à l'adresse [api.istex.fr](http://api.istex.fr). Mais cette réponse est incomplète, voire inadaptée pour qui souhaite simplement consulter ou imprimer un document scientifique. Dans les faits, le personnel d'un établissement peut accéder aux ressources Istex de multiples manières – par son portail d'établissement, ou au moyen d'une extension pour Firefox, par exemple –, sans jamais interroger directement l'API elle-même. Dès lors, on peut reformuler la question autrement : comment ceci est-il rendu possible? En quoi une API permet et facilite la diffusion de ressources? Et quels sont les divers modes d'accès?

## UNE BRIQUE SIMPLE, STANDARDISÉE ET OUVERTE

En français, API (*Application Programming Interface*) signifie « interface de programmation applicative » : un système ou un ensemble de composants per-

mettant d'interagir avec un service. Dans notre cas, le service est la plateforme Istex, qui propose un ensemble de fonctionnalités pour accéder aux ressources acquises auprès des éditeurs. Se voulant universel, ce service est disponible sur Internet à travers le protocole standard HTTP.

Comme de nombreux services « orientés ressources », l'API Istex se base sur une architecture REST<sup>1</sup> et utilise un format simple et léger : JSON<sup>2</sup>.

Ce service web est donc une « brique » simple, standardisée et ouverte<sup>3</sup>, qui va servir de base à tout site web ou à tout logiciel qui souhaite donner à l'utilisateur final un accès aux ressources Istex.

L'API Istex gère un grand nombre de fonctionnalités courantes : contrôle d'accès, recherche de documents avec possibilité de filtres, tris et facettes, choix du format de la réponse, récupération des documents dans leur forme d'origine, normalisée ou enrichie<sup>4</sup>, export d'un ensemble de résultats, etc. Toutes ces fonctionnalités peuvent donc être proposées dans les interfaces de l'utilisateur final.

[1] REST est une architecture logicielle souvent utilisée dans les API web pour sa simplicité et son adéquation avec le protocole HTTP.

[2] JSON (JavaScript Object Notation) est un format léger, permettant de stocker des objets typés (booléens, textes, tableaux...).

[3] Sa documentation complète est disponible sur [api.istex.fr/documentation](http://api.istex.fr/documentation)

[4] Références bibliographiques, catégorisations, etc.

## ISTEX DANS LES ÉTABLISSEMENTS

À ce jour, plusieurs établissements pilotes proposent déjà l'accès aux ressources Istex, notamment les universités de Lorraine, de Rennes et de Saint-Étienne. Les personnels de ces universités peuvent utiliser la partie documentaire de leurs portails pour rechercher parmi les ressources Istex et accéder aux documents en texte intégral. Ceci a été rendu possible par une intégration de l'API, complètement transparente pour l'utilisateur.

Ces intégrations ont été réalisées à l'aide de *plugins* génériques, notamment pour les logiciels Drupal et uPortal. Ces briques logicielles étant disponibles librement, elles peuvent être réutilisées par d'autres établissements qui disposeraient des mêmes types de portail.

Quant aux établissements dont les sites ne s'appuient sur aucun logiciel ou espace numérique de travail particulier, ils peuvent intégrer l'accès aux ressources Istex grâce à des *widgets* web génériques et paramétrables – une opération qu'un webmestre réalise facilement en insérant quelques lignes de code informatique.

À plus long terme, les établissements ayant souscrit à un outil de découverte (EBSCO, OCLC...) pourront sélectionner les bouquets Istex et afficher un lien vers le plein texte hébergé sur la plateforme Istex.

[5] Cette déclaration se base sur les fichiers KBART fournis par la base Bacon.

## TROIS ÉQUIPES POUR UNE PLATEFORME

Le développement de la plateforme est réalisé à l'Inist-CNRS de Vandoeuvre-lès-Nancy, où le travail est mené par trois équipes aux objectifs complémentaires et coordonnés.

- **ISTEX-DATA** : vérification et curation des données livrées par les éditeurs en amont de la mise en ligne. Son expertise des formats lui permet également de mettre au point les transformations XSLT vers les formats pivots MODS et TEI. L'équipe est également en train de bâtir une « chaîne d'OCRisation » (reconnaissance optique de caractères), qui permettra d'ajouter ou d'améliorer le plein texte « brut » des documents, matière indispensable pour les travaux de text & data mining.

- **ISTEX-API** : mise à disposition des documents. Ce travail s'articule autour de trois aspects : une « chaîne

d'ingestion » qui prépare, reformate et indexe les données, une API qui expose les documents sur Internet, et des outils qui facilitent l'exploitation de l'API (l'extension pour Firefox, par exemple).

- **ISTEX-RD** : recherche et développement pour l'amélioration des données. L'objectif est de mettre au point des outils permettant d'enrichir les données initiales en repérant ou extrayant de nouvelles informations. Les données produites (références bibliographiques ou entités nommées, par exemple) sont ensuite reversées dans l'API et mises à disposition de la communauté.

Ces trois équipes travaillent en étroite collaboration, en s'appuyant sur la méthode agile Scrum. Elles ajustent en permanence leurs objectifs en fonction des retours venant des utilisateurs.

## LE PLEIN TEXTE À PORTÉE DE MAIN

Deux fonctionnalités de l'API Istex combinées entre elles facilitent l'accès aux documents sans bouleverser les habitudes des usagers.

- **Identification** : depuis peu, l'accès à l'API Istex peut se faire au moyen de l'authentification délivrée par le fournisseur d'identités de son établissement (par exemple, Janus, pour le CNRS). En effet, l'API s'appuyant sur la fédération d'identités Éducation-Recherche de Renater, l'ensemble de la communauté scientifique bénéficie dès à présent de cette facilité.

- **Recherche plein texte** : une autre fonctionnalité intéressante de l'API est son résolveur de liens, compatible avec la norme OpenURL. À partir de métadonnées simples (titre, auteur...) ou d'identifiants standards (DOI, PMID...), elle permet de savoir si un document est présent dans la base de documents Istex et, si oui, de faire un rebond vers le plein texte.

Comment ces fonctionnalités sont-elles utilisées par le grand public ? En premier lieu, par les extensions pour les navigateurs web Chrome et Firefox. Disponibles sur GitHub et sur les boutiques d'applications, ces extensions analysent les pages web visitées en recherchant des identifiants standards DOI, PMID et PII. Une fois la disponibilité dans la plateforme Istex vérifiée, un lien vers le plein texte PDF s'affiche. L'authentification par fédération d'identités est automatiquement demandée en cas de besoin. L'installation de ces extensions se fait en quelques clics, ce qui nous permet d'élargir le spectre d'utilisateurs Istex.

De manière similaire, des liens vers la plateforme Istex sont établis depuis le moteur de recherche Google Scholar. Dès lors que l'utilisateur a sélectionné Istex dans la partie « liens vers les bibliothèques » de ses paramètres, des liens OpenURL vers le résolveur de l'API Istex sont automatiquement affichés pour tous les documents ayant été déclarés à Google<sup>5</sup> comme faisant partie du « bouquet Istex ». Comme pour les extensions, si la résolution aboutit, l'utilisateur est redirigé vers le plein texte PDF, avec les mécanismes de contrôle d'accès présentés précédemment.

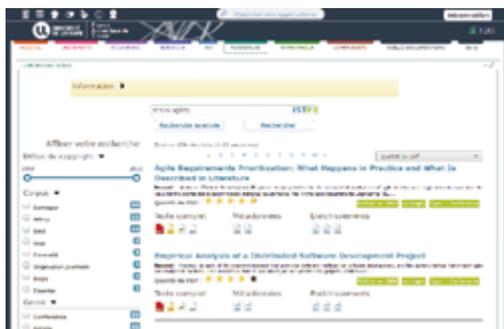
### TEXT & DATA MINING

Une spécificité d'Istex en tant qu'archive documentaire est la possibilité d'utiliser le texte intégral comme matière première pour des travaux de recherche, en lui appliquant par exemple des techniques de *text & data mining*. Il a fallu pour cela que l'API soit facilement « moissonnable » par les chercheurs. Plusieurs actions ont été menées en ce sens. La première a été de mettre à la disposition de la communauté un outil de moissonnage automatique. Écrit en langage NodeJS, il est disponible sur GitHub

et s'utilise en interface « ligne de commande ». Ses nombreuses options permettent de cibler un sous-corpus de la même manière qu'on le ferait en interrogeant l'API – filtrer sur l'ensemble des champs, trier, choisir les formats de sortie – ou encore de limiter le nombre de documents souhaités. Ce script fonctionne sur l'ensemble des systèmes d'exploitation et a déjà été utilisé pour extraire plusieurs millions de documents.

Pour ceux qui ne sont pas familiers avec la ligne de commande et qui souhaitent récupérer des volumétries inférieures à 10 000 documents, l'API Istex propose également une fonctionnalité d'export au format Zip. Tout comme avec l'outil précédent, l'utilisateur peut spécifier finement sa requête et choisir les formats souhaités. Les utilisateurs les plus techniques pourront, quant à eux, réaliser eux-mêmes leurs scripts ou programmes de moissonnage dans leur langage favori. Un partage avec l'ensemble de la communauté est d'ailleurs vivement apprécié.

**CLAUDE NIEDERLENDER**  
Responsable du projet Istex-API  
Inist-CNRS  
claude.niederlender@inist.fr



## TUTORIELS ET CODE SOURCE

Une série de tutoriels à destination des webmasters ou des utilisateurs les plus techniques a été réalisée par le service formation de l'Inist. Interactifs, ils permettent d'apprendre à l'aide de cours et d'exercices, de manière ludique, à interroger l'API. Ils sont disponibles sur le site de l'Inist.

Dans une optique de partage et d'émulation de la communauté, le code source d'un certain nombre d'outils spécifiques d'Istex a été mis en ligne sur le site GitHub à l'adresse [github.com/istex](https://github.com/istex). On y retrouve notamment les plugins pour Drupal et uPortal, les extensions pour navigateurs, les scripts de moissonnage de l'API, les widgets pour sites web ou le code source d'un mini-site de démonstration. Toute contribution est bienvenue.



# À sept mois du clap de fin, Istex approche les 20 millions d'articles

**La politique d'acquisition de ressources documentaires à l'échelon national dans le cadre du projet Istex, porté par le CNRS, Couperin, l'Université de Lorraine et l'Abes, a donné lieu à la constitution d'un fonds documentaire unique, mis à disposition à la fois sur les plateformes des éditeurs et via l'API Istex.**

Le programme d'achats de licences nationales a débuté par des bases de données qui recèlent des documents à valeur patrimoniale. Les documents concernés sont de natures hétérogènes. Ce sont d'abord les grammaires et les dictionnaires de Classiques Garnier Numérique, qui permettent un voyage dans l'histoire de notre langue. Ce sont aussi les deux bases complémentaires, Early English Books Online (EEBO) et Eighteen Century Collection Online (ECCO), qui offrent une vue globale sur les publications de langue anglaise – mais pas seulement – sur plus de trois siècles. Les collections rétrospectives de revues scientifiques : un autre grand axe que le projet Istex poursuit en privilégiant la profondeur de collection et l'adossement aux abonnements actuels des bibliothèques. Par la signature de douze contrats avec des éditeurs de stature internationale comme Springer, Wiley, Cambridge University Press, Institute of Physics ou encore Emerald, tous les établissements français ayant des missions d'enseignement supérieur et de recherche accèdent aux

archives de plus de 8000 revues dans tous les domaines disciplinaires : médecine, chimie, physique, économie, mais aussi sociologie, histoire, droit, linguistique, etc. La couverture chronologique a été étendue le plus possible pour se rapprocher des publications courantes.

Mais Istex, c'est aussi des *e-books* : les achats réalisés à ce jour (10000) portent sur des titres en sciences, en chimie, en science politique et en droit international.

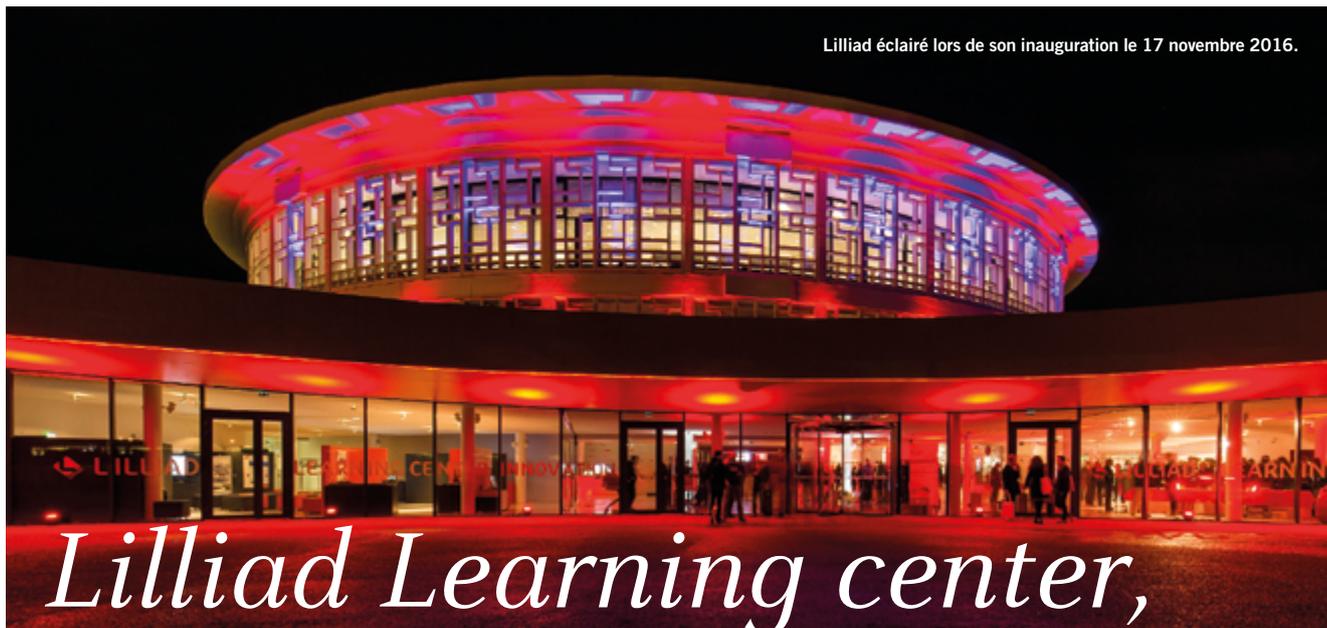
Les mois qui restent jusqu'à la fin du projet – prévue au 31 août 2017 – seront employés pour parfaire cet ensemble, notamment en augmentant le contenu francophone et en renforçant le corpus en sciences humaines et sociales. Les derniers achats seront réalisés au début de l'été.

**CAROLE MELZAC**  
Responsable du service Achats  
documentation électronique, Abes  
carole.melzac@abes.fr



### POUR EN SAVOIR PLUS

**Vous pouvez retrouver toutes les informations sur chaque ressource sur le site [LicencesNationales.fr](https://www.licencesnationales.fr)**



Lilliad éclairé lors de son inauguration le 17 novembre 2016.

© Atmosphère Photo

# Lilliad Learning center, au cœur de la Cité scientifique

Lilliad a ouvert au public en septembre dernier, à l'issue d'un projet de près de dix ans et de deux années de travaux. Visite des espaces et des services de cette bibliothèque universitaire d'un nouveau genre.

Lilliad a pour noyau l'ancienne bibliothèque de l'Université des sciences et technologies de Lille (USTL). Au sein du campus de la Cité scientifique, le bâtiment a dès l'origine constitué le symbole de l'université, son cœur. Bâtiment de type « Pailleron », rond, ouvert et sans murs porteurs, mais reposant sur une forêt de piliers métalliques, la bibliothèque universitaire se composait de trois niveaux – sous-sol partiellement enterré, rez-de-chaussée, premier étage. L'identité de la structure reposait sur deux caractéristiques qui formaient la signature architecturale du lieu : une coupole en pavés de verre typique des années 1960 d'une part, des claustras décoratifs – moucharabieh – venant épouser la peau extérieure du bâtiment de l'autre. L'articulation de la bibliothèque avec les autres bâtiments du campus constituait également une caractéristique remarquable : une bibliothèque ronde, au centre du campus, et un ensemble de bâtiments d'enseignement et de recherche rayonnant autour d'elle.

## UNE FILLE DES TRENTE GLORIEUSES

Construite au milieu des années 1960, et bien que très moderne lors de son ouverture, la bibliothèque était fille de son temps : elle était caractérisée par des espaces de travail ouverts, mais proportionnés pour un public universitaire alors peu nombreux.

À partir des années 80, avec la démocratisation de l'accès à l'université et le fort enrichissement des collections, la bibliothèque s'est progressivement révélée moins adaptée aux usages comme aux attentes. Cette inadaptation s'est doublée d'un vieillissement du bâtiment, accompagné d'un décrochage rapide au regard des normes de sécurité en matière de construction.

Au début des années 1990, l'université travaille à un premier projet de rénovation-extension du bâtiment : il s'agissait alors d'étendre les surfaces pour faire face à l'afflux étudiant, de moderniser les usages – mise à niveau des réseaux par exemple – mais pas d'en élargir les missions.

Ce projet de rénovation-extension ira jusqu'au concours, mais ne sera *in fine* pas mené à terme. En 2006, après deux tentatives et une forte augmentation des coûts de construction, les financements initialement obtenus dans le cadre du CPER<sup>1</sup> d'alors ne sont plus suffisants : il faut initier un nouveau projet.

## DE LA BU AU LEARNING CENTER

Le nouveau projet vise à réaliser plus qu'une « simple » bibliothèque, aussi bien adaptée soit-elle aux nouveaux usages. Il s'agit désormais d'élargir les missions et le périmètre de l'ancienne bibliothèque universitaire, dans ses espaces comme dans ses services. Pour ce faire, Lilliad comporte trois espaces

## LILLIAD EN CHIFFRES

- 12 600 m<sup>2</sup> de superficie totale
- 50 salles de travail (4 à 20 personnes)
- 2 salles de conférence (292 et 120 places)
- 1 café internet (90 places)
- Wi-Fi dans tout le bâtiment
- 2 000 prises électriques
- 13 km de stockage de livres et périodiques

principaux - une bibliothèque, un complexe événementiel et un espace de valorisation de la recherche partenariale, l'Xperium. Bien qu'ils puissent fonctionner indépendamment, ces espaces sont traités non pas de manière juxtaposée mais intégrée (accueil, pilotage, informatique, administration, communication, logistique...). En effet, ces trois espaces participent de la même finalité : la valorisation et la diffusion des savoirs.

Cet élargissement des missions se traduit par une diversification des usagers qui va bien au-delà du public universitaire traditionnel. Le public lycéen tout d'abord, dans un double souci d'attractivité des futurs étudiants vers l'université et de transition douce entre le lycée et celle-ci. Les entreprises et les acteurs régionaux de l'innovation ensuite, à qui il s'agit d'offrir un cadre de rencontre privilégié avec le monde universitaire. Le grand

public enfin, dans un souci d'ouverture de l'université sur la Cité. Tous ces publics concourent *de facto* à la création d'un lieu inédit, porteur d'une alchimie nouvelle au cœur du campus.

## **XPERIUM, VITRINE DE LA RECHERCHE EN TRAIN DE SE FAIRE**

Espace scénographié, Xperium peut être décrit comme une vitrine de la recherche partenariale de l'université. À travers une thématique déclinée en une dizaine d'expériences et de manipulations, il s'agit de présenter la recherche développée au sein des laboratoires de l'université. Les thématiques se succèdent par cycles de deux ans. Initié en octobre 2016, le cycle intitulé « Quels défis pour REV3R ? Des innovations pour un monde durable » présente des expériences scientifiques, interactives et ludiques sur le thème de la 3<sup>ème</sup> révolution industrielle. Xperium s'appuie sur une direction scientifique et une équipe pédagogique composée d'enseignants-chercheurs et de doctorants, qui assurent la présentation des expériences. Parmi les nouveaux publics attendus, deux sont particulièrement ciblés : les lycéens et les entreprises. Ouvert depuis le 10 octobre 2016, Xperium a déjà accueilli plus de 1 500 visiteurs.

## **UN LIEU POUR L'ÉVÉNEMENTIEL**

L'intégration de cet espace au sein d'un ensemble plus vaste incluant bibliothèque et lieux de vie participe d'une volonté affirmée de favoriser la rencontre entre le monde socio-économique et l'université. Implanté autour d'un patio central, le complexe événementiel constitue un microcosme en soi et répond à l'ambition de concentration favorable à l'émergence d'une ambiance propice à l'innovation : grand et petit amphithéâtres équipés, deux salles de commission, un espace d'exposition/cocktail déjeuner. S'ajoute un espace dit « de promotion de l'innovation », dans lequel sont regroupés un ensemble de services, en appui à une politique volontariste pour faire de Lilliad une vitrine de l'innovation au sein de l'université ou en partenariat avec elle, à travers des événements, des expositions, des actions de promotion et de valorisation, organisées en propre ou en relation avec ses partenaires du monde socio-économique. En matière événementielle, outre une programmation spécifique dédiée à l'innovation, l'espace accueille les manifestations à caractère scientifique, institutionnel et partenarial de l'université,



➔ L'Atrium, espace de sociabilité, est situé au premier étage de la bibliothèque.

et verra en 2017 la mise en œuvre d'une politique locative externe. Ouvert depuis le 15 novembre 2016, le complexe événementiel a déjà accueilli plusieurs milliers d'utilisateurs, à l'occasion de plus de 20 événements.

## **UNE BIBLIOTHÈQUE REPENSÉE**

La bibliothèque – et les espaces de convivialité associés – occupe une place de choix puisque, symboliquement, elle est implantée dans l'enceinte de l'ancien bâtiment rénové alors que les nouveaux espaces – complexe événementiel et Xperium – intègrent l'extension.

Permettant aux usagers de travailler – espaces bibliothèque, de se détendre – espaces sociabilité, de se restaurer – café, il s'agit là du cœur de vie de Lilliad, ouvert de 8 h à 20 h, selon des horaires adaptés à la fréquentation du campus.

L'accueil comporte deux niveaux : l'un dédié à l'orientation générale est situé à l'entrée de Lilliad, l'autre dédié aux renseignements, se trouve au premier étage.

Autour d'un espace de sociabilité – atrium central sous un puits de lumière qui distribue la circulation des usagers, ce premier étage forme une bibliothèque en miniature, avec des espaces de travail ouverts, des salles de travail en groupe, des collections documentaires, des espaces de consultation internet, des imprimantes et des copieurs. La bibliothèque occupe également le deuxième étage du bâtiment et offre au total 850 places de travail et plus de 80 000 documents en libre accès.



## **TRAVAIL EN GROUPE**

Parce qu'un « learning center » est, étymologiquement, un « lieu pour apprendre », Lilliad comporte de nombreuses salles de travail en groupe, favorisant les nouveaux modes de travail en commun, ainsi qu'une salle d'innovation pédagogique. Pensée avec l'ensemble des acteurs concernés au sein de l'université, cette dernière offre aux enseignants un lieu propice à l'innovation pédagogique, aux chercheurs en sciences de l'éducation un terrain idéal d'observation des pratiques pédagogiques émergentes. Les usagers plébiscitent aujourd'hui le bâtiment, dans toutes ses dimensions (bibliothèque, sociabilité, restauration, valorisation de la recherche, espace événementiel), avec une fréquentation journalière comprise entre 3 000 et 4 000 entrées.

### **JULIEN ROCHE**

Directeur du SCD  
de l'université de Lille  
Sciences et Technologies  
julien.roche@univ-lille1.fr

[1] CPER : contrat de plan État-région, aujourd'hui contrat de projet État-région.

## Projet SGB mutualisé : les sites pilotes tout

Les sites pilotes du projet, regroupés en trois « vagues », s'apprêtent à signer au mois d'avril les marchés subséquents de l'accord-cadre SGBm. Un point sur la mise en œuvre du processus et l'accompagnement de l'Abes.

En octobre dernier, après seize mois d'un dialogue compétitif poussé avec les principaux éditeurs des systèmes de gestion de bibliothèques (SGB) de nouvelle génération, l'accord-cadre du marché SGBm, conclu notamment afin de maintenir un bon niveau de concurrence technologique et économique, a retenu quatre offres : Decalog-Data management, Ex-Libris, Linagora-Bibliore et OCLC. Dès décembre dernier, le processus débutait, les sites pilotes étant naturellement les premiers à se ré-informatiser au cours de l'année 2017. Regroupés en trois « vagues », ils ont rédigé les cahiers des charges des premiers marchés subséquents. En effet, si le cahier des clauses techniques particulières de l'accord-cadre définit les besoins techniques de l'ensemble des signataires du groupement de commande<sup>1</sup>, les documents des marchés subséquents permettent de sélectionner, parmi les titulaires de cet accord, l'offre la plus adaptée aux besoins spécifiques des sites constituant chacune des vagues. Ces premiers marchés seront signés en avril 2017<sup>2</sup>.

### Solo, duo et sextet

Un groupe pilote étant par définition hétérogène, la constitution de « vagues » s'est révélée un exercice relativement complexe, qui a donné

lieu, au final, à plusieurs stratégies. Ainsi, la Direction de la documentation de l'Université de Bordeaux a constitué une vague « solo », composée de cinq établissements bordelais ; le SICD de Toulouse et la BIU de Montpellier ont formé une vague « duo », composée de quatorze établissements. Pour leur part, les SCD de Caen, Clermont-Ferrand, Le Havre, Littoral, Rouen et la Bibliothèque Universcience se sont regroupés dans une « vague à six », probablement la plus complexe à mettre en œuvre.

### Défrichage et anticipation

Cette période de dialogue compétitif avec les fournisseurs a permis aux sites pilotes de défricher le terrain, donnant aux experts des sites – et de l'Abes – une expérience précieuse, partagée de différentes manières.

Ainsi, la « J-SGBm » – journée annuelle d'informations et d'échanges réservée aux membres du Groupement de Commande qui s'est tenue le 12 décembre dernier à la BULAC pour la 2<sup>ème</sup> année consécutive – a permis des échanges fructueux entre les experts des sites pilotes et de l'Abes et les membres des GC. Ce fut l'occasion d'aborder les sujets difficiles, d'anticiper la préparation des réorganisations, des migrations, des marchés subséquents, et de

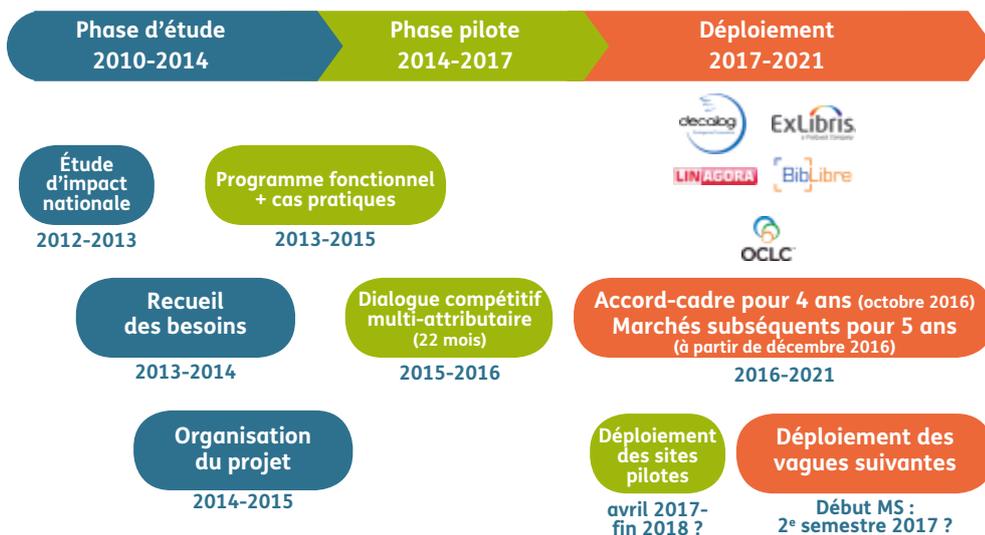
→ Le terrain ayant été défriché par les sites pilotes, des outils partagés pour accompagner le déploiement seront proposés.



Pixabay (CCO Public Domain)

# près de leur Graal ?

## DÉROULEMENT DU PROJET



vérifier la pertinence des dispositifs d'assistance ou de mutualisation envisagés.

### Retours d'expérience, FAQ et chantiers en cours

Parmi les dispositifs de concertation avec les sites pilotes, l'accompagnement sur les marchés subséquents et l'assistance ont déjà fait leurs preuves. Cette assistance montera progressivement en puissance, avec, par exemple, le guichet ABESstp, sur lequel seront publiées des « foires aux questions ». ABESstp, à travers les réponses à des interrogations précises, capitalisera l'expérience des sites pilotes.

Par ailleurs, pour accompagner la préparation de la migration des données des catalogues, certains services proposés par l'Abes ont évolué, une évolution profitant à l'ensemble des établissements du réseau Sudoc.

Parmi les dispositifs en cours de gestation, citons :

- un dispositif de sauvegarde national pour les données des « établissements SGBm » ;
- une plateforme mutualisée pour les formations et supports SGBm ;
- une coordination du recueil des besoins et des initiatives locales comme des développements sur les applications nationales.

Les chantiers sont nombreux et leurs livrables très attendus.

### Nouvelles vagues

Il va sans dire que les résultats des négociations des trois premiers marchés subséquents seront déterminants pour les marchés suivants. En effet, ce sont eux qui fixeront des repères sur les marges de négociation et la facilité d'appropriation.

À partir de mi-mars, les grands établissements pourront vérifier si les tarifs négociés sont accessibles, les réponses pour les plus petits établissements étant attendues pour fin avril.

Les prochaines sessions de déploiement devraient débuter en juin, puis suivre un rythme de deux par an, en janvier et juin. Plusieurs vagues par session étant possibles, un déploiement de plus de quinze établissements par an est envisagé, toute la difficulté résidant dans le fait de constituer des vagues d'établissements aux attentes comparables et en mesure de mobiliser leurs ressources au même rythme.

**SERGE GENOT**

Directeur du projet SGBm  
Abes  
serge.genot@abes.fr

[1] La liste des établissements signataires du groupement d'achats est parue dans le numéro 77 d'*Arabesques*.

[2] Depuis le lancement du projet SGBm, le marché des SIGB évolue : développement du logiciel libre, arrivée de nouveaux produits, concentration accrue des acteurs. Il y a un an, Proquest a racheté son principal concurrent, Ex-Libris. Ebsco a annoncé son soutien à Folio, plateforme open source qui est une évolution du logiciel américain Quali-Ole, et dont on attend des résultats concrets pour septembre 2017.

# EAD en bibliothèques : le choix de la modularité

Le cadre de gouvernance de l'EAD en bibliothèques se reconfigure, pour mieux prendre en compte la mosaïque des besoins en traitements rétrospectifs et la variété des attentes.

Depuis la mise en ligne, début 2013, d'un Guide national de bonnes pratiques EAD<sup>1</sup>, l'animation du catalogue EAD en bibliothèques a été amenée à prendre de nouvelles voies. À la gestation du schéma XML EAD3, dont les conditions d'application en France restent incertaines, ont fait pendant, sur la scène nationale, les réflexions du groupe de travail interministériel chargé de projeter un dispositif national de production en EAD, qui a remis, fin 2015, son étude à son comité de pilotage et au CSB<sup>2</sup>. Cette étude de faisabilité<sup>3</sup> a défini sept ensembles de besoins fonctionnels qu'appellent les pratiques de description de manuscrits et d'archives (non administratives) et a pointé, comme première urgence à traiter, l'absence d'environnements de catalogue mutualisés et adaptés aux besoins des établissements des collectivités territoriales (directement desservis par le département de la coopération de la BNF). Elle a souligné que les bonnes pratiques et l'usage de XMetaL-XMax (*plug-in* web commun à Pixml et Calames) serviront d'appui aux efforts de convergence; et que la perspective d'un référentiel national d'entités offrira des intersections essentielles avec le signalement des autres objets culturels et scientifiques.

## Une coordination interministérielle à quatre têtes

Des deux principaux scénarios examinés en 2015 (répliqués de Pixml ou de Calames Prod pour mettre à jour la base Manuscrits et archives du CCFr), c'est celui de la BNF qui a été retenu, le renforcement des collaborations techniques entre la BNF et l'Abes ainsi que l'objectif d'un rapprochement progressif des environnements de catalogue étant conservés.

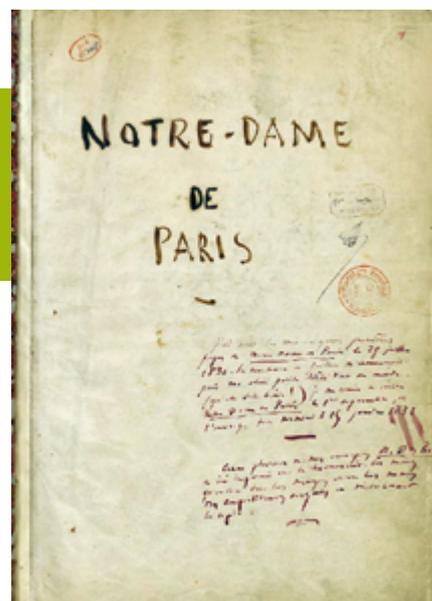
Prenant le relais du comité de pilotage, le comité exécutif a réinscrit à parité des représentants issus des périmètres culture

et ESR, de la BNF et de l'Abes au sein de quatre groupes de travail chargés de spécifier les évolutions en cours et d'adapter les pratiques EAD. En novembre 2016, leurs attributions ont été ainsi fixées : logiciels d'édition XML (pilote : Franck Bellugeon, BNF); interopérabilité des données (pilote : Jean-Marie Feurtet, Abes); analyse des réseaux de production et d'utilisation de l'EAD (pilote : Patrick Latour, bibliothèque Mazarine); interfaces de saisie et ergonomie des outils (pilote : Florent Palluault, médiathèque François-Mitterrand de Poitiers).

## La souplesse plutôt qu'un « couteau suisse »

Dans cette nouvelle distribution, l'adaptation des environnements Pixml et Calames reste un enjeu central, mais les efforts ne se concentrent plus en direction des seules problématiques d'outillages XML. Un temps entretenue par la possibilité d'une réappropriation du plugin collaboratif XML développé par l'Equipex Bibliissima, l'idée d'un « couteau suisse » pour l'EAD en bibliothèques a fait long feu. Il s'agit plutôt de définir un cadre commun, interstructurel et suffisamment adaptable pour convenir à une diversité d'acteurs et de cycles de vie des données, en termes de modèle économique, de gouvernance et d'accompagnement des réseaux.

Les premières opérations de signalement régional exhaustif des manuscrits champardenois menées par l'équipe CCFr et Interbibly ont mis en évidence la mosaïque des besoins en traitements rétrospectifs, la variété des attentes des producteurs potentiels et la gageure que représenterait la quête d'un environnement technique unique et véritablement polyvalent. Parallèlement, forts d'une expérience quasi décennale, la BNF et le réseau Calames ont multiplié les interdépendances entre, d'une part,



➔ Première page du manuscrit de Victor Hugo, conservé à la BNF.

Source : Gallica/BNF

L'EAD est un format de catalogage utilisé pour la description des manuscrits et des documents d'archives.

leurs données EAD, et, d'autre part, les numérisations patrimoniales, le traitement des archives de la recherche et des fonds iconographiques ou hybrides. Pour autant, ces données restent peu portables vers d'autres environnements (bibliothèques numériques, SIGB, portails, outils de gestion, bases RDF, projets de publications spécifiques...), difficiles à représenter dans d'autres formats plus courants au sein des mêmes communautés (Dublin Core, Marc...), sans probation encore dans le traitement émergent des archives numériques, et en attente d'une modélisation (Records in Context<sup>4</sup>) qui devrait favoriser leur désenclavement à moyen terme. Mappings, protocoles d'échanges, modèles de saisie, et d'autres ressources encore restent largement à mutualiser et à encadrer de recommandations communes.

Les livrables des quatre groupes de travail sont attendus d'ici à l'automne 2017.

**JEAN-MARIE FEURTET**

Responsable de Calames, Abes  
jean-marie.feurtet@abes.fr

[1] [bonnespratiques-ead.net](http://bonnespratiques-ead.net) Pour mémoire, le site officiel du format Encoded Archival Description est maintenu par la Library of Congress : [loc.gov/ead](http://loc.gov/ead)

[2] Le CSB (Comité stratégique bibliographique) a succédé en 2011 au Comité « Sudoc-SI BNF-CCFr » pour la coordination des actions bibliographiques nationales.

[3] [abes.fr/Media/Fichiers/Calames/Etude\\_faisabilite\\_outil\\_national\\_EAD\\_nov2015](http://abes.fr/Media/Fichiers/Calames/Etude_faisabilite_outil_national_EAD_nov2015)

[4] [ica.org/fr/egad-ric](http://ica.org/fr/egad-ric)

# Cercles-Bacon : à vous de jouer !

Sous la forme de chantiers qualité, les Cercles confient aux réseaux partenaires des opérations essentielles jusqu'ici réalisées par l'Abes.

Le service rendu par l'Abes ne pourrait s'accomplir sans l'aide des professionnels de la documentation qui constituent nos réseaux. Nous œuvrons pour les réseaux, avec les réseaux. Cercles-Bacon en est la parfaite illustration.

## Métadonnées à la source

Revenons d'abord à la naissance de Bacon, qui répondait au besoin, exprimé par les établissements, d'améliorer la gestion de la documentation électronique. L'Abes, lorsqu'elle a construit cette base de connaissance nationale, a décidé d'accompagner les éditeurs scientifiques francophones de façon à les convaincre d'appliquer la recommandation KBART (Knowledge Bases and Related Tools, portée par la NISO<sup>1</sup>).

Il fallait non seulement obtenir des progrès quant au signalement et à la gestion des accès dans les outils de découverte, comme KBART le préconise, mais également – et c'était là un pari ambitieux – améliorer la qualité des métadonnées à la source, autrement dit chez les éditeurs. L'ensemble du circuit allait ainsi bénéficier des améliorations tout en gardant sa cohérence.

Soulignons en effet que Bacon ne modifie jamais les métadonnées transmises par les éditeurs : une différence entre les informations contenues dans les fichiers KBART et celles des bases de données des éditeurs interromprait l'accès. Partant du principe qu'une information inaccessible est une information inutile, nous préférons conserver les métadonnées en l'état et transmettre les erreurs aux éditeurs pour qu'ils les corrigent.

## Interlocuteurs des éditeurs

À ce jour, l'Abes se charge de sensibiliser les éditeurs à Bacon et au suivi de la recommandation KBART ; l'agence évalue les métadonnées transmises

et communique le résultat à l'éditeur ; elle maintient une base de données contenant les métadonnées au format KBART et expose les fichiers via le site web et les webservices.

Avec l'ouverture des Cercles-Bacon, certaines tâches seront désormais reportées sur les réseaux. Chaque référent Cercles-Bacon sera responsable de la collecte des fichiers auprès de l'éditeur, de leur évaluation, de la rédaction du rapport d'évaluation et du chargement des fichiers dans la base Bacon. Les rôles ainsi redistribués, l'Abes pourra revoir ses priorités quant à la gestion du projet.

Cette transmission de tâches présente des avantages, dont le premier concerne le statut de l'interlocuteur auprès de l'éditeur. En effet, les personnels de bibliothèques sont les acquéreurs des ressources électroniques. Du point de vue des éditeurs, ce sont eux qui décident d'acheter ou non une ressource, ce sont eux leurs véritables clients. Ils disposent en conséquence d'un pouvoir de persuasion plus important que l'Abes. De plus, les référents auront accès au corpus traité, ce qui, là encore, n'est pas le cas de l'Abes. Or l'analyse est plus pertinente si l'analyste dispose de cet accès.

## Traversez la passerelle !

Pourquoi n'avoir pas confié cette mission aux réseaux dès le départ ?

En réalité, les traitements ont été internalisés pour une seule et unique raison : il fallait prendre le temps de réaliser les outils et de se confronter aux traitements à appliquer aux métadonnées.

Aujourd'hui les outils sont prêts et le moment est venu de passer le relais. Du laboratoire où il a été conçu jusqu'à l'atelier où il a été réalisé, Bacon arrive aujourd'hui à maturité<sup>2</sup>.

Cercles-Bacon est une invitation à s'approprier Bacon, une passerelle entre

les Bibliothèques et l'Abes : un atelier participatif auquel les membres des réseaux seront chaleureusement conviés dans les mois à venir.

Le premier établissement à avoir répondu favorablement à notre appel est la BU de Poitiers, en la personne de Léa Maubon, responsable de la documentation électronique. Elle aura à sa charge le corpus Brepols. Qu'elle en soit ici remerciée.

**CYRIL LEROY**

Gestionnaire de métadonnées,  
Abes  
cyril.leroy@abes.fr

[1] [www.niso.org/workrooms/kbart](http://www.niso.org/workrooms/kbart) (en anglais).

[2] Lire la série d'articles à ce sujet sur Punktokomo : <https://punktokomo.abes.fr/2016/09/12/openrefine-au-service-de-bacon-quelle-evaluation-pour-les-fichiers-kbart-1-introduction>



Pixabay (CCO Public Domain)

# Urfist de Bordeaux : l'open science par la pratique

**A** Bordeaux, depuis trois ans, chercheurs et doctorants attentifs à éviter les pièges de la « *bad science* » se forment à des logiciels conçus par leurs pairs. Et leurs besoins sont souvent comblés au-delà de leurs espérances.

Dans les domaines les plus divers, les publiants reçoivent, de la part des financeurs comme de leurs établissements, des incitations de plus en plus fortes pour œuvrer en faveur de l'*open science*. S'il n'en existe pas de définition canonique, elle consiste, comme son nom l'exprime, en un mouvement d'ouverture appliqué à l'ensemble des activités de recherche. Il s'agit donc de s'assurer de la transparence des méthodes employées; mais aussi d'assainir les processus de communication scientifique, de garantir la disponibilité et la réutilisabilité des données en encourageant notamment le recours aux outils numériques qui favorisent la collaboration. L'*open science* soulève des défis techniques, juridiques et culturels, voire éthiques : comment mener une recherche intègre et transparente tout en satisfaisant aux critères d'évaluation actuels et aux exigences de sa discipline ?

## Valeur ajoutée... et avérée !

Par les perspectives qu'il ouvre, le contexte s'avère favorable à de nouvelles formes de collaboration entre professionnels de l'information scientifique et technique et chercheurs. Mais les professionnels de l'IST doivent relever un double défi. D'une part, leur offre de services ne doit pas être perçue comme s'inscrivant sur un terrain uniquement prescriptif, théorique. Si les enjeux de qualité et d'intégrité de la recherche font consensus parmi les chercheurs, les moyens pour atteindre ces objectifs demeurent difficiles à définir, tant ils dépendent de pratiques et de valeurs inhérentes à chaque discipline. Et surtout, ainsi que le soulignent Paul Smaldino et Richard McElreath<sup>1</sup>, les pratiques de recherche restent grandement tributaires de modalités d'évaluation mettant l'accent sur la publication régulière de résultats dits innovants dans des revues à forte visibilité. Quelle peut être la contribution des professionnels de l'IST à l'effort collectif en faveur d'une recherche plus transparente, alors que tout concourt au développement de ce que Smaldino et McElreath nomment la « *bad science* » ? D'autre part, les

[1] Paul E. Smaldino et Richard McElreath, « The Natural Selection of Bad Science », arXiv:1605.09511 [physics, stat], 2016.

[2] Gilles Bastin et Milan Bouchet-Valat, « R.TeMiS, un logiciel libre pour l'analyse textuelle fondé sur R », présenté au Congrès de l'Association française de sociologie, Nantes, 2013.



professionnels de l'IST soulignent la difficulté de convaincre les chercheurs de la valeur ajoutée de leurs compétences.

Enfin, les services de formation se confrontent à une autre interrogation : à l'heure de la désintermédiation et des agendas surchargés, la formation en face à face ne serait-elle pas un archaïsme ? Pas si sûr. En trois ans et à moyens constants, le nombre de chercheurs et enseignants-chercheurs formés par l'Unité régionale de formation à l'information scientifique et technique (Urfist) de Bordeaux a quadruplé. Sur la même période, les effectifs de doctorants formés ont, quant à eux, doublé. En 2016, 69 % des usagers formés sont des doctorants et des chercheurs.

## Les chercheurs parlent aux chercheurs

L'Urfist de Bordeaux fait partie d'un réseau comprenant six autres unités, qui sont implantées à Lyon, Nice, Paris, Rennes, Strasbourg et Toulouse. Services interuniversitaires, les Urfist ont pour mission la formation des doctorants, chercheurs, enseignants-chercheurs et personnels d'appui à la recherche. Ils accueillent des usagers de toutes les disciplines. Le ministère de l'Éducation, de l'Enseignement supérieur et de la Recherche détermine pour tout le réseau les domaines de compétence et les objectifs que chaque unité décline en un programme d'actions qui lui est propre.

Depuis 2014, l'Urfist de Bordeaux met l'accent sur des formations à des logiciels conçus par des chercheurs pour des chercheurs, gratuits et souvent open source. On peut citer Scientifig (annotation et formatage de figures), TXM (analyse de corpus textuels), FactoMineR (analyse exploratoire de données), R.TeMiS (création, édition, manipulation de corpus sous R), RQDA (analyse qualitative de données) ou encore DtmVic (analyse de données statistiques et textuelles). Ces logiciels sont élaborés dans le cadre de projets de recherche ou pour répondre aux besoins spécifiques de leurs concepteurs. Leurs modalités de diffusion sont extrêmement variables : des équipes organisent régulièrement

des formations sur leur site universitaire ou bien à l'extérieur; des chercheurs déposent directement leur logiciel sur une plateforme ou sur leur site web, sans médiation complémentaire. L'Urfist participe à l'élargissement et à la diversification des communautés d'utilisateurs de ces logiciels en les diffusant en dehors des réseaux habituels de leurs concepteurs.

Dès que possible, l'Urfist de Bordeaux fait appel au développeur ou à l'un des concepteurs du logiciel pour animer la formation. Dans le cas d'outils peu diffusés, le niveau d'attente des usagers peut être modéré voire faible. Ainsi, de nombreux questionnaires d'évaluation complétés par les chercheurs témoignent d'un hiatus entre la faiblesse de leurs attentes et l'utilité de la formation quant à leurs besoins : « Sans *a priori* sur l'intérêt du logiciel, je repars avec la certitude de son utilité pour mon travail », « Je n'avais pas de grandes attentes par rapport à cette formation. J'ai été très positivement surprise. Le logiciel est vraiment intéressant et sa mise dans le contexte des autres logiciels existants a été bien faite et extrêmement utile. Merci! », « Très bonne présentation générale d'un outil peu connu, gratuit, apparemment facile à utiliser. » Il est à noter, également, que le profil du formateur contribue pour beaucoup à légitimer la perception du logiciel présenté.

### **Publics pluriels et *barcamp***

Bien que largement tournées vers la pratique, les formations ne se réduisent pas à des séances de manipulation d'interfaces. Le choix d'un logiciel *open source* plutôt que propriétaire ouvre aussi un questionnement sur les processus et les méthodes de recherche. On touche ici à des questions épistémologiques. Ainsi que le rappellent les concepteurs de R.TeMiS (R Text Mining Solution), Gilles Bastin et Milan Bouchet-Valat, « les logiciels disponibles sont fortement ancrés dans les contextes théoriques dans lesquels ils ont été élaborés ». Ils précisent que R.TeMiS, qui applique les méthodes de la statistique textuelle sur des corpus de nature différente, « permet à l'utilisateur de s'affranchir plus facilement d'un environnement statistique déterminé », et a été conçu pour « promouvoir une approche ouverte et réflexive des corpus de données textuelles »<sup>2</sup>. Par ailleurs, les formations étant ouvertes aux publics de toutes disciplines, elles donnent l'occasion aux participants d'appréhender d'autres types de pratiques. Les publics de ces formations sont pluriels, voire inattendus. Ainsi, archéologues et spécialistes de la préhistoire sont venus massivement se former au logiciel Scientifig, conçu par Benoît Aigouy, chercheur en biologie. La formation des doctorants et des personnels d'appui à la recherche joue par ailleurs un rôle

fondamental dans l'évolution des pratiques au sein des unités de recherche.

En quête de modalités de collaboration plus directes entre acteurs de la recherche autour de l'*open science*, l'Urfist de Bordeaux a conçu et piloté, le 20 octobre 2016, le premier *barcamp* sur l'*open access* du site universitaire bordelais. Se sont associés à l'Urfist : le CNRS, l'Inria, Sciences Po Bordeaux, l'Université de Bordeaux, l'Université de Bordeaux-Montaigne et le consortium Couperin. Il s'agissait de concevoir un événement centré sur des pratiques plutôt que sur un projet d'établissement en particulier et de proposer un format à même de refléter une pluralité de points de vue. Le choix s'est rapidement porté sur la formule du *barcamp*, qui confère aux participants un rôle actif dans le déroulement de la manifestation. Sur la base des propositions des participants, trois ateliers ont été proposés : « Marre de prêcher dans le désert? Renouvelons l'accompagnement vers le libre accès », « Le libre accès a-t-il besoin des institutions? Nos institutions ont-elles besoin du libre accès? » et « D'un échec annoncé à un projet gagnant : la clé de vos succès pour le libre accès ». Durant une journée, chercheurs, professionnels de l'IST et personnels d'appui à la recherche ont ainsi pu débattre ensemble.

Ainsi, en promouvant des logiciels de recherche gratuits et si possible *open source*, l'Urfist de Bordeaux ancre ses services sur le terrain de la pratique, sans pour autant adopter un point de vue techniciste. Former les acteurs de la recherche à des alternatives aux produits propriétaires contribue, à son échelle, à l'émergence d'une recherche plus transparente. Formations, ateliers et journées d'étude représentent autant d'espaces de dialogue entre chercheurs de disciplines variées. Son prochain défi? Renforcer la formation des chercheurs aux enjeux juridiques de l'*open science*.

SABRINA GRANGER

Conservateur des bibliothèques  
Responsable de l'Urfist de Bordeaux  
sabrina.granger@u-bordeaux.fr

## **UN GIS EN 2017**

**La collaboration inter-Urfist va s'intensifier grâce à la fondation, cette année, d'un groupement d'intérêt scientifique (GIS). Le réseau des Urfist avait créé son association dès 2010. Mais l'émergence de thématiques nouvelles, la reconfiguration du paysage universitaire et les besoins croissants de coordination rendent nécessaire une nouvelle structure de coopération, qui permettra aux Urfist de développer des partenariats avec d'autres acteurs de l'IST. Sa priorité : structurer une activité de recherche au niveau national et international autour de thèmes relatifs à l'IST, tout en continuant de développer des actions de formation innovantes.**

Une rubrique qui vous présente la diversité des professionnels qui constituent les réseaux de l'Abes.

(Portrait)

## Yves TOMIC

Responsable du système d'information documentaire de la bibliothèque de l'université Paris-Dauphine

### **Quelles sont vos fonctions actuelles ?**

La bibliothèque de l'université Paris-Dauphine détient une collection d'ouvrages académiques et de revues en langue étrangère de haut niveau en sciences économiques et gestion. J'y dirige le service du système d'information documentaire, qui réunit quatre personnes. Ce service administre un large portefeuille d'applications. En 2016, notre service a été mobilisé autour du projet de nouveau portail, unifiant site institutionnel et portail documentaire dans une seule et même interface, et développé en interne. En outre, je suis responsable de la mission Indicateurs relatifs aux activités de la bibliothèque, et je suis membre du conseil scientifique de l'université.

### **Comment résumeriez-vous les grandes étapes de votre parcours professionnel ?**

C'est grâce à la connaissance de plusieurs langues slaves que je suis entré, en 1995, dans le monde des bibliothèques, plus précisément à la bibliothèque de la Sorbonne. Cette première expérience m'a permis d'intégrer ensuite la Bibliothèque de documentation internationale contemporaine (BDIC) où je suis devenu, en 1997, responsable du secteur Balkans.

Au cours de ces années, je menais également des activités de recherche au sein de l'Association française d'études sur les Balkans et de sa revue *Balkanologie*. Cette recherche sur les pays de l'ex-Yougoslavie m'a conduit à produire des études pour les ministères des Affaires étrangères et de la Défense, ainsi que pour le Tribunal pénal international pour l'ex-Yougoslavie. Et c'est aussi à travers ce travail d'historien que j'ai appréhendé les enjeux de la documentation académique côté utilisateur.

À la BDIC, j'étais responsable du site Internet et effectuais déjà de nombreux développements. Mais c'est en 2011 que je me suis orienté plus clairement vers l'informatique documentaire en intégrant le Service commun de documentation de l'université Paris-Sud, où ma mission a été le déploiement du SIGB Koha. En octobre 2014, j'ai été recruté par la bibliothèque de l'université Paris-Dauphine afin de gérer son système d'information documentaire et mettre en place un service qui s'y consacre.

### **À quand remontent vos premiers contacts avec l'Abes, et dans quel contexte ?**

J'ai noué des liens avec des collègues de l'Abes autour des applications que j'ai développées. En effet, elles exploitent les services web mis à disposition par l'agence.

La première application a été installée en 2014 : il s'agit de l'outil DoMyBiblio, qui édite des listes bibliographiques à partir de numéros PPN ou ISBN. A suivi le service de contrôle qualité des notices bibliographiques, CheckSudoc.



Enfin, le portail UnivDoc permet d'interroger les données du Sudoc dans le cadre d'un outil de découverte, fournissant aussi des résultats de CrossRef, ScienceDirect, Springer, JournalTocs, Gallica. J'ai eu l'occasion de présenter aux collègues, lors des Journées Abes de 2014, les fonctionnalités qu'il était possible de développer avec les services web de l'agence.

### **Quels défis majeurs, d'après vous, aura à relever l'Abes dans les prochaines années ?**

L'Agence devra mener à bien la modernisation du Sudoc (l'interface publique et les outils de description des données) tout en gérant le(s) projet(s) de SGBm qui n'étaient pas prioritaires au regard du premier défi.

### **Qu'appréciez-vous le plus dans votre métier ?**

La créativité, l'inventivité et l'expérimentation qu'il est possible de déployer. L'université Paris-Dauphine s'y prête particulièrement bien car l'établissement s'inscrit dans une démarche forte d'innovation et d'expérimentation. En témoigne le nouveau portail que nous avons réalisé en interne et qui intègre la recherche dans le Sudoc, par exemple. Nous proposons une articulation entre les données de l'outil de découverte dont nous exploitons l'API et les données provenant du catalogue du Sudoc par le biais de ses différents services web.

### **Qu'est-ce qui vous énerve le plus ?**

La démission de certaines bibliothèques face à la chose informatique, perçue comme compliquée – alors qu'il suffit de recruter un ingénieur ou deux ayant les compétences requises afin de maintenir et faire évoluer un système d'information.

### **Si l'Abes était un animal, ce serait... ?**

Un animal d'envergure, au vu du volume de métadonnées qu'elle gère !

### **Votre expression favorite ?**

« On n'est jamais mieux servi que par soi-même. » Je crois que l'on répond mieux aux besoins des utilisateurs lorsque l'on met en place des solutions au plus près d'eux. Il ne faut pas avoir peur des développements réalisés en interne.

yves.tomic@dauphine.fr