

# Ar(*abes*)ques

AVRIL - MAI - JUIN 2017

DOSSIER

## Autorités, identifiants, entités *L'expansion des référentiels*

**PLEINS FEUX SUR** • Le site Richelieu : les bibliothèques de l'École des chartes  
et de l'Institut national d'histoire de l'art

**ACTUALITÉS** • Tout sur les Journées Abes

85  
abes  
agence bibliographique  
de l'enseignement supérieur



## (Dossier) AUTORITÉS, IDENTIFIANTS, ENTITÉS *L'expansion des référentiels*

Le volume et la diversité des métadonnées en circulation dans les systèmes d'information (de l'enseignement supérieur, de la recherche, de la culture) exigent de repenser le rôle des référentiels d'autorité. Considérés comme données de confiance au service du développement de l'*open data* et du web sémantique, ils constituent un capital précieux, une garantie d'indépendance, tout en interrogeant en profondeur les pratiques catalographiques « classiques ». S'ouvrant sur un éclairage terminologique apporté par les principaux experts du domaine, ce dossier présente d'abord la stratégie de l'Abes en la matière, principalement au travers d'IdRef – Identifiants et référentiels pour l'ESR. Il y est ensuite question du Fichier national d'entités (FNE), projet mené par la BnF et l'Abes, du rôle du Répertoire national des structures de recherche (RNSR), puis du « graphe culture » du MCC visant à l'interconnexion de données et référentiels culturels, ou encore de la belle expérience de science participative du Muséum de Paris. C'est un tour d'horizon des grands référentiels nationaux pour comprendre les acquis... et ce qui reste à accomplir!

## 22 (Pleins feux sur...)

### Le site Richelieu : les portes s'ouvrent sur l'art et l'histoire

CAMILLE DÉGEZ-SELVES ET ANNE-ELISABETH BUXTORF  
Bibliothèque des Chartes / Bibliothèque de l'INHA



© Clément Olivier Ouadah / INHA, juin 2016

## 24 (Actualités)

## 28 (Portrait)

04 Des référentiels et de leur usage aujourd'hui  
V. BOULET

06 3 fois sur le métier remettons les notions  
F. MISTRAL, O. ROUSSEAU, Y. NICOLAS, P. LE PAPE

08 IdRef, les autorités en conquête et en partage  
F. MISTRAL

10 Qualinca et IdRef : l'intégration est en cours !  
A. LE PROVOST

11 Nous avons les coffres, il nous faut les bonnes clés !  
J. POUMEYROL

12 Ils utilisent IdRef  
• *Persée, partenaire confirmé* V. BOULÉTREAU  
• *ADUM, plus de visibilité pour les doctorants* C. MORALES  
• *Conditor, le futur pot commun de métadonnées* A. CORET  
• *AureHAL et son idHAL rassembleur* B. MARMOL ET B. KUNTZIGER  
• *SyMoGIH, de l'UMR 5190 – Larhra, et les « objets historiques »*  
• *Prelib et la littérature bretonne*  
N. BLANCHARD, J.-B. PRESSAC ET M. THOMAS  
• *Caplab, pour un suivi de l'activité des unités* R. THOUY

15 Le RNSR, colonne vertébrale du SI recherche  
I. POULIQUEN, I. KABLA-LANGLAIS ET X. CHEN

16 Le futur FNE : vers une vraie coproduction  
F. JOHANNIC-SETA  
Vers un nouveau Rameau ? F. MENARD

18 La culture dessine son « graphe » M.-V. LEROI

20 L'herbier du Muséum : retour d'expérience sur la coconstruction de référentiels  
C. BESOMBES, S. CHAGNOUX ET G. ILLIEN

## 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 : Bob May/Flickr (CC BY-NC-SA 2.0)  
[www.flickr.com/photos/alternative\\_illustrations/16421670432](http://www.flickr.com/photos/alternative_illustrations/16421670432)

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

# Incroyable mai(s) vrai !

Chère lectrice, cher lecteur, vous tenez entre les mains un trésor.

En effet, cette livraison d'*Arabesques*, fabriquée avec patience, minutie et générosité par tous ses auteurs et l'équipe de rédaction, vous révélera de grands secrets. Vous y découvrirez que désormais les référentiels font autorité !

Les plus avisés d'entre les professionnels savaient déjà que de grandes choses se passaient dans les officines bibliothéconomiques nationales. Mais aucun, j'en suis sûr, n'avait idée de la folle ambition, de la quantité colossale de travail, d'intelligence et d'expertise collectives, ni des résultats stupéfiants déjà obtenus par celles et ceux qui œuvrent à la transfiguration des autorités (auteurs, collectivités, sujets, lieux géographiques) en immenses référentiels interopérables et exposés sur le web. Aujourd'hui les autorités, créées par l'intelligence humaine, se partagent et circulent entre les machines, pour enrichir les métadonnées de tous les objets bibliographiques, et bien au-delà, de toutes les entités du réel. Ainsi en sont enrichies la description du monde et la recherche de tous les types d'informations et de connaissances utiles aux hommes.

Vous lirez dans les pages qui suivent tout ce que les acteurs nationaux ont réalisé, souvent en fort aimable



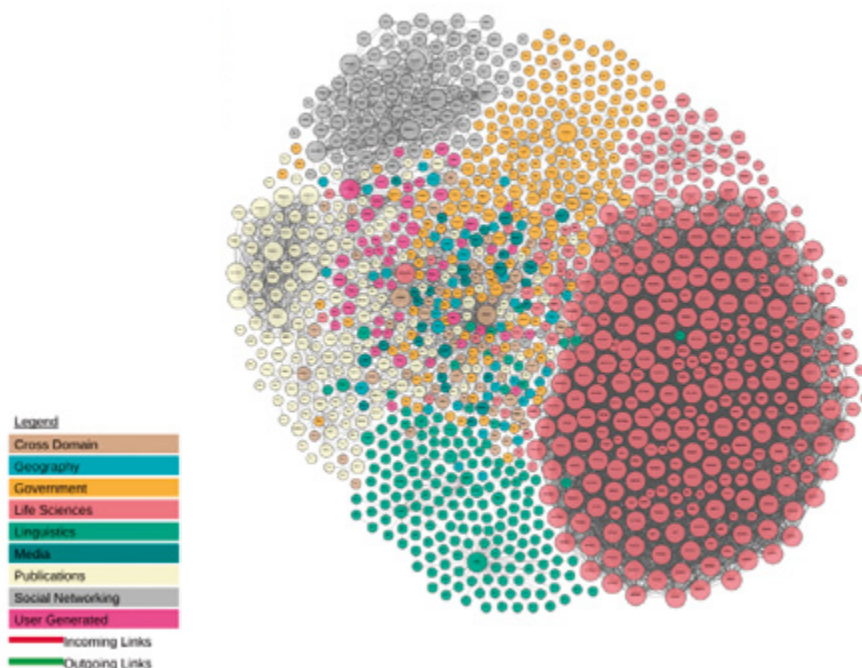
collaboration, pour valoriser le travail de catalogage informatisé de 50 années réalisé par les milliers de bibliothécaires, documentalistes, archivistes, informaticiens et spécialistes de l'IST que nous sommes.

## « Désormais, les référentiels font autorité ! »

De nombreuses réalisations sont encore à venir, décrites également dans ce dossier, qui alimentent le rêve et stimulent la créativité. Ce qui est aussi remarquable, c'est que ces travaux menés à l'échelle nationale s'inscrivent dans une démarche mondiale menée par les mêmes corporations et contribuent à ce qui constitue sans doute le grand œuvre de ce siècle informationnel, le web des données ouvertes et liées<sup>1</sup>, LOD de son petit nom.

DAVID AYMONIN  
Directeur de l'Abes

[1] [https://fr.wikipedia.org/wiki/Linked\\_open\\_data](https://fr.wikipedia.org/wiki/Linked_open_data)



LOD cloud diagram 2017 : représentation des jeux de données publiés au format Linked Data (image réalisée et maintenue par A. Abele, J.P. McCrae, P. Buitelaar, A. Jentzsch et R. Cyganiak), interactive sur [lod-cloud.net](http://lod-cloud.net).

# Des référentiels et de leur usage aujourd'hui

**Le catalogage se transfigure : d'un monde où l'on normalisait des notices pour les échanger, on passe à un monde où ce que l'on expose, ce sont des données structurées pour les partager. Le classement hiérarchique cède alors le pas au contrôle des éléments de description d'une ressource.**

**E**n 1627, Gabriel Naudé, dans son *Advis pour dresser une bibliothèque*, comparait le travail de classement d'un bibliothécaire au rôle d'un médecin. Afin de pouvoir maîtriser la « quantité infinie » des publications, il appelle à établir un ordonnancement, à des fins de sélection.

Le classement, perçu comme un instrument de domination et de sélection dans un univers fini, sert à dire ce qui est digne ou non de figurer dans une bibliothèque : « Toutefois pour ne laisser cette quantité infinie ne la définissant point, et aussi pour ne jeter les curieux hors d'esperance de pouvoir accomplir et venir à bout de cette belle entreprise, il me semble qu'il est

avec l'essor du web, qui implique non seulement une croissance exponentielle du nombre de ressources, mais renforce encore le caractère composite et évolutif d'une ressource, qui n'est plus que la décomposition et la recombinaison à l'infini de données.

La notion de web de données (*linked data*) consacre l'explosion de l'idée de « ressource ». Il s'agit de faciliter l'exploitation de données structurées, c'est-à-dire de données qualifiées et encadrées, de sorte que des machines puissent les interpréter facilement. Le web de données consiste à exposer ces données structurées sur le web et les relier entre elles, ce qui permet d'accroître leur visibilité et leur réutilisation. En outre, le boom de la production de données, de l'ordre de 2,5 trillions d'octets chaque jour, provenant souvent, par le retraitement automatisé, de données d'origines diverses, implique l'essor de nouvelles technologies de stockage et de traitement, appelées Big Data (ou mégadonnées). La manière dont les données utilisées ou produites par les bibliothèques s'insèrent dans le Big Data est un enjeu désormais identifié par l'IFLA<sup>2</sup>.

## Peut-on encore « ordonner la quantité de drogues suivant la qualité d'icelles »?

*à propos de faire comme les Médecins, qui ordonnent la quantité de drogues suivant la qualité d'icelles, et de dire que l'on ne peut manquer de recueillir tous ceux qui auront les qualités et conditions requises pour être mis dans une Bibliothèque*<sup>1</sup>.

Dans cette logique, la notion de description et de classement est vécue de manière hiérarchique et verticale.

Gabriel Naudé, même s'il n'emploie pas le terme, conçoit le rôle du bibliothécaire comme un médiateur : un patient a besoin du médecin pour se soigner ; un utilisateur a besoin d'un bibliothécaire pour accéder à des ressources. Un certain nombre d'éléments structurants amènent à repenser de telles conceptions.

### EXPLOSION DE L'IDÉE DE « RESSOURCE »

Les pratiques et les usages ont d'ores et déjà évolué en remettant en cause l'idée d'un classement hiérarchique. L'utilisateur prend des concepts ou des notions et navigue entre elles pour accéder à des ressources. Les modes de recherche sur le web incitent à une telle évolution. La logique de graphe, composé d'entités reliées par des relations qualifiées, prend donc le pas sur celle de classification hiérarchisée.

La question de la « quantité infinie » évoquée par Gabriel Naudé est devenue encore plus critique

### DE LA NOTICE NORMALISÉE À LA DONNÉE STRUCTURÉE

Nous sommes donc en présence d'une contradiction apparente : comment concilier la nécessité de données structurées, soulignée par les réalisations du web de données, et l'essor exponentiel des données produites, impulsé par le Big Data ?

Un peu de dialectique permet de la dépasser : c'est grâce à la structuration renforcée des données que celles-ci pourront être traitées plus rapidement et plus efficacement. Si l'on applique ce principe aux données produites ou utilisées par les bibliothèques, cela conduit à passer d'un monde où l'on normalisait des notices pour pouvoir les échanger à un monde où l'on structure des données pour pouvoir les partager. Le concept de contrôle bibliographique universel, au cœur de l'écosystème international des données de bibliothèque, s'en trouve redéfini. Afin de prendre en compte ces différents éléments structurants, la France s'est lancée dans le programme de la « transition bibliographique »<sup>3</sup>, dont l'objectif est de se doter de règles et d'outils permettant de mettre en pratique ce changement de

[1] Sur Gallica, [gallica.bnf.fr/ark:/12148/bpt6k1041429f](http://gallica.bnf.fr/ark:/12148/bpt6k1041429f), p. 41-42.

[2] Le rapport de développement stratégique de l'IFLA de 2013 (<https://trends.ifla.org/literature-review/big-data>) a amené en 2016 la constitution d'un groupe d'intérêt spécial consacré au Big Data.

[3] [www.transition-bibliographique.fr](http://www.transition-bibliographique.fr)

[4] La version définitive de ce modèle conceptuel fusionnant FRBR-ER (FRBR OO n'est pas concerné), FRAD et FRSAD est attendue par l'automne 2017. Voir le projet diffusé pour enquête internationale ([www.ifla.org/files/assets/cataloguing/frbr-irm/frbr-irm\\_20160225.pdf](http://www.ifla.org/files/assets/cataloguing/frbr-irm/frbr-irm_20160225.pdf)) et une présentation du modèle par Pat Riva (congrès IFLA 2016) (<http://library.ifla.org/1084/7/207-riva-fr.pdf>)

[5] Gabriel Naudé, *op. cit.*, chapitre 9, « quel doit être le but principal de cette bibliothèque ».

paradigme. Concrètement, cela suppose d'appliquer le modèle de données FRBR, désormais sous sa forme dite « IFLA-LRM »<sup>4</sup>, d'en tirer un code et des règles de catalogage en s'inspirant de RDA, de former la communauté professionnelle aux enjeux et aux normes produites dans ce cadre, et de pousser les systèmes et les outils à s'adapter.

## PÉRIMÈTRE ÉLARGI

Pleinement concernés par la transition bibliographique, les référentiels recouvrent le périmètre des données d'autorité. Cependant, la notion de référentiel va plus loin, car elle annexe l'ensemble des éléments contrôlés d'une description de ressource (langues, types de document comme la nouvelle zone 0 de l'ISBD...).

Ainsi, la notion de données d'autorité, opposée à celle de données bibliographiques, n'est plus pertinente aujourd'hui. Nous avons désormais affaire à des entités dont les éléments de description, les données, sont contrôlées par des référentiels. Par exemple, pour décrire une entité « manifestation », nous avons besoin des entités « expression », « œuvres », « personne » accompagnées de liens qualifiés (« auteur », « sujet »). Il nous faut également une entité « date », une entité « langue »... qui ne font pas partie du périmètre traditionnel des données d'autorité.

Cela implique d'élargir et d'affiner le périmètre des données structurées. Les bibliothèques en possèdent, de qualité, bien structurées. Le format Marc étant un peu « le web de données de Monsieur Jourdain », il ne s'agit donc pas de crier à la « fin du catalogage », mais d'accompagner sa transfiguration. D'autres organismes publics en possèdent également : les institutions patrimoniales (archives, musées...), les éditeurs publics (DILA, éditeurs universitaires...). La mutualisation de ces données – qui permet tout à la fois de les confronter, donc de renforcer la confiance qu'elles inspirent à des tiers susceptibles de les récupérer, et de faire masse – est devenue une nécessité... et un atout majeur. En effet, la production mutualisée et la réutilisation des données publiques constituent les fondements d'un véritable service public de la donnée.

Différentes initiatives sont prises dans ce sens, à l'image de la perspective d'un fichier national d'entités (voir l'article « FNE » page 16), fusionnant dans un premier temps les données d'autorité de la BnF et de l'Abes et offrant à ses utilisateurs une base de coproduction mutualisée.

Gabriel Naudé disait également : « *car de s'imaginer qu'il faille après tant de peine et de despense cacher toutes ces lumières sous le boisseau, et condamner tant de braves esprits à un perpétuel silence et solitude, c'est mal reconnoître le but d'une bibliothèque* »<sup>5</sup>. Les « lumières » sont les données ; les « braves esprits » les catalogueurs qui les produisent. Dans le nouvel écosystème international des données,



↑ **Robert Marcaire Médecin**, lithographie coloriée d'Honoré Daumier. Du temps de Naudé et jusqu'à récemment, le lecteur avait besoin du bibliothécaire comme le malade du médecin. Changement de paradigme !

la généralisation et la mutualisation des référentiels, par leur inscription dans le web de données et dans le Big Data, est un outil pour remplir cet objectif.

**VINCENT BOULET**

*Chef du service des référentiels, BnF  
vincent.boulet@bnf.fr*



### POUR EN SAVOIR PLUS

- Définition et estimations quantitatives du Big Data : [lebigdata.fr/definition-big-data](http://lebigdata.fr/definition-big-data)

- Sur le site de la Fédération internationale des associations et institutions de bibliothèques (IFLA) se trouvent de nombreux documents utiles pour comprendre le changement de paradigme.

La définition du web de données, avec des exemples concrets d'application dans le monde des bibliothèques, a été étudié au cours de la conférence « *Linked Data in Libraries: Let's make it happen!* » : <http://ifla2014-satdata.bnf.fr/program.html>

Sur le concept de contrôle bibliographique universel, voir l'article de Françoise Bourdon et Gildas Illien « *À la recherche du temps perdu, retour vers le futur : CBU 2.0* » <http://library.ifla.org/956/1/086-illien-fr.pdf>

En réponse à trois questions centrales du dossier, des experts de l'Abes livrent ici leurs réflexions, issues de leur expérience quotidienne. François Mistral est responsable d'IdRef, Olivier Rousseaux est chef du service Métadonnées, Yann Nicolas, expert métadonnées, et Philippe Le Pape, chargé de mission Normalisation.

# 3 fois sur le métier remettons les notions

## 1. Quelle distinction faites-vous entre « référentiel » et « fichier d'autorité » ?

**OLIVIER ROUSSEAU :** De « référentiel », le Larousse donne la définition suivante : « ensemble auquel doivent appartenir les éléments, les solutions d'un problème posé ». Du fait qu'ils organisent et contraignent les métadonnées produites par les réseaux de signalement animés par l'Abes, les fichiers d'autorité appartiennent bien à un ensemble : fonctionnant comme un tout, avec des règles d'interdépendance et d'interopérabilité de ses constituants, ils permettent le contrôle de la cohérence des métadonnées bibliographiques.

**PHILIPPE LE PAPE :** Les deux termes s'appliquent à des données « de référence », statut consacré soit par une labellisation, soit par l'usage. On pourrait nommer « référentiel » un ensemble de « données outils » (modèles, règles de catalogage, formats, vocabulaires normalisés...) auquel on recourt pour garantir qualité et interopérabilité des métadonnées qu'on produit et/ou administre. Et on nommerait « données d'autorité » celles des métadonnées qu'on produit ou/et administre qui ont vocation à faire référence à leur tour. Dans le Sudoc, et plus encore

dans theses.fr, cette définition s'applique aussi (à mon sens) aux données « bibliographiques ».

**FRANCOIS MISTRAL :** De « référentiel », je retiens l'idée d'un jeu de données suffisamment fiables pour être utilisées en confiance et en produire ou en agréger d'autres. De « fichier d'autorité », je retiens la double dimension de contrôle et de légitimité à assurer ce contrôle, par une pratique fondamentale : la mention explicite des sources. Ce point établi, je voudrais surtout souligner leur horizon de convergence : la qualité. Les bibliothécaires produisent ainsi des données d'autorité qui ont toute légitimité à devenir des référentiels.

**YANN NICOLAS :** Les référentiels des uns sont les données centrales des autres : si, à l'Abes, par « référentiel », on entend « données permettant de décrire les documents en minimisant le recours à du texte libre », pour un autre système d'information, nos documents correspondent à des entités secondaires... On observe également la tendance à transformer les référentiels « listes fermées » en référentiels « entités identifiées » : on passe de « littéraux » (chaînes de caractères simples ou typées) à des « ressources » pouvant elles-mêmes comporter des attributs, être décrites. Connecter nos données à ces référentiels les enrichit, par transitivité, de beaucoup d'autres informations. Autre tendance observée : nos autorités traditionnelles se normalisent, rejoignent d'autres référentiels. Cette normalisation est à la fois intellectuelle et technique. Le paradigme « web sémantique » est le vecteur principal de cette normalisation : tout devient ressource, identifiée de manière univoque à l'échelle universelle grâce à sa (ou ses !) URI, et ce sont les triplets RDF qui en parlent le mieux...

## 2. Comment la montée en puissance du rôle des référentiels impacte-t-elle votre travail et les pratiques de catalogage des professionnels des réseaux ?

**O. R. :** L'Abes est confrontée à ces questions pour avoir posé comme principe, dès la mise en place du Sudoc, la réutilisation de référentiels existants : Rameau ou FMeSH pour les accès matière, fichiers d'autorité de la BnF pour les accès auteur ou titre uniforme ; codes ISO ou Unimarc. Aujourd'hui, la



réflexion porte sur la manière d'appréhender les métadonnées de référentiels tiers (Sirene, RNSR... au niveau national; VIAF, au niveau international). Pour le catalogage au quotidien, l'expansion des référentiels offre des opportunités d'améliorer les méthodes de liage automatique, ou de développer des outils d'aide à la décision.

**P. LP :** Dans le Sudoc, la normalisation (et l'emploi de référentiels) a toujours été cruciale, et avec les règles RDA, le recours à des vocabulaires contrôlés (type de contenu, de médiation, de support...) s'accroît encore.

**F. M. :** Ma démarche au poste de responsable IdRef se déroule selon trois axes : le réseau des catalogueurs (dont les Correspondants autorité); l'outillage visant à accroître leur maîtrise de la production des données d'autorité; et la dissémination multicanaux, multiformats, dans et hors les murs, de ces données. Il s'agit de rendre plus attractifs les données mais aussi les services associés, notamment les alignements<sup>1</sup> afin d'en faire un levier pour démarcher des partenaires et intégrer de nouveaux projets, l'un des apports « spécifiquement Abes » étant la mutualisation.

**Y. N. :** Je ne pense qu'à ça ! Ces dernières années, nous avons travaillé sur les projets Qualinca et Hub de métadonnées. Qualinca produit des algorithmes qui auscultent et améliorent la qualité des liens entre notices bibliographiques et notices d'autorité. Entre Sudoc et IdRef, par exemple. Côté Hub de métadonnées, nous récupérons des données en provenance des éditeurs, nous les homogénéisons et enrichissons grâce aux référentiels de toutes sortes (langues, auteurs, sujets, types de document...). Nous remplaçons (ou complétons) des mots par des identifiants : un nom d'auteur par un lien vers une URI IdRef, ISNI, Orcid, etc. Dans les deux cas, nous connectons l'information bibliographique à des référentiels, pour mieux la structurer et la connecter à l'échelle du web et pas seulement à un catalogue, fût-il national !

### **3. Comment imaginez-vous le rôle des référentiels dans le paysage de l'IST voire au-delà ?**

**O. R. :** Je ne le vois pas évoluer radicalement dans l'immédiat. La tendance est aux rapprochements de référentiels existants. Mais un référentiel tiers est à aborder avec prudence, et chaque fois, les mêmes questions devront être examinées : à quels objectifs, quelles contraintes répond-il ? Sur quel modèle de données est-il fondé ? Quelles en sont les règles d'alimentation ? Couvre-t-il nos besoins en granularité des données et traçabilité des futures évolutions ? Quelle gouvernance technique et scientifique en résultera-t-il, et sera-t-elle adaptée à notre contexte ? Un exemple : le projet de FNE amorcé en mars 2017 entre la BnF et l'Abes répond à ces questions en recherchant une solution de fusion des

fichiers d'autorité existant de part et d'autre au profit d'un fichier national unique géré en coproduction.

**P. LP :** On va vers une importance grandissante des identifiants de confiance dans lesquels le « nom », la « forme d'autorité », les données elles-mêmes se trouvent de plus en plus ramassés : le passage des métadonnées bibliographiques de systèmes fermés au Web renforce la nécessité de les normaliser et de les étiqueter selon des systèmes d'identification qui, pour être efficaces, doivent jouir d'une large reconnaissance.

**F. M. :** Les données produites par les bibliothécaires sont promises à un grand avenir, l'enjeu de l'information se situant dans sa structuration. Avec un point délicat : les catalogueurs sont à la fois acteurs et spectateurs aux premières loges de ce phénomène. Un accompagnement est à inventer... On imaginerait volontiers les référentiels jouer un rôle structurant pour l'IST. Ils pourraient amener une reconfiguration des missions des opérateurs nationaux, façonnée à leur image : tout entière de spécialisation et de coopération, pour un service rendu à la communauté ESR (et au-delà) de haut niveau.

**Y. N. :** Vu de l'Abes, en caressant du regard le paysage un peu cacophonique de l'IST en France, j'espère une politique publique des référentiels claire. Que chacun joue sa partition, maintienne et mette à disposition les référentiels de son ressort. Qu'on évite les doublons ! Par exemple, que le RNSR soit, de droit et de fait, reconnu comme LE service public national qui fournit identifiants et attributs de référence pour les laboratoires français. Ce qui n'empêche pas des clients comme STAR et theses.fr de gérer leurs propres attributs complémentaires, en sus des attributs RNSR, à des fins propres, bibliographiques ou pas. En revanche, accrochons nos attributs aux identifiants RNSR ! Même chose pour les autorités de type Entreprise (avec Sirene, de l'INSEE). L'Abes doit être un bon client des référentiels des autres, et un bon fournisseur de référentiels : son positionnement, son organisation et son capital de données la rendent légitime. C'est déjà le cas pour le référentiel des thèses françaises ou celui des chercheurs français.

**[1]** L'alignement, auquel nous nous référons à de multiples reprises et qui est illustré dans les exemples du dossier, consiste à établir un lien, terme à terme, entre les entités de deux (ou plusieurs) bases, au moyen d'un identifiant unique et pérenne, de façon manuelle ou automatique.



**POUR EN SAVOIR PLUS**

**Les réponses publiées ici ne sont qu'un bref résumé des réflexions de leurs auteurs. Retrouvez-les *in extenso* et accompagnées de nombreux liens sur <https://pункtokomo.abes.fr/2017/04/20/autorites-vs-referentiels-3-questions-aux-experts-de-labes/>**

Fidèle à sa mission initiale – constituer une base de catalogage commune aux bibliothèques universitaires –, l'Abes, tout en diversifiant ses activités, a conçu IdRef. Basé sur les autorités Sudoc, IdRef propulse les autorités pour un réseau de plus en plus élargi de producteurs et d'utilisateurs.

# IdRef, les autorités en conquête et en partage

À l'origine, l'Abes a été créée afin de constituer le Sudoc en tant que « coopérative de catalogage » des bibliothèques des universités françaises. Depuis cette initiative fondatrice, les activités de l'Agence se sont diversifiées et ont donné lieu à la conception de nouveaux catalogues et bases de données à contenu spécifique : thèses électroniques (2006), archives et manuscrits (2007), acquisition de masse d'archives d'éditeurs scientifiques (2011). Afin de contrebalancer cette stratégie de diversification, contraire à la vision initiale d'une base de données Sudoc exclusive et complète, l'Abes a choisi de miser sur le pouvoir des données d'autorité. Il a donc été décidé de leur accorder leur autonomie. Les autorités Sudoc sont bien vivantes, IdRef est là pour le dire haut et fort !

## ET LE SUDOC ENFANTE IDREF...

Conçue nativement de manière générique et ouverte de sorte que n'importe quel partenaire puisse y accéder en lecture et en écriture, IdRef met à la disposition des catalogueurs une interface web qui permet de lier manuellement ou de créer de nouvelles autorités à partir de leur propre environnement. Ce dispositif d'interopérabilité de type API est le premier d'une liste plus conséquente, qui fait d'IdRef un objet à facettes et à façon :

- base vivante et administrée de notices d'autorité identifiées de façon unique et pérenne ;
- site web pour rechercher, consulter et rebondir vers des catalogues documentaires ;
- outil professionnel *full web* pour produire des données d'autorité ;
- portail d'accès aux données ouvertes sous licence Etalab et réutilisables par micro-webservices, requêttables par le moteur de recherche Solr et moissonnables en OAI-PMH ;
- « interprète » des données en de multiples formats standards, pour les humains comme pour les machines (html, xml, rdf, json).

Rappelons également que, de par sa conception, IdRef agit pour le web de données : voici maintenant huit ans que les données d'autorité sont exposées au format RDF, à la volée pour une parfaite fraîcheur. En 2017, la perspective de disposer d'un *triple store*<sup>1</sup> (Sparql End Point) est proche, ce qui renforcera sa contribution au web de données et ses relations « sémantiques » avec d'autres institutions amies, comme Persée (data.persee.fr) ou la BnF (data.bnf.fr).

## ... QUI RALLIE DE NOUVEAUX PARTENAIRES...

Le dynamisme des réseaux Sudoc, Thèses et Calames attire vers IdRef de nouveaux partenaires aux profils divers et aux finalités variées, parfois « exotiques ». L'ensemble dessine une entreprise intégratrice, un projet fédérateur. Dans un paysage de l'IST pouvant paraître fragmenté, il existe donc aussi des formes unificatrices, des forces capables de préserver la diversité en rassemblant.

Ainsi se côtoient : Persée, portail d'accès libre et gratuit à des revues et des publications scientifiques numérisées ; Adum, réseau Accès doctorat unique et mutualisé de gestion des écoles doctorales ; le Larhra (laboratoire de recherche historique Rhône-Alpes) de l'UMR 5190, spécialisé en histoire moderne et contemporaine et particulièrement investi dans le champ des humanités numériques ; la plate-forme ORI-OAI de gestion de la production numérique institutionnelle ; Prelib, projet de recherche en littérature de langue bretonne de la MSH Bretagne. Au-delà, IdRef entend devenir la principale source d'identification des chercheurs français. À ce titre, IdRef apporte une contribution tout à fait particulière au projet Conditor (voir p. 12). De même, la coopération avec des projets orientés gestion de la recherche, comme Caplab (voir p. 14), ouvre de nouveaux champs de coopération interinstitutionnelle et de nouveaux débouchés pour la production de métadonnées des différents réseaux.

## ... ET S'APPUIE SUR SES CATALOGUEURS EXPERTS...

Ces nouvelles coopérations renforcent l'esprit qui préside au projet IdRef : la mutualisation, la qualité des données, l'expertise des producteurs, l'accompagnement réseau. L'activité de catalogage des auto-

→ Un nouveau logo pour IdRef !  
Les établissements qui réutilisent les données IdRef sont invités à apposer ce logo sur leur site.





rités qui, pendant de nombreuses années, est restée dans les coulisses du Sudoc ou à l'ombre des BU, est au cœur des intentions d'IdRef : son utilité et le savoir-faire qu'elle nécessite sont établis bien au-delà de notre secteur professionnel traditionnel. En outre, l'accroissement du nombre de contributeurs et la réutilisation des données dans des contextes ou par des utilisateurs diversifiés sont autant d'opportunités d'amélioration et d'enrichissement. Pensons, par exemple, à la réduction des micro-erreurs ou à l'agrégation de contenus qui engendre de facto un enrichissement de l'information : il y va de la consolidation du signalement, qui profite autant aux professionnels qu'aux utilisateurs finaux.

Le renouveau de la fonction Correspondant autorité, interlocuteur de l'Abes sur le terrain, est emblématique de cette situation. Cet expert ès données d'autorité et problématiques de catalogage voit sa fonction traditionnelle revisitée à l'aune des enjeux scientifiques et politiques en matière de visibilité de la production scientifique. Les fonctions de ce professionnel, compétent sur les questions d'identification des auteurs et d'attribution fidèle de leurs publications, tendent à se déplacer vers les services à la recherche des établissements universitaires : archives institutionnelles, pages de chercheurs, applications bibliométriques...

Accompagner ces changements fait donc partie intégrante du projet IdRef, qu'il s'agisse de l'animation du réseau des Correspondants autorité sur ces problématiques professionnelles bien spécifiques, ou du soutien et de la promotion de leur implication dans les projets innovants de type « service à la recherche ». Deux exemples éloquentes : à Lille 2, où le Correspondant autorité est l'administrateur Sampra, application à vocation bibliométrique ; à l'INP de Toulouse, où le Correspondant autorité est le gestionnaire de l'archive ouverte OATAO.

### ... POUR ÉLARGIR LE CERCLE DE LA MUTUALISATION

L'interopérabilité est un élément central de la démarche opérationnelle d'IdRef. Désireux de donner des « preuves de concept », IdRef a procédé au cours de l'année 2016 à une intégration des références bibliographiques des catalogues Calames, theses.fr et HAL. Il s'agissait de montrer la fonction de pivot des identifiants et les bénéfices d'un adossement des catalogues à un référentiel commun.

Ouvrir grand l'accès à l'interopérabilité est devenu un impératif pour nos partenaires. À l'aide des puissants algorithmes d'alignement développés dans le cadre du projet de recherche Qualinca, IdRef offre un service capable d'identifier les personnes, notamment pour la communauté ESR. C'est ainsi qu'en 2016 ont été identifiés les publiants dans plusieurs archives institutionnelles (dont Spire de Sciences Po, SAM de l'Ensam ou Okina de l'Université d'Angers), des

enseignants-chercheurs/chercheurs d'établissements universitaires (dont Toulouse 2, l'INSA de Rouen ou l'Université de Lorraine) ou disposant d'un identifiant dans HAL (IdHAL).

Redistribuer l'ensemble de ces alignements constitue la prochaine étape du projet IdRef. Une démarche ouverte, globale et coordonnée semble en effet le meilleur moyen d'éviter la création de monopoles ou une situation de concurrence sauvage entre référentiels. Ainsi passons-nous par la construction de passerelles avec les acteurs internationaux. Dans cette optique, IdRef est partie prenante de trois projets majeurs en matière d'identification : VIAF, ISNI (voir encadré ci-dessous) et Orcid.

Entérinée en mai 2016 par la signature d'un Memorandum of Understanding, la coopération avec Orcid vise à plus d'interopérabilité entre les deux applications. Dans le même temps, des opérations d'alignement IdRef-Orcid ont été programmées pour l'identification des personnes présentes dans chacun des référentiels et, corrélativement, pour le rapprochement des identités et productions propres à chacun des environnements.

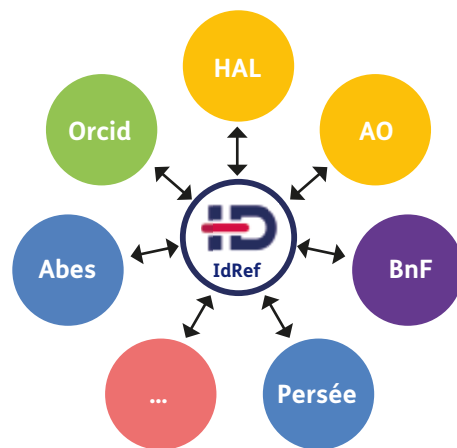
IdRef est équipé de tout l'outillage nécessaire pour être un hub multi-applicatifs à la disposition des porteurs de projet, des SI d'établissements et opérateurs de l'IST.

**FRANÇOIS MISTRAL**

Responsable IdRef, Abes  
mistralf@abes.fr

**YANN NICOLAS**

Expert métadonnées, Abes  
nicolas@abes.fr



➔ IdRef et ses partenaires, une entreprise fédératrice.

[1] Un *triple store* permet de stocker et de récupérer un très grand nombre de données RDF (*Resource Description Framework*). Celles-ci sont stockées sous la forme de triplets associant un « sujet » (la ressource à décrire, un « prédicat » (un type de propriété) et un « objet » (une donnée ou une autre ressource, la valeur de la propriété), identifiés notamment par une URI.

## ● ● ● IDREF – FACTS AND FIGURES

**D**ébut avril 2017, IdRef comptait près de **3,4 millions de notices d'autorité**, dont 2,69 millions de notices Personnes, 307 000 notices de Collectivités, 77 000 notices de Lieux géographiques ; ainsi que deux référentiels d'indexation matière : Rameau et FMeSH.

Au cours des cinq dernières années, plus de 180 000 notices d'autorité sont créées en moyenne par an : le dynamisme des contributeurs à IdRef, catalogueurs des bibliothèques et gestionnaires des thèses, n'est plus à prouver !

À l'appui d'IdRef vient un réseau de **180 Correspondants autorité**, experts et interlocuteurs dédiés aux autorités.

- Dissémination des autorités Personnes :
- 2 600 000 notices dans VIAF
  - 1 900 000 d'identifiants ISNI assignés
  - 328 138 éléments avec la propriété « IdRef »
  - 280 000 notices liées dans theses.fr
  - 60 000 notices liées dans Calames
  - 50 000 notices liées dans Persée
  - 750 000 notices alignées avec celles de la BnF

## ● ● ● QUALINCA ET IDREF : L'INTÉGRATION EST EN COURS !

**Disposer de descriptions d'entités fiables et uniques, les associer sans erreur et en masse à des ressources de tout type : tels sont les enjeux de Qualinca. Bilan des travaux menés à l'Abes.**

À l'Abes, l'automatisation du liage des ressources documentaires aux autorités n'est pas un sujet nouveau. Explorée dès 2010 dans le cadre de SudocAD, qui offrait une entrée en matière aux résultats prometteurs, cette question a été approfondie grâce à la participation à Qualinca (« qualité et interopérabilité de grands catalogues documentaires »), projet Agence nationale de la recherche initié en 2012 et officiellement terminé depuis septembre 2016. Si le cheminement théorique fut long et sinueux, il se révèle aujourd'hui payant. Les premiers fruits sont mûrs pour la récolte, gageons que ce ne seront pas les derniers ! Aujourd'hui, les briques fonctionnelles issues du projet Qualinca se répartissent autour de trois axes de travail : alignement, diagnostic qualité et production assistée.

Le premier axe concerne **l'alignement** : grâce à un programme dédié, il est désormais possible d'aligner sur IdRef – ainsi que sur d'autres référentiels tels ISNI, VIAF, Orcid – des entités « personnes » issues de différentes sources (catalogues documentaires, archives institutionnelles, entrepôts OAI-PMH, bases d'articles, annuaires, référentiels eux-mêmes...). À ce jour, l'alignement à IdRef de 50000 autorités Persée constitue le plus bel exemple.

### VALIDER ET CORRIGER

Le deuxième axe de travail concerne le **diagnostic qualité**. L'objectif de multiplier les liens ne suffit pas, il est primordial de s'assurer de leur fiabilité. En matière de processus automatisés, le principe de prudence est bien sûr requis. Ainsi, chaque nouvel algorithme est testé en confrontant ses résultats à des « vérités terrain », autrement dit à des alignements déjà existants et validés par des experts humains. Néanmoins, que faire du rétrospectif ? Le fait que nos bases soient truffées d'anomalies (liens erronés, doublons, autorités mêlées) risque d'engendrer de la confusion pour les algorithmes qui se nourrissent de ces liens. Les conséquences ? Au mieux, moins

de liens créés ; au pire, malgré la plus grande prudence, la création de nouveaux liens erronés.

Comment sortir de ce dilemme ? La collaboration étroite avec des chercheurs de l'équipe GraphIK de Montpellier<sup>1</sup> a guidé l'Abes vers la résolution du problème. Il en a résulté SudoQual, un outil de validation et de correction des liens préexistants entre le catalogue Sudoc et IdRef. Produit d'une réflexion scientifique exigeante, SudoQual est un programme sophistiqué, au paramétrage complexe ; bien qu'il s'agisse encore d'un prototype, il est possible d'en exploiter certaines « pièces détachées », comme par exemple, un petit programme dérivé capable de détecter des anomalies grâce à deux règles logiques simples qui peuvent s'énoncer ainsi : « *Il existe une anomalie si : 1) au moins deux autorités distinctes ont des ressources liées dont les titres sont identiques ; 2) au moins deux autorités distinctes ont des ressources liées dont les co-contributeurs sont identiques.* »

### CONTRÔLER ET ASSISTER

La **production assistée** constitue le troisième axe de travail, considéré comme le volet le plus important : comment donner aux catalogueurs des moyens de contrôle et d'assistance à la mesure des enjeux actuels (gestion de la masse, qualité, exhaustivité) ? À partir d'une interface imaginée initialement pour visualiser les diagnostics produits par SudoQual, il a

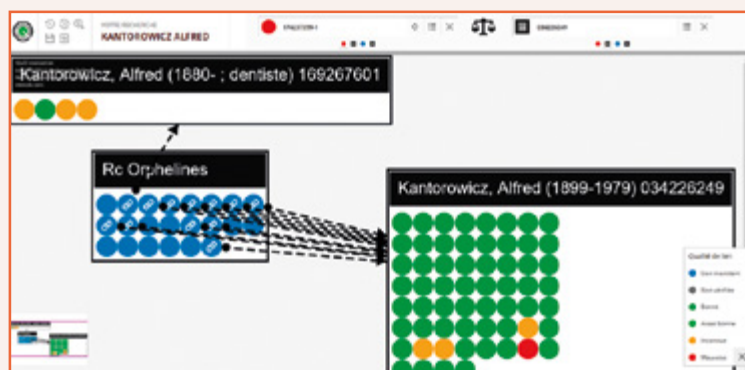
été décidé d'aller plus loin. À la représentation visuelle et synthétique des autorités enrichies de leurs liens bibliographiques grâce à IdRef, un principe essentiel a été ajouté : l'interactivité. En effet, l'utilisateur doit rester acteur du système, ses connaissances dépassant et complétant celles du programme informatique. Il en résulte un outil d'aide à la décision, conçu pour faciliter le contrôle qualité et fiabiliser la production des liens.

Reste à présent à définir les scénarios concrets d'intégration de cet outil dans les environnements de travail quotidien. Au vu de ces résultats, le projet Qualinca a déjà servi les objectifs portés par IdRef. En se concentrant sur les entités personnes, il a mis au point, éprouvé et validé des méthodes génériques qui pourront s'appliquer à d'autres types d'entités, comme les collectivités ou les œuvres. Au-delà du liage, des perspectives ont été ouvertes : création, fusion, scission d'autorités. De son côté, le réseau IdRef est plus mûr, renforcé, mieux armé qu'en 2012. Il semble désormais naturel de poursuivre la convergence et de promouvoir l'intégration de Qualinca et d'IdRef... ou l'intégration de Qualinca dans IdRef.

**ALINE LE PROVOST**

Expert métadonnées - équipe Qualinca  
le-provost@abes.fr

[1] GraphIK (Graphs for Inferences on Knowledge) est une équipe commune de l'Inria (Sophia Antipolis), de l'Université de Montpellier, du CNRS et de l'INRA <https://team.inria.fr/graphik/>



➔ Diagnostic posé par SudoQual pour un sous-ensemble circonscrit par l'appellation « Kantorowicz, Alfred » : les billes sont des points d'accès dans les notices du Sudoc, les boîtes sont des autorités IdRef. Catalogueurs, à vous de jouer !

# Nous avons les coffres, *il nous faut les bonnes clés !*

À la Direction de la documentation de l'Université de Bordeaux, Jérôme Poumeyrol est impliqué dans différents projets (open access, bibliométrie) qui intègrent les référentiels pour identifier les auteurs de publications scientifiques et les structures de recherche. Membre du CA de l'ADBU, il anime la commission Recherche et documentation de l'association.

Dans l'usage quotidien des professionnels de l'IST, la distinction entre « référentiel » et « base d'autorité » est, me semble-t-il, de moins en moins tranchée. Sans doute les projets réalisés par les bibliothèques avec d'autres partenaires (notamment nos collègues des directions des systèmes d'information en université) nous amènent-ils à utiliser de plus en plus fréquemment des termes propres aux systèmes d'information. Ainsi, « référentiels » en lieu et place de « bases d'autorité » et de « notices d'autorité ». Cependant, il nous faut parfois revenir aux fondamentaux de la profession pour expliciter à nos interlocuteurs la valeur ajoutée de la production intellectuelle de bases et notices d'autorité.

La différence qui demeure, à mon sens, est celle de la richesse de l'information qui est rassemblée dans ces dernières. S'y ajoute la notion de « certification » d'une information créée par un professionnel. Celle-ci ne peut être proposée avec autant de qualité par les référentiels – d'auteurs ou de structures de recherche, par exemple – des bases de données citationnelles commerciales. Mais, plus que celle de la distinction, la vraie question est peut-être celle de la connexion entre référentiels et bases d'autorité. Nous avons les coffres, mais il faut les bonnes clés !

## LAISSER LA PLACE À UNE GÉNÉRATION SPONTANÉE ?

Les grands éditeurs scientifiques et les fournisseurs internationaux de données informationnelles n'ont pas attendu pour créer leurs référentiels – signe évident de leur importance dans l'économie de la connaissance. Le problème est celui de leur multiplicité : quelle est la valeur d'un référentiel si l'on en compte plusieurs pour un même objet ? Lier un ensemble d'objets, d'entités ou de productions en permettant la traçabilité est fondamental et, pour cela, le référentiel s'impose ; mais doit-on laisser la place à une sorte de « génération spontanée » ?

L'exemple des identifiants (ID) auteurs est éclairant : après les ID de producteur commerciaux de bases citationnelles (ResearchID, ScopusID), ceux d'archive ouverte nationale (IdHAL), c'est maintenant le phénomène Orcid. L'adhésion massive à cette proposition d'ID montre qu'elle répond bien à un besoin de la communauté des chercheurs et des éditeurs. L'alimentation et la gestion par le chercheur lui-même avec une ergonomie efficace est intéressante, mais elle pose aussi des questions.

La puissance publique a donc un rôle majeur à jouer en ce domaine hautement stratégique. Identifiants et référentiels sont (parmi d'autres) les leviers de la mise en œuvre du web sémantique, car c'est sur leur qualité que s'appuiera l'exploitation publique et commerciale des données. L'origine « certifiée » et la garantie sur la construction de la donnée passent donc par des organismes publics et des standards internationaux.

## UNE DOUBLE ATTENTE QUANT AU RÔLE DE L'ABES

Pour les établissements de l'ESR, l'attente vis-à-vis du rôle de l'Abes est double. D'une part, il s'agit de consolider le travail réalisé depuis près de vingt ans par le réseau des bibliothèques membres du Sudoc : un travail collectif qui a permis de constituer un ensemble de notices d'autorité à la fiabilité reconnue. Ce rôle de coordinateur et de garant de la qualité de l'alimentation des bases reste incontournable, parallèlement à l'adaptation aux évolutions du web et l'accompagnement du réseau dans ces changements.

D'autre part, forte de cet historique, l'Abes doit continuer, par sa capacité d'innovation, à faire fructifier ce gisement informationnel.



© Geralt/Fixabay (CCO)

L'exemple d'IdRef est intéressant : cette application progresse par cercles concentriques. Les liens qui se développent avec Orcid vont, à mon avis, dans le bon sens : des autorités alimentées en proximité par des professionnels et certifiées par un opérateur national mais avec une visibilité internationale et d'autres interconnexions proposées par une « non profit organization ».

Les structures documentaires de l'ESR qui participent au Sudoc ne peuvent, je pense, que souhaiter voir l'Abes jouer pleinement son rôle dans ce fameux et tant attendu « alignement des référentiels ». Cela suppose bien évidemment de se rassembler et d'être cohérent au niveau national avant de nouer des partenariats avec d'autres acteurs internationaux. Les évolutions et discussions en cours prennent cette direction, ce qui laisse entrevoir des résultats positifs dans les prochains mois.

**JÉRÔME POMEYROL**

Responsable du service Soutien à la recherche  
Direction de la documentation  
Université de Bordeaux  
jerome.poumeyrol@u-bordeaux.fr



## PERSÉE, partenaire confirmé

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

**Infrastructure de recherche associée au dispositif CollEx, Persée développe et maintient une plateforme dédiée à la production et la diffusion de collections et de corpus documentaires numériques élaborés à partir de fonds patrimoniaux.**

Au-delà de la mise à disposition de documents scientifiques, Persée vise à les intégrer pleinement dans un ensemble de connaissances (désormais mondialisé) en les rendant interopérables. Cela ne peut se faire qu'en mettant en œuvre des formats et des référentiels normalisés.

Persée n'a pas vocation à définir et maintenir de tels référentiels, aussi nous appuyons-nous sur des « produits » mis à disposition par des institutions partenaires. S'agissant des auteurs, parmi les référentiels envisageables, nous avons choisi celui proposé par l'Abes, IdRef, pour plusieurs raisons :

- son périmètre – issu du Sudoc, il couvre la plupart des auteurs scientifiques français

sur une très large période ;

- son ouverture – les données d'IdRef sont réutilisables, leur production et leur curation s'inscrivent dans un contexte international ;
- ses services – les différents outils – site web, API – facilitent l'exploitation et la mise à jour des données ;
- son dynamisme – IdRef bénéficie d'évolutions régulières, et son équipe est à l'écoute et réactive.

Depuis mars 2016, les liens entre les auteurs de Persée et IdRef sont établis soit au fil de la production – tout auteur « rencontré » au cours du processus de documentation est lié à une autorité IdRef et, si celle-ci n'existe pas, Persée peut la créer dans IdRef ; soit *a posteriori*, par un alignement automatique effectué grâce aux outils développés par l'Abes (Qualinca) ou par un documentaliste de l'équipe Persée ayant la possibilité d'alimenter IdRef (ajouts, compléments d'information, corrections, etc.).



En mars 2017, sur les 115 000 auteurs de Persée, la moitié sont liés à des autorités d'IdRef, ce qui permet dans un second temps de déduire de nouveaux liens vers d'autres référentiels (data.bnf.fr, VIAF, ISNI, dbPedia, Wikipedia, RePEc, etc.).

Outre la mise à disposition de données plus riches, d'outils de recherche, de parcours et de rebonds étoffés, les gestionnaires des différents systèmes d'information impliqués bénéficient d'une amélioration significative de la qualité des données primaires et des procédures automatisées de collecte d'informations complémentaires.

**VIVIANE BOULÉTREAU**

Responsable du pôle informatique et développements  
UMS Persée

(Université de Lyon – ENS de Lyon – CNRS)  
[viviane.bouletreau@persee.fr](mailto:viviane.bouletreau@persee.fr)

## CONDITOR, le futur pot commun de métadonnées

[www.bibliothequescientifiquenumerique.fr/conditor](http://www.bibliothequescientifiquenumerique.fr/conditor)

**Lancé officiellement le 5 décembre 2016, Conditor est un projet multipartenarial<sup>1</sup> porté par la DIST-CNRS et soutenu par le MENESR.** Sa vocation : mettre en place un référentiel national de la production scientifique de la recherche publique française comportant :

- un « pot commun » de métadonnées alimenté par des sources multiples en provenance de l'enseignement supérieur et de la recherche (HAL, archives institutionnelles, bases bibliométriques, systèmes d'information recherche...), de sources internationales et de référentiels choisis collégialement ;
- un outillage adapté pour constituer ce réservoir et diffuser les métadonnées – validées ou à valider – vers les systèmes d'information de l'ESR.

En tant que référentiel national, Conditor se doit de garantir fiabilité et interopérabilité

sémantique entre ces sources bibliographiques diverses. Pour ce faire, il s'appuiera sur des référentiels communs adaptés et reconnus. Pour identifier les structures, le Répertoire national des structures de recherche (RNSR) sera utilisé. Pour identifier les auteurs, Conditor s'appuiera sur le référentiel IdRef porté par l'Abes. En effet, outre la richesse des données d'autorité produites par les réseaux de signalement (Sudoc, Thèses, Calames), les identifiants IdRef, grâce aux puissants algorithmes d'alignement développés au sein du projet de recherche Qualinca, sont alignés sur les identifiants IdHAL et Orcid, ce qui participe à l'objectif global d'enrichissement des corpus. Les apports prévus sont multiples : qualité et complétude des réservoirs institutionnels par le croisement de différentes sources, identification pérenne des acteurs (personnes

et structures), détection exhaustive de la production nationale, fiabilité des indicateurs de la science, facilitation du dépôt du texte intégral et des activités de gestion de la recherche pour les chercheurs...

D'ici deux ans, l'outillage nécessaire ainsi qu'un premier réservoir de métadonnées seront à disposition. Objectif à cinq ans : la mutualisation complète des travaux de référencement et l'intégration aboutie au sein des systèmes d'information.

**ANNIE CORET**

Responsable IST DIST/CNRS  
[annie.coret@cnrs-dir.fr](mailto:annie.coret@cnrs-dir.fr)

**[1]** Partenaires du projet : Abes, Amue, CCSD, Inist-CNRS, InSHS, Service d'appui à la politique et à la prospective scientifiques du CNRS, Huma-Num, Inra, Inria, Irstea, MENESR (équipes ScanR et RNSR), OST du HCERES, universités de Bordeaux, Paris-Dauphine, Paris-Diderot et UPMC.

# Ils utilisent IdRef... Ils utilisent IdRef...



## ADUM, plus de visibilité pour les doctorants

[www.adum.fr](http://www.adum.fr)

**Adum (Accès doctorat unique et mutualisé) est un outil collaboratif à la disposition des écoles doctorales, établissements, unités de recherche, bibliothèques, et des chercheurs, doctorants et docteurs.** Il réunit dans un annuaire en ligne les profils de compétences des doctorants et docteurs.

Les données sont sous la responsabilité des écoles doctorales et des établissements.

Quant à l'outil, il est développé par l'association Contact.

La base Adum compte 26 708 doctorants, 36 968 chercheurs/enseignants-chercheurs et plus de 50 000 docteurs issus des 135 écoles doctorales qui l'utilisent.

L'Abes et Contact collaborent depuis 2007 sur les échanges avec les applications STEP et STAR pour le signalement des thèses en cours et des thèses soutenues suivant les

recommandations TEF (thèses électroniques françaises). Nous avons aligné le référentiel « Personnes » avec IdRef. Ainsi, lors des flux d'échanges de métadonnées vers STEP et STAR, celles-ci sont enrichies par l'identifiant PPN IdRef.

En 2015, Adum a « connecté » IdRef dans ses programmes. Au vu du grand nombre de chercheurs dans Adum, nous avons souhaité aligner nos deux bases ; aussi avons-nous choisi une solution de *back-office* utilisant les algorithmes développés par l'Abes. L'alignement des auteurs IdRef avec ceux identifiés dans Adum forme des « couples » qui rendent interopérables nos systèmes. De plus, il permet de détecter des anomalies et consolider les corpus.

À partir d'exports d'Adum, l'Abes a effectué un premier travail d'alignement des données avec des résultats très encourageants, puisque positifs dans 1 567 cas sur un

échantillon de 1 578 chercheurs. Importé dans Adum et dans les flux de données vers STAR et theses.fr, cet alignement établit un lien direct entre les thèses déclarées et les encadrants. Nous l'avons depuis étendu à l'ensemble des encadrants de thèses « actifs » au cours des cinq dernières années, ainsi qu'aux docteurs qui viennent de soutenir leur thèse présents dans Adum et theses.fr. Les profils personnalisés des docteurs Adum sont donc maintenant connectables aux pages personnelles des mêmes docteurs de theses.fr. Jeu gagnant-gagnant pour nos institutions, l'opération est également une amélioration du service rendu à nos communautés.

**CATHERINE MORALES**

Association Contact – Adum  
Université de Montpellier  
[cathy@adum.fr](mailto:cathy@adum.fr)



## AUREHAL et son IdHAL rassembleur

<https://hal.archives-ouvertes.fr/page/mon-idhal>

**AureHAL est l'application de gestion des six référentiels utilisés dans HAL : domaines, structures, revues, projets ANR, projets européens et auteurs.** Le référentiel « auteurs » est formé par l'ensemble des formes auteur présentes dans les dépôts HAL. Une forme auteur est identifiée par nom, prénom, e-mail, URL de site, établissement d'appartenance et identifiant IdHAL qui peut lui être attaché. Une même personne peut disposer de plusieurs formes auteur (nom-prénom, nom-initiale du prénom), deux personnes pouvant posséder une même forme auteur (homonymes ou mêmes nom et initiale de prénom). Une forme auteur n'est donc pas un identifiant d'auteur.

Pour lier un compte utilisateur à un auteur, HAL utilise IdHAL : un chercheur authentifié rassemble ses formes auteur avec un identifiant unique. Il peut aussi aligner son compte avec d'autres identifiants (IdRef, Orcid...). Cet

alignement est réalisé par l'auteur lui-même, qui ne connaît pas toujours ses identifiants. HAL offre aussi la possibilité de stocker des identifiants auteur lors de dépôts avec l'API Sword. L'identifiant est alors associé au couple forme auteur/document. À partir de l'application Star, l'Abes fournit systématiquement l'identifiant IdRef de l'auteur d'une thèse. Cet IdRef est visible (et cliquable) sur la page de la thèse, sans être rattaché à un utilisateur de la plateforme HAL.

L'amélioration de la qualité de ses référentiels – et notamment le référentiel auteurs – constitue l'un des objectifs de HAL. Pour cela, l'alignement des identifiants constitue un enjeu majeur. Il permet, lors de rapprochements de diverses données, de mieux identifier les doublons. Obtenir les différents identifiants des auteurs est donc particulièrement important.

Le CCSD travaille dans l'optique d'améliorer

ces alignements. Deux identifiants sont particulièrement visés : au niveau national, IdRef, pour améliorer et automatiser le travail d'alignement en cours avec l'Abes, qui récupère les comptes avec IdHAL et essaye d'associer l'IdRef correspondant ; au niveau international, Orcid.

Au-delà des actions entreprises avec l'Abes, le projet Conditor devrait permettre, en croisant de nombreuses sources bibliographiques, d'aligner les différents identifiants trouvés via les publications.

**BRUNO MARMOL**

Responsable plateformes productions  
CCSD /Inria  
[bruno.marmol@inria.fr](mailto:bruno.marmol@inria.fr)

**BÉNÉDICTE KUNTZIGER**

Responsable assistance utilisateurs et  
coordination des portails  
CCSD /CNRS – INRIA – Université de Lyon  
[benedicte.kuntziger@ccsd.cnrs.fr](mailto:benedicte.kuntziger@ccsd.cnrs.fr)



## SyMoGIH, de l'UMR 5190 – Larhra, et les « objets historiques »

<http://symogih.org> - <http://siprojuris.symogih.org>

Depuis une dizaine d'années, le laboratoire de recherche historique Rhône-Alpes, (Larhra) développe le projet technique et scientifique SyMoGIH. Cet enfant des *digital humanities* propose une ontologie générique de représentation des données historiques autorisant leur interopérabilité et leur publication sélective. Sur cette base a été mise en place une plateforme collaborative utilisée par plusieurs dizaines de chercheurs. Elle permet un stockage cumulatif des données par le recours à des référentiels d'« objets histo-

riques » et à un dictionnaire de types d'unités de connaissance, alimenté grâce à la modélisation évolutive de l'information historique. C'est ce recours à des référentiels qui rend le matériau historique réutilisable.

Outre la publication de ses référentiels sous forme d'URI liées à des notices d'autorité, SyMoGIH s'est engagé dans une collaboration avec l'Abes. Après le branchement d'IdRef et un test d'alignement concluant mené sur le corpus « Acteurs » du projet Siprojuris, Système d'information des professeurs de droit

(1804-1950), la poursuite de l'alignement, voire l'intégration, des autorités SyMoGIH avec les IdRef doit faciliter l'ouverture de notre entrepôt vers d'autres réservoirs de qualité, tout en enrichissant les IdRef.



**PIERRE VERNUS**  
Larhra  
[pierre.vernus@ish-lyon.cnrs.fr](mailto:pierre.vernus@ish-lyon.cnrs.fr)



## PRELIB et la littérature bretonne

<http://mshb.huma-num.fr/prelib/>

Prelib est une base de données prosopographique des acteurs de la littérature bretonne et de leur production littéraire, portée depuis 2012 par le Centre de recherche bretonne et celtique (CRBC, EA 4451/UMS 3554) à Brest. L'objectif de ce corpus est de comprendre les raisons qui ont conduit des individus à écrire et à publier en breton. C'est en cherchant un

outil pour aligner les individus de notre base avec Wikidata que François Mistral, responsable d'IdRef, nous a proposé d'aligner les individus de notre base avec le référentiel de l'Abes. Après quelques essais basés sur un export en XML de nos données, nous avons poursuivi l'alignement avec OpenRefine et le service de réconciliation `refine_viaf` mis à

disposition sur GitHub. Un script en Python permet d'interroger le web service `merge_inv` et de mettre à jour les PPN devenus obsolètes à la suite des fusions de notices.

**NELLY BLANCHARD**  
**JEAN-BAPTISTE PRESSAC**  
**MANNAIG THOMAS**  
CRBC



## CAPLAB, pour un suivi de l'activité des unités



Au deuxième trimestre 2017, l'Agence de mutualisation des universités et établissements (Amue) va

lancer, en partenariat avec le CNRS, le projet national Caplab de description et suivi des projets et activités des laboratoires.

S'inscrivant dans la trajectoire de construction d'un système d'information de la recherche au niveau national, Caplab viendra compléter Geslab (application de gestion financière des laboratoires) et Dialog (application de demande de ressources des laboratoires). Les données relatives aux personnels affectés dans les laboratoires (chercheurs, enseignants-chercheurs, ITRF), comme les métadonnées des productions/réalisations (notices de publications, thèses), seront recueillies dans les bases/référentiels/solutions externes

à Caplab et associées aux projets et activités gérés dans cette nouvelle application. Il est bien entendu envisagé que ces métadonnées proviennent de Conditor.

Ces besoins rendent l'usage d'un identifiant unique Auteur/Personne indispensable. IdRef pourrait répondre à ce double besoin, car l'Abes est en mesure de construire et proposer, avec cet identifiant et les données associées, un véritable référentiel national des notices d'autorité des chercheurs. Naturellement lié au référentiel national des thèses et aux ressources acquises par les bibliothèques universitaires françaises, IdRef est également pressenti comme l'identifiant auteur pivot/de référence du projet national Conditor, futur référentiel des notices de publication de la recherche française. Enfin, il est préconisé par le Cadre de cohérence de la

recherche publié en mai 2016 par le MENESR. La propagation de cet identifiant unique pourra encore s'accroître via le futur Référentiel partagé recherche (REFP) au sein du SI Recherche national, et vers les SI des établissements. Ainsi positionné stratégiquement, cet identifiant deviendra incontournable pour Caplab, dès lors qu'il pourra être aligné avec certains identifiants, notamment l'identifiant ORCID.

Enfin, les services d'alignement mis à disposition par IdRef pourraient s'avérer utiles dès les premières phases de construction de Caplab afin de recueillir les données des auteurs/personnes nécessaires au lancement de l'application dans les sites pilotes.

**ROMAIN THOUY**  
Amue  
[romain.thouy@amue.fr](mailto:romain.thouy@amue.fr)

# Le RNSR, colonne vertébrale du SI recherche

En délivrant un identifiant unique à chacune des structures qui mettent en œuvre la politique de recherche française, ce répertoire national leur permet d'articuler leurs systèmes d'information et de disposer d'éléments fédérateurs pour rapprocher les données aux niveaux local, régional et national.

La politique de recherche française s'inscrit dans un système complexe, composé d'universités, organismes de recherche, grandes écoles et autres établissements de type CHU, Fondations, GIP, etc. Au sein de ce dispositif à plusieurs niveaux, l'unité de recherche constitue l'entité de base : le plus souvent sous tutelle mixte, elle peut être liée à une ou plusieurs universités ou organismes de recherche, travailler au sein d'une fédération, d'une alliance et relever d'un ou plusieurs ministères de rattachement.

## IDENTIFIER EN TOUTE CERTITUDE

Par l'attribution d'un identifiant national unique à chacune des structures de recherche, le Répertoire national des structures de recherche (RNSR)<sup>1</sup>, reconnu comme « outil de référence » dans le Cadre de cohérence des SI Recherche<sup>2</sup>, favorise l'articulation entre les systèmes d'information de ces différentes institutions aux niveaux local, régional et national.

Obligatoire pour les activités relevant de l'administration publique – de type contrôle, allocation de moyens et de contrats –, le périmètre du RNSR a été étendu en 2015 à la **totalité de la recherche publique**, indépendamment des ministères de tutelle. Désormais, « tout acteur de la recherche publique française, partie prenante d'un système national d'information sur la recherche publique, doit intégrer cet identifiant national dans son SI. »

Mis en place il y a une dizaine d'années, le RNSR était alimenté à l'origine de façon centralisée à partir des données collectées par le MENESR dans le cadre de la contractualisation avec les universités. À la suite de la disparition de la dotation spécifique au titre de la recherche prévue par la loi LRU, le mode d'alimentation a été refondé. Ainsi, en 2013, chaque organisme se dote d'un correspondant qui, en dehors des campagnes annuelles d'actualisation, dispose des droits nécessaires pour une mise à jour des données au fil de l'eau, ce qui a amélioré considérablement la complétude et la fraîcheur des informations.

## SIMPLIFICATION, TRANSPARENCE, OUVERTURE

Si le RNSR offre aux chercheurs une nette **simplification** des tâches administratives – comme, par exemple, le pré-remplissage des formulaires d'évaluation ou des appels à projets –, le répertoire est également au service de la **gouvernance et du pilotage de la recherche**. Ainsi, diverses initiatives ont, par leur réussite, mis en évidence le rôle pivot joué par le RNSR pour structurer les systèmes d'information des différents opérateurs autour d'un objet identifié et partagé. Il est utilisé par de nombreux services comme Galaxie (publication des postes d'enseignants-chercheurs et procédure d'avancement) ; l'application Pélican (dépôts de dossiers dans les procédures d'évaluation par le HCERES) ;

## LES CHIFFRES DE FÉVRIER 2017

- 3 793 structures de recherche actives, dont
  - 3 264 unités de recherche
  - 39 équipes
  - 489 structures collaboratives sous la tutelle de 293 opérateurs de recherche.
- 2 360 structures mixtes (sous cotutelles)
- 1 433 structures d'un seul opérateur
- correspondants RNSR : 256 membres actifs.

En 2016, 28 nouveaux membres et 10 nouveaux opérateurs.

le SI des appels à projets de l'ANR, l'archive ouverte HAL... L'identifiant RNSR est en cours d'implémentation pour des usages diversifiés tels les référentiels partagés IdRef et Conditor, l'ARB (annuaire de la recherche biodiversité), les SI de l'OST (Observatoire des sciences et techniques), du Genci (Grand Équipement national de calcul intensif), de l'Ifremer... Servant de base au moteur de recherche ScanR<sup>3</sup>, les données du RNSR sont également disponibles sous forme de jeux de données sur la plateforme Open Data du MENESR<sup>4</sup>, ce qui permet une véritable capitalisation des informations par le rapprochement entre données de différentes sources, publiques ou privées.

### ISABELLE POULIQUEN

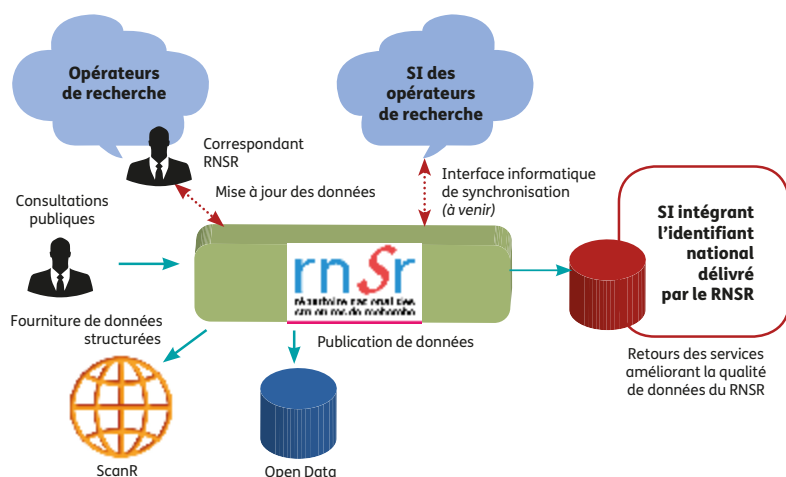
Responsable de la cellule d'urbanisation des SI  
isabelle.pouliquen@recherche.gouv.fr

### ISABELLE KABLA-LANGLAIS

Sous-directrice SIES  
isabelle.kabla-langlois@recherche.gouv.fr

### XIAOFENG CHEN

coordonnateur des systèmes d'information  
SD SIES, MENESR  
xiaofeng.chen@recherche.gouv.fr



[1] <https://appliweb.dgri.education.fr/rnsr/>

[2] [https://esr-wikis.adc.education.fr/ca2co/index.php/Cadre\\_de\\_cohérence\\_Recherche](https://esr-wikis.adc.education.fr/ca2co/index.php/Cadre_de_cohérence_Recherche)

[3] <https://scanr.enseignementsup-recherche.gouv.fr>

[4] [data.enseignementsup-recherche.gouv.fr/pages/home](https://data.enseignementsup-recherche.gouv.fr/pages/home)

# Le futur FNE : vers une vraie coproduction

À l'heure du web de données, les référentiels, or noir des bibliothèques, constituent un enjeu stratégique de visibilité et d'appropriation. La construction commune par l'Abes et la BnF d'un Fichier national d'autorité (devenu « d'entités ») a été jugée en 2015 prioritaire par le Comité stratégique bibliographique.

**Quels sont les objectifs du projet FNE ? En premier lieu, par la mutualisation de la production, il doit permettre de réaliser des économies d'échelle. Aujourd'hui, les deux agences bibliographiques gèrent deux fichiers d'autorité, sous deux formats distincts. Une première convergence est d'ores et déjà à l'œuvre :** le fichier d'autorité de la BnF est versé dans IdRef, dont les données sont partiellement alignées dans le fichier BnF, tous deux disponibles dans le fichier d'autorité international virtuel Vial. À présent, il importe de renforcer cette mutualisation par une vraie coproduction entre les deux agences et leurs réseaux. Elle apportera de la fluidité de création, de gestion et de réutilisation des données d'autorité au sein d'un même outil, et donnera les conditions de généraliser les bonnes pratiques sur les autorités. Intégrer des applications exogènes aux fichiers d'autorité – comme le Fichier national des propositions Rameau – assurera une meilleure réactivité face à l'émergence de concepts issus de la recherche.

## UNE COMPLÉTUDE BIENVENUE

Cette mutualisation contribuera à améliorer la fiabilité et à moderniser les autorités et les pratiques.

En effet, l'expérience en gestion de réseau et en coproduction de référentiels de l'Abes et l'expertise de la BnF en politique de qualité des données ne pourront que se compléter. De même, le FNE bénéficiera d'une couverture issue, d'une part, du dépôt légal et donc représentative de la

production culturelle nationale et, de l'autre, de l'Information scientifique et technique (notamment par le portail des thèses). La masse critique se prête en outre aux traitements automatisés ou semi-automatisés.

Outil de production mutualisé qui décloisonnera les deux silos actuels de données d'autorité, le FNE favorisera la réutilisation des données – avant tout, par les réseaux animés par l'Abes et la BnF, bien sûr. S'il n'est pas prévu d'interface publique de

recherche et de consultation, des API, bien recensées et documentées, permettront aux catalogueurs de lier, dans leur application de production, les données bibliographiques aux données d'autorité ; les usagers pourront ainsi consulter les données d'autorité rapatriées au sein des interfaces publiques, avec d'autant plus de pertinence que les données seront rafraîchies en temps réel. Mais l'accroissement de la visibilité des référentiels augmente aussi la responsabilité quant à l'information pérenne et contrôlée. Aussi la politique de qualité des données devra-t-elle faire l'objet de toutes les attentions. Une partie des énergies économisées par la mutualisation de la production pourront s'y redéployer.

Enfin, le principe de l'*open data* est réaffirmé, les deux acteurs ayant d'ores et déjà placé leurs données sous licence Etalab, et le recours aux solutions *open source*, privilégié.

## QUE TRADUIT LA BASCULE SÉMANTIQUE ?

Si l'ouverture à des usages autres que strictement bibliographiques sous-tend le projet, les référentiels intégrés au FNE continueront à s'appuyer sur une logique de signalement de ressources documentaires effectivement ou potentiellement mises à disposition par les producteurs – par exemple, pour Rameau, où le vocabulaire fait l'objet d'une arborescence thématique et systématique. Il ne sera pas exclu toutefois de compléter les données présentes dans le FNE par alignement avec les référentiels d'autres métiers (éléments biographiques issus de l'INSEE, numéro SIRET d'entreprises faisant l'objet d'une notice « collectivités »...).

Le FNE repose sur le concept d'entités, et non plus sur celui, très bibliothéconomique, d'autorité. Cette bascule sémantique traduit tout à la fois l'ancrage résolu dans le *linked open data*, mais aussi dans le modèle IFLA-LRM. De fait, si le premier cercle des référentiels qui alimenteront le FNE est constitué des champs traditionnels des autorités<sup>1</sup>, à terme, les expressions, les lieux, voire les laps de temps pourraient les rejoindre.

Le FNE pose également la problématique du passage d'une unité « notice » à une unité « donnée » : le modèle de production devra intégrer cette dimension et définir le niveau nécessaire de granularité pour une coproduction transparente, gage de qualité des données. La BnF travaille également à faire évoluer

↘ Le recours aux solutions *open source* sera privilégié pour la constitution du fichier national d'entités.



© Opensource.com/Wikipedia (CC BY-SA)

[1] Personnes, collectivités, sujets et œuvres.



en ce sens le format InterMarc pour intégrer les avancées de la transition bibliographique et gérer une part croissante de données exogènes, comme les données des éditeurs, dans son catalogue.

## UN CAS D'ÉCOLE

Les premiers producteurs – en termes de volumétrie et de déploiement – ont vocation à être les catalogueurs de la BnF et du réseau Abes, quelles que soient les applications utilisées : Sudoc ou Calames, PIXML ou le futur outil de catalogage de la BnF. Mais l'objectif est bien d'ouvrir rapidement cet outil aux bibliothèques territoriales, dont les référentiels restent le plus souvent dans les outils locaux et dont le versement constituerait une plus-value, notamment en termes de couverture géographique. Le groupe de travail national en charge de l'étude de faisabilité pour un outil mutualisé en EAD avait d'ailleurs considéré la mise en place d'un FNE comme un préalable absolu pour assurer la cohérence et la qualité des données.

Décloisonner les entités signifie également ouvrir potentiellement le FNE à d'autres acteurs. Il faut auparavant déterminer si cette ouverture passera par des alignements gérés au niveau des pilotes opérationnels du FNE, ou si la possibilité sera offerte d'une réelle coproduction d'entités « au fil de l'eau ». Sont envisagés, par exemple, des partenariats avec les archives, les musées ou la Direction de l'information légale ou administrative. En parallèle de l'instruction avec l'Abes, la BnF mène, avec les Archives nationales et le Service interministériel des archives de France, le programme de recherche Pilote d'interopérabilité des autorités archivistiques françaises (PIAAF). Son objectif est de prototyper l'alignement des données d'autorités « collectivités » de la BnF et d'autorités archivistiques d'administrations productrices. Au-delà du rapprochement bibliothèques-archives, ce projet constituera un cas d'école, dont les conclusions pourront être réexploitées pour un versement du Répertoire national des structures de recherche dans le FNE.

Les calendriers de l'Abes et de la BnF convergent pour une ouverture du FNE à l'horizon 2020, qui est également, pour la BnF, celui du déploiement d'un nouvel outil de production pour ses données bibliographiques. S'il est encore trop tôt pour présenter une feuille de route détaillée, l'année 2017 devrait permettre de préciser les fonctionnalités attendues et de poser les premiers jalons en termes de modèle et de gouvernance des données et d'architecture technique. Trois groupes de travail BnF-Abes se réuniront. Le premier sera consacré à la constitution du bloc de données de départ du FNE, en partant des alignements des jeux de données. Le second déterminera la plus-value du FNE dans un contexte de coproduction, en l'illustrant par plusieurs cas d'usage parmi les plus représentatifs.

Enfin, le troisième groupe pilotera une prestation de *benchmark*, qui étudiera les outils d'alignement et de coproduction de référentiels existants et le fonctionnement d'une sélection de fichiers nationaux d'entités. Les livrables de ces trois groupes seront présentés au Comité stratégique bibliographique de décembre 2017, qui validera la feuille de route pour les années suivantes.

FRÉDÉRIQUE JOHANNIC-SETA

Directrice du Département des métadonnées, BnF  
frederique.johannic-seta@bnf.fr

{BnF

## UN NOUVEAU RAMEAU ?

Le référentiel sujet Rameau, largement connu et utilisé dans le monde des bibliothèques françaises et francophones, a mis en route depuis 2015 quelques changements peu spectaculaires en apparence, mais qui pourraient bien lui donner, par un soudain effet cumulatif, un tout nouveau visage. Ouverture, interopérabilité et restructuration des données en sont les mots-clés.

• **Interopérabilité.** À ses équivalences historiques avec les LCSH ou Laval RVM, Rameau en a ajouté, au fil du temps, bien d'autres<sup>1</sup>, toutes signalées de façon hétérogène dans les notices. Il fallait remédier à cette dispersion, en incluant ces équivalences et en élargissant le spectre des référentiels avec lesquels les autorités Rameau se lieraient. Une nouvelle zone permet désormais de se lier à de nombreux référentiels via une URI<sup>2</sup>, lorsqu'elle existe.

• **Collaboration Inist-Rameau.** Rameau s'est engagé depuis 2007 dans un long processus, entièrement manuel, de révision terminologique domaine par domaine. En collaboration avec l'équipe Terminologie-TAL de l'Inist, une méthodologie innovante d'échange et d'enrichissement réciproque est actuellement mise à l'épreuve pour deux sous-domaines<sup>3</sup>.

Des experts du Centre ont aligné deux fichiers en SKOS (TermSciences/Rameau) pour dégager des équivalences exactes qui ont été contrôlées par les deux partenaires. Chaque notice a ensuite fait l'objet d'un enrichissement.

• **Translittérations.** Le centre Rameau a décidé que ses notices d'autorité pourraient être enrichies, dans le cas des termes empruntés à des langues non latines, par l'ajout des translittérations ISO (qui permettent l'interopérabilité des données), et de caractères dans la langue d'origine<sup>4</sup>. Les chantiers successifs qui s'y consacrent convergent avec les pratiques des autres services d'autorité de la BnF.

• **Révision de la syntaxe.** Un groupe de travail national a été créé en mars 2016, avec pour mandat de formuler, dans un rapport attendu pour le printemps 2017, des préconisations d'évolution de Rameau – en particulier de sa syntaxe –, qui répondent aux demandes des utilisateurs. Il faut adapter Rameau à l'ensemble des transformations de son domaine (web de données, modélisation IFLA-LRM, projet de Fichier national des entités, nouveaux outils de recherche et de catalogage) ; le rendre plus simple et plus intuitif ; valoriser sa richesse sémantique pour en maximiser les possibilités d'exploitation et de réutilisation dans tous les contextes possibles. À suivre.

FLORENCE MENARD

Responsable du Centre national Rameau  
florence.menard@bnf.fr

[1] Par exemple, des classifications chimiques ou organologiques.

[2] *Uniform resource identifier*.

[3] « Taxonomie des poissons » et « Transferts de chaleur ».

[4] Par exemple, la notice du mot sanscrit « Vr vddhi » (« augmentation »).

# La culture dessine son « graphe »

En soutenant des projets d'interconnexion de données et référentiels culturels, le ministère de la Culture et de la communication vise à constituer une base de connaissances, appelée « graphe culture », recourant aux technologies du web sémantique. Ce travail d'harmonisation et de modélisation permettra de décloisonner les différents secteurs concernés.

**E**n lien avec la mission Etalab, créée en 2011, le ministère de la Culture et de la Communication (MCC) coordonne la politique d'ouverture et de partage des données culturelles. Une plateforme dédiée, [data.culturecommunication.gouv.fr](http://data.culturecommunication.gouv.fr), a été lancée fin 2016, afin de valoriser les biens culturels et offrir à leurs producteurs des outils de réutilisation, tels que la cartographie ou la data-visualisation. Dans la lignée de cette politique, le MCC s'est très rapidement emparé de la question de l'application du web sémantique au domaine dont il a la responsabilité. Cet enjeu stratégique a conduit le ministère à sensibiliser ses services et établissements publics à ces technologies. Certains établissements publics, comme la Bibliothèque nationale de France (BnF) ou

services et établissements publics à produire, gérer, exposer et lier ses référentiels culturels en utilisant les technologies du web sémantique.

## LE PROGRAMME HADOC : DÉCRIRE N'IMPORTE QUEL BIEN CULTUREL

La constitution du « graphe culture » s'inscrit dans la continuité du programme Hadoc, conduit par le ministère depuis 2008, qui vise à l'harmonisation de la production des données culturelles. Le ministère a lancé ce programme en partant du constat que plusieurs secteurs métiers, tels que l'inventaire du patrimoine, les monuments historiques ou encore les musées, des secteurs certes différents mais néanmoins proches, fonctionnaient totalement en

**Des secteurs proches, comme l'inventaire du patrimoine, les monuments historiques ou les musées, fonctionnaient en silos et ne communiquaient pas entre eux**

silos et ne communiquaient pas entre eux. Le modèle de données Hadoc a vocation à décrire n'importe quel bien culturel. À terme, l'ensemble des descriptions ainsi géné-

le Centre Pompidou, ont été pionniers en la matière. La BnF a lancé la plateforme [data.bnf.fr](http://data.bnf.fr) dès 2011, dans une dynamique d'ouverture, de structuration et de liage des données.

## UNE FEUILLE DE ROUTE FORMALISÉE EN NEUF AXES

Afin de définir une stratégie sur le web sémantique, globale à l'ensemble de ses services et établissements, le ministère a constitué en 2013 des groupes de travail réunissant des experts de chaque secteur (bibliothèques, archives, musées, musique, audiovisuel et spectacle vivant). Leurs réflexions communes ont abouti à la production d'une feuille de route<sup>1</sup>, publiée en 2014 et formalisée selon neuf axes opérationnels et stratégiques, qui couvre les aspects tant techniques ou technologiques qu'organisationnels. Cette feuille de route permet de répondre aux enjeux actuels du web des données liées en définissant notamment un programme de formation des agents du MCC, un partenariat de R&D signé avec l'Inria<sup>1</sup> ou encore des preuves de concept permettant d'expérimenter et démontrer l'interopérabilité sémantique des données culturelles.

Un des neuf axes stratégiques et opérationnels de la feuille de route propose la constitution d'un « graphe culture ». L'objectif pour le ministère est d'inviter ses

rées est amené à constituer le référentiel des biens culturels. Cet important travail d'harmonisation et de modélisation a concerné la partie descriptive du bien culturel, mais également la partie « référentiel » qui l'alimente. Les services concernés ont donc travaillé à mettre en cohérence les thésaurus spécialisés utilisés par les différentes bases de données sectorielles.

## L'APPLICATION GINCO : GÉRER LES VOCABULAIRES SPÉCIALISÉS

Ginco<sup>2</sup>, acronyme de Gestion informatisée de nomenclatures collaboratives et ouvertes, a été développé par le MCC dans le cadre du programme Hadoc pour proposer un outil complet de création et de gestion de vocabulaires spécialisés. Ginco implémente la norme ISO-25964<sup>3</sup> pour la production des vocabulaires. Cette norme, publiée en deux volets, en 2011 et en 2013, révisé les normes de conception de thésaurus et d'interopérabilité dans la perspective du web des données liées. L'ensemble des services du MCC collaborent à la création, l'harmonisation et l'alignement des référentiels par cet outil, qui permet ensuite d'exposer l'ensemble des vocabulaires scientifiques et techniques en SKOS. La plateforme de diffusion de Ginco, distincte de celle de production, permet une consultation en ligne, un téléchargement au format SKOS ou des requêtes complexes en SPARQL. Ces référentiels

[1] Le MCC et l'Inria ont signé une convention-cadre, le 12 décembre 2016, pour soutenir des projets de R&D culturels innovants.

[2] [culturecommunication.github.io/ginco/](https://github.io/ginco/)

[3] ISO 25964-1(2011) « Thésaurus pour la recherche documentaire » et ISO 25964-2(2013) « Interopérabilité avec d'autres vocabulaires ».

[4] Isni, code international normalisé des noms (en anglais *International Standard Name Identifier*), [www.isni.org](http://www.isni.org)

[5] Viaf, fichier d'autorité international virtuel (en anglais *Virtual International Authority File*): <https://viaf.org>

du MCC sont alignés avec d'autres référentiels nationaux de la sphère culturelle, comme Rameau de la BnF, ou extérieurs et internationaux, comme celui du Getty ou Iconclass.

## LE PROJET ONOMA : IDENTIFIER LES ACTEURS CULTURELS

Au même titre que les vocabulaires spécialisés sur les domaines, les sujets représentés ou les entités abstraites, le champ des acteurs historiques fait également l'objet d'un chantier important dans le cadre du programme Hadoc. Le projet Onoma (en cours) vise à créer un référentiel des acteurs, en s'appuyant sur les technologies du web sémantique. La notion d'« acteurs » recouvre les auteurs, les créateurs, les producteurs, ou toute personnalité intervenant dans le cycle de vie d'un bien culturel. La clé de voûte du projet repose sur l'attribution d'un identifiant pérenne et unique pour ces acteurs. Un des neuf axes stratégiques de la feuille de route propose notamment de déployer le dispositif Isni<sup>4</sup>. La base Isni, qui s'est constituée à partir de la base Viaf<sup>5</sup>, comporte plus de 7,5 millions d'identifiants sur des auteurs patrimoniaux et contemporains. La BnF est devenue une agence d'enregistrement en 2014 et assure l'attribution et la diffusion d'identifiants Isni. L'enjeu pour le MCC, à travers le projet Onoma, est de déployer largement les identifiants Isni pour les auteurs et acteurs historiques peu connus ne disposant pas d'identifiants et de diffuser et interconnecter les données relatives à ceux déjà identifiés. Cette démarche contribuera à la visibilité et au rayonnement des auteurs français à l'échelle nationale et internationale.

L'ensemble de ces référentiels culturels, formalisés selon les standards du web sémantique et interconnectés, participe à la constitution de la base de connaissances appelée le « graphe culture ». L'enjeu pour le MCC est de favoriser l'enrichissement de ce graphe en soutenant des projets d'interconnexion de données et référentiels culturels, mais aussi de veiller à sa maintenance et à son actualisation. La perspective de ce « graphe culture » marque l'engagement du MCC pour une coproduction des connaissances, destinée à désenclaver les bases de données institutionnelles et casser les fonctionnements en silos en misant sur les approches participatives.

### MARIE-VÉRONIQUE LEROI

Chargée de mission au département  
de l'innovation numérique  
Ministère de la Culture et de la Communication  
marie-veronique.leroi@culture.gouv.fr



© Library of Congress Prints and Photographs Division



Woman on tricycle, followed by men on penny-farthings, Hy Sandham, 1887. Une feuille de route (<http://vu.fr/3r6>) a été publiée par le MCC, en 2014, pour répondre aux enjeux actuels du web de données.

## ● ● ● ARCHIPÉDIE, UNE ENCycLOPÉDIE DÉDIÉE À L'ARCHITECTURE

[archipédie.citedelarchitecture.fr](http://archipédie.citedelarchitecture.fr)

La Cité de l'architecture et du patrimoine (Capa) s'est engagée, ces dernières années, dans la refonte et l'harmonisation de ses sites Web et de ses bases de données. Archipédie, l'un des premiers résultats de cette démarche, se présente comme une encyclopédie numérique collaborative, consacrée à l'architecture moderne et contemporaine, et entend s'adresser au plus grand nombre. La connexion interinstitutionnelle de données a pu être expérimentée, dans la mesure où les notices d'architectes d'Archipédie sont liées par l'identifiant Isni aux notices correspondantes de la base Agorha, de l'Institut national de l'histoire de l'art (Inha). Ainsi une recherche sur un architecte permet de consulter la notice qui lui est consacrée, mais également celle sur ses œuvres ou les collectifs auxquels il a pris part.

## ● ● ● DOREMUS, UN PROJET DE REPRÉSENTATION SÉMANTIQUE D'ŒUVRES MUSICALES

[www.doremus.org](http://www.doremus.org)

Doremus est un projet financé par l'Agence nationale de recherche et réunissant la BnF, la Philharmonie de Paris, Radio France, et des partenaires universitaires et privés. Doremus s'attache à formaliser et décrire de la façon la plus complète une œuvre musicale. Ce projet s'inscrit pleinement dans la feuille de route pour le Web sémantique du MCC et participe à la constitution du « graphe culture ». Prévu pour être finalisé en 2018, il mettra à disposition des modèles et des référentiels pour la représentation sémantique d'œuvres musicales, afin de rendre interopérables les catalogues d'œuvres existants.

# L'herbier du Muséum : retour d'expérience sur la co-construction de référentiels

**Pour les bibliothécaires, cataloguer une collection et en établir les références – à travers les fichiers d'autorité – est l'affaire de spécialistes du métier, souvent soumis à l'exercice solitaire et érudit d'une expertise, à la fois bibliothéconomique et disciplinaire.** L'expérience du Muséum national d'histoire naturelle (MNHN), qui s'intéresse à la conservation et au signalement de collections de spécimens naturalistes et aux publications et aux archives qui documentent ces collections, interroge cette vision des choses et ouvre le débat sur la possibilité de partages et de pratiques collaboratives qui présentent un intérêt économique réel pour les institutions patrimoniales en même temps qu'elles encouragent le développement des sciences participatives ou citoyennes.

La rénovation de l'Herbier de Paris a été l'occasion de numériser massivement les spécimens de ses collections de botanique. À la faveur d'un financement exceptionnel, poursuivi aujourd'hui dans le cadre de l'infrastructure de recherche e-ReColNat, plus de 5 millions d'images ont été mises en ligne entre 2008 et 2012. Une opération de cette ampleur a conduit à repenser la pratique classique de catalogage régulier par des gestionnaires de la collection doublé d'une prise de vue des spécimens les plus précieux. En effet, le catalogue, malgré ses plusieurs centaines de milliers d'entrées dans la base de données initiale, s'est vu complété en un temps record de plusieurs millions d'entrées pour lesquelles l'information se réduisait souvent à une photographie et un numéro d'inventaire.

Comme la plupart des images numérisées permettent de lire l'étiquette du spécimen qui contient les informations à cataloguer, le Muséum a souhaité expérimenter auprès du grand public afin de voir si le succès des sciences participatives des observatoires naturalistes pouvait être étendu à la transcription d'étiquettes pour les collections. C'est ainsi que le site [lesherbonautes.mnhn.fr](http://lesherbonautes.mnhn.fr) est né, offrant à chacun la possibilité de contribuer à la création d'une base de données scientifiques à partir de la transcription des textes présents sur les photographies des pages d'herbiers.

## DANS LA PEAU D'UN HERBONAUTE

Participer à la création d'une base de données scientifique destinée à faire progresser la connaissance des évolutions de la biodiversité, telle est la proposition faite à l'internaute qui s'aventure sur le site Les Herbonautes. Il lui suffit pour cela de choisir l'une des « missions » proposées sur le site. Chacune correspond à un corpus thématique, constitué autour d'un groupe de plantes et d'un projet de recherche, qu'explique un paragraphe de mise en contexte et de présentation des enjeux de l'étude de cette sélection de spécimens.

Un clic sur le bouton *Contribuer* fait apparaître au centre de l'écran une page d'herbier numérisée, tandis que s'affiche à gauche le menu à partir duquel l'internaute réalisera ses contributions. Dans ce menu, une liste de champs apparaît – pays, région, date, récolteur, déterminateur, localité, géolocalisation, numéro de récolte, herbier propriétaire : les références indispensables à l'identification d'une espèce, en botanique comme dans la pratique taxonomique en général. Le néophyte ne peut cependant accéder qu'au premier de ces domaines, l'identification du pays d'origine du spécimen, qu'il doit déterminer à partir des informations figurant sur la page. Le nom du pays est sélectionné dans un menu déroulant renvoyant à un référentiel géographique préétabli, évitant au contributeur de se soucier de l'orthographe ou de la casse. L'encart d'aide précise que le pays



© Muséum national d'histoire naturelle (CC BY)

↑  
Planche de  
*Silene noctiflora*,  
spécimen intégré dans  
la mission « Silènes du  
Maghreb » sur le site des  
Herbonautes.  
Herbier Muséum de  
Paris, herbier Delpature.

sélectionné doit être le pays actuel, ce qui peut occasionner des recherches pour des spécimens historiques. Des outils sont mis à disposition pour faciliter cette recherche : consignes de lecture et d'écriture, liens utiles à la localisation des toponymes, à l'identification des noms de personnes, mais aussi les propositions des autres contributeurs sur le même spécimen ainsi que les commentaires qu'ils ont pu laisser à la suite d'une identification difficile. L'outil de visionnage de la page d'herbier permet un zoom d'excellente qualité ainsi qu'une navigation aisée sur la page, afin de permettre à l'internaute de déchiffrer les étiquettes dans les meilleures conditions possibles. En cas de doute, il peut moduler sa réponse en cochant « Je l'ai déduit », ou « J'hésite ». Il peut également noter que l'étiquette ne présente pas d'information sur le pays de provenance du spécimen, passer son tour en demandant à changer de spécimen, ou avertir que la mauvaise qualité de la numérisation rend la photographie inutilisable.

Pour enregistrer sa contribution, le contributeur crée un identifiant, qui est souvent un pseudonyme associé à une adresse de courriel. Ce compte personnel lui confère son identité d'herbonaute. Il peut à tout moment consulter le tableau de suivi de ses contributions. Son profil peut être complété d'un avatar et d'une courte description, détails appréciés lors des interactions avec les autres herbonautes. Lors de sa troisième contribution, l'utilisateur est soumis à un premier quiz pour tester ses compétences mais aussi apprendre à décoder correctement les informations qui figurent habituellement sur une page d'herbier. S'il passe le test avec succès, l'herbonaute reçoit un badge et accède au niveau supérieur, qui lui ouvre un nouveau champ de contribution nécessitant une analyse plus précise ou des recherches plus complexes. Les badges collectés au fur et à mesure de sa pratique s'affichent sur son profil à côté du nombre de contributions et de spécimens examinés, témoignant de son expérience et de ses connaissances auprès des autres membres de la communauté.

## LA FOULE A DU TALENT

Lancé fin 2012 sans beaucoup de publicité, le site a très rapidement rencontré son public. Aujourd'hui, près de 2 700 herbonautes ont contribué à documenter 186 000 spécimens en répondant à plus de 2,4 millions de questions. Cette participation est encouragée par des aspects empruntés à l'univers du jeu

vidéo : déblocage de niveau, classement, badges, discussions entre joueurs. Mais les photographies d'herbier ont leur propre pouvoir de fascination. La beauté de certaines plantes et des étiquettes font voyager dans le temps et l'espace en racontant l'histoire de collectes aux quatre coins du monde depuis plusieurs siècles. Autre incitation, le regroupement de spécimens en « missions » qui ont une unité thématique, dont on peut suivre la progression jusqu'à l'objectif annoncé et qui peuvent être terminées en quelques semaines. Certaines de ces missions sont animées par des scientifiques à qui l'aide des herbonautes sert à documenter des spécimens nécessaires à leur recherche. Les participants apprécient beaucoup cet échange direct avec les botanistes.

Des comparaisons entre des transcriptions effectuées par des professionnels et par Les Herbonautes ont montré que le grand public était capable de transcriptions d'une qualité équivalente, voire supérieure à celles effectuées par les agents du MNHN. Ce résultat inattendu s'explique en partie par la sélection des participants, qui ne répondent qu'à des questions simples : seuls les plus motivés ont accès aux questions les plus délicates, après avoir répondu à plusieurs centaines de questions simples et passé des tests pédagogiques. Mais la plus grande force de la « foule » vient de la redondance des réponses, qui permet leur consolidation. Chaque question est posée à plusieurs personnes et en cas de désaccord, les herbonautes sont encouragés à en discuter sur un forum jusqu'à l'obtention d'un consensus. Certains ont ainsi développé des compétences en transcription des écritures anciennes ou en géographie coloniale, qui ne sont pas forcément les domaines d'expertise des chercheurs en botanique.

## VERS DE NOUVEAUX MODÈLES DE COLLABORATION

À l'heure où les barrières entre les institutions s'estompent afin de mutualiser les compétences nécessaires à la publication sur le Web de référentiels utiles à des bases de données d'intérêt national ou international, l'expérience probante des Herbonautes conduit le Muséum à envisager de dupliquer et d'étendre cette approche collaborative du catalogage en lien avec la numérisation. La nouvelle organisation du Muséum, mise en œuvre en 2017, réunit dans une même direction générale déléguée les services en charge du signalement des collections

naturalistes et documentaires. C'est l'occasion tout à fait inédite de réunir une masse critique de compétences bibliographiques et scientifiques autour d'un même ensemble disciplinaire cohérent – la taxonomie et, plus largement, les sciences du vivant. Le Muséum souhaite par ailleurs adopter et promouvoir les principes de l'ouverture et du partage des données publiques. La proposition de l'Abes et de la BnF de constituer un fichier national d'entités ouvert aux contributions de leurs réseaux respectifs apparaît comme un cadre particulièrement stimulant pour imaginer de nouvelles formes de collaboration. Si le Muséum pouvait ainsi alimenter les futurs hubs de métadonnées portés par les opérateurs nationaux de données produites conjointement par ses bibliothécaires, ses chercheurs et les internautes qu'il a su embarquer dans les sciences participatives, ce serait sans nul doute une belle aventure à tenter. L'intention de cette collaboration est posée, il reste à en construire la mécanique.

### CHLOÉ BESOMBES,

*Conservatrice des bibliothèques  
Chargée de mission architecte systèmes  
et données, MNHN, Direction des bibliothèques  
et de la documentation  
chloe.besombes@mnhn.fr*

### SIMON CHAGNOUX

*Ingénieur, Responsable applications  
scientifiques, MNHN, DSI  
simon.chagnoux@mnhn.fr*

### GILDAS ILLIEN

*Conservateur général des bibliothèques  
Directeur des bibliothèques  
et de la documentation, MNHN  
gildas.illien@mnhn.fr*



### POUR EN SAVOIR PLUS

L'expérience des Herbonautes a trouvé un écho dans plusieurs publications qui peuvent être lues en ligne :

- E. Pérez et M. Pignal. « Numériser et promouvoir les collections d'histoire naturelle ». Bulletin des bibliothèques de France (BBF), 2013, n° 5, p. 27-30. <http://urlz.fr/57Le>. ISSN 1292-8399.
- G. Le Bras, M. Pignal et al., 2017, "The French Muséum national d'histoire naturelle vascular plant herbarium collection dataset" <http://vu.fr/3ra>

Le 15 décembre 2016, après des décennies de débats et de travaux, le site Richelieu accueillait le public dans ses magnifiques espaces rénovés. Abritant depuis plusieurs siècles la Bibliothèque nationale, il réunit désormais la Bibliothèque nationale de France, l'Institut national d'histoire de l'art et l'École nationale des chartes. Visite d'un lieu historique entièrement repensé.

# Richelieu : les portes s'ouvrent sur l'art et l'histoire

Ce campus d'un genre unique se veut outil documentaire, scientifique et culturel dédié à l'art, à son histoire et au patrimoine. Il ambitionne d'être un cadre de travail exceptionnel pour les chercheurs et, dans le même temps, ouvert à un public plus large désireux de découvrir ce site et ses collections. L'origine du projet remonte à la fin des années 1980. En 1988, François Mitterrand annonce une future « Très Grande Bibliothèque ». Cette décision entraîne la création de la BnF, qui depuis héberge les Imprimés sur son site de Tolbiac. La question de l'avenir du site Richelieu, vidé d'une grande partie de ses collections, se pose alors. Plusieurs scénarios sont envisagés. Il est finalement décidé d'y réunir les départements spécialisés de la BnF, la bibliothèque de l'INHA et celle de l'ENC. Si chaque institution reste autonome, elles entament une réflexion sur l'accueil des publics, la politique documentaire, culturelle et scientifique, et les modalités de fonctionnement au sein d'un espace qu'elles seront amenées à partager.

## SUR LE PASSÉ GLORIEUX PASSE UN SOUFFLE NOUVEAU

Ce site est exceptionnel. Son histoire est à la fois celle de la Bibliothèque nationale et de la richesse de ses collections, mais aussi un ensemble patrimonial au sein duquel a œuvré une succession d'architectes qui ont tour à tour embelli, rationalisé, modifié, densifié, modernisé, modelé les espaces afin de répondre aux différents usages selon les époques. Mais ce passé prestigieux a son revers : à la fin des années 1990, l'endroit est devenu vétuste. Les conditions de sécurité et de conservation ne sont plus assurées. Les circulations sont malcommodes, les bureaux, étroits, et l'accueil du public, malaisé. Le bâtiment n'a jamais connu de rénovation d'ensemble depuis le XIX<sup>e</sup> siècle et ne correspond plus aux besoins des utilisateurs. C'est pourquoi un projet de rénovation du site est lancé. Il s'agit non seulement de redéfinir

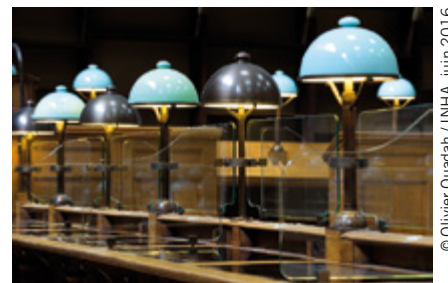
les espaces, mais aussi de leur donner un nouveau souffle, un nouveau destin. S'ouvre alors une période d'études de programmation préalables (2000-2006). Si, dès 1998, la salle Labrouste est attribuée à l'INHA, alors encore en gestation, la question du partage des lieux entre les institutions est difficile. Malgré le départ des Imprimés de la BnF, la place manque vite. Il faut rogner sur les espaces d'exposition ou de stockage. En parallèle s'élaborent aussi les grands principes d'organisation des travaux. La BnF souhaite être toujours ouverte au public. Un plan en deux phases est donc conçu. Le site est coupé en deux zones. Tandis que la zone 1 sera en chantier, la zone 2 accueillera l'ensemble des activités du site, et réciproquement.

## 15 ÉTAGES DE CHANTIER PRÈS DES SALLES DE LECTURE

En 2006, une convention permettant de lancer le chantier est signée entre les ministères de la Culture et de l'Enseignement supérieur. L'Opérateur du patrimoine et des projets immobiliers de la culture (OPPIC) pilote les travaux. Le budget alloué est de 232 millions d'euros. En 2007, l'agence des architectes Bruno Gaudin et Virginie Brégal est choisie. Les parties classées de la zone 1 sont confiées aux architectes des monuments historiques Jean-François Lagneau et Patrice Girard. En 2009, le projet de rénovation est validé par les tutelles et le permis de construire, déposé. En 2010, la zone 1 est vidée de ses occupants : bibliothécaires et collections se resserrent dans la zone 2 ou partent provisoirement sur d'autres sites. Le chantier peut débuter. Il commence en 2011 par la mise en place d'une paroi coupe-feu entre les deux zones, créant, pour plusieurs années, deux mondes voisins, mais totalement étanches : d'un côté, un chantier de plus de 30 000 m<sup>2</sup> sur 15 niveaux, abritant plus de 200 ouvriers de tous les corps de métiers ; de l'autre, la vie presque habituelle des salles de lecture de



© Jérôme Delatour / INHA, décembre 2016



© Olivier Ouadah / INHA, juin 2016

la BnF et de l'INHA. Cette première étape est prévue pour durer deux ans, il en faudra cinq. Malgré les difficultés, le chantier avance et permet de faire des découvertes sur l'histoire du site à mesure que le bâtiment se dévoile. Au printemps 2016, l'OPPIC livre la zone 1 à la BnF. La commission de sécurité donne son feu vert. Les trois institutions peuvent enfin s'installer.

La BnF et l'INHA, dont les espaces sont interdépendants, décident de rédiger ensemble un marché de transfert. Il s'agit de déménager 50 km linéaires de collections et 300 agents, dans un délai incompressible de neuf mois. Un planning serré est établi, tandis qu'en parallèle, les équipes de chaque institution mènent de front la préparation des collections et le maintien de l'activité. L'École des chartes, de son côté, planifie le déménagement de sa



bibliothèque – encore abritée à la Sorbonne – pour la fin de l'année 2016 (la livraison des rayonnages neufs a pris du retard). Pour l'INHA et la BnF, le transfert commence en août. Les salles de lecture ferment en septembre (INHA) et octobre (BnF), pour rouvrir aux lecteurs en zone rénovée, le 15 décembre 2016. En coulisse, les chaînes de transferts continuent jusqu'à la fin du mois de mars 2017. La bibliothèque de l'École des chartes, quant à elle, ferme le 15 octobre 2016, pour rouvrir le 6 mars 2017.

Le site a été inauguré le 11 janvier 2017 par le président de la République. Les trois institutions ont, à cette occasion, organisé un week-end portes ouvertes : plus de 14 000 personnes s'y sont pressées, prêtes à patienter deux heures et plus pour voir ou revoir ces bibliothèques.

### HEUREUX RAPPROCHEMENT

Désormais voisines, les bibliothèques de l'École nationale des chartes et de l'Institut national d'histoire de l'art sont vouées à tisser des collaborations nouvelles. Les équipes des deux institutions disposent en effet de bureaux à proximité immédiate et de plusieurs espaces partagés, à savoir une salle de détente et deux ateliers, au sein desquels travaillent les personnels de l'atelier de reliure et de restauration de l'INHA ainsi que la restauratrice de l'ENC. Certaines orientations similaires de leur poli-

tique de services tendent déjà à rapprocher les deux bibliothèques. En particulier, la priorité donnée au libre accès est au cœur de chaque projet ; dans les deux cas, celle-ci s'est appuyée sur une importante réorganisation des collections et sur la mise en place d'une signalétique claire et lisible, faisant apparaître les ensembles thématiques – les deux bibliothèques ont à cet effet travaillé avec le même signaléticien. L'INHA et l'ENC ont également profité de la rénovation du bâtiment pour moderniser leurs équipements et offrir de nouveaux services aux lecteurs. Les salles de lecture disposent désormais de places de travail équipées de prises électriques et connectées, y compris en wi-fi, et de nouveaux matériels de reprographie et de numérisation en libre-service.

### CIRCULEZ, IL Y A TOUT À LIRE !

Une complémentarité existe également entre les collections, puisque la bibliothèque de l'ENC dispose de collections en histoire de l'art et en archéologie, particulièrement pour la période médiévale en France, au titre des enseignements dispensés par l'École dans ces deux disciplines, sans toutefois prétendre au niveau d'exhaustivité pratiqué à la bibliothèque de l'INHA. Il est donc prévu d'instaurer une véritable coopération documentaire, afin de mieux articuler les acquisitions et la politique de conservation des deux bibliothèques.

Ce rapprochement ouvre également la voie à des projets communs de valorisation, à commencer par les collections de l'historien de l'art médiéval Léon Pressouyre (1935-2009), dont les archives ont été données à l'INHA, tandis que la bibliothèque a été attribuée à l'ENC. Au-delà de cette proximité évidente, le projet de rénovation du site Richelieu prévoit l'ouverture d'une circulation entre les deux bibliothèques, rendue possible par la création d'une porte entre le magasin central de l'INHA et la salle de lecture de la bibliothèque de l'ENC au rez-de-chaussée. Cette porte, actuellement réservée à l'usage des seuls personnels, devrait être ouverte au public au terme d'une réflexion sur l'impact de cette ouverture en termes de flux de lecteurs et de collections et sur la nécessaire adaptation des infrastructures techniques (systèmes antivols, SIGB, etc.). La création d'une circulation nouvelle pourrait également déboucher sur l'association de l'ENC au dispositif de carte commune qui existe désormais pour les lecteurs de l'INHA et de la BnF.

**CAMILLE DÉGEZ-SELVES**

*Directrice de la bibliothèque de l'École nationale des chartes  
camille.degez-selves@enc-sorbonne.fr*

**ANNE-ELISABETH BUXTORF**

*Directrice de la bibliothèque de l'Institut national d'histoire de l'art  
anne-elisabeth.buxtorf@inha.fr*

(Pleins feux sur... LE SITE RICHELIEU À PARIS)

# Bibliothèque des Chartes

## Adieu Sorbonne, bonjour Petits-Champs

**C**réée en 1846, vingt-cinq ans après la fondation de l'École, la bibliothèque de l'École nationale des chartes vient d'emménager, le 6 mars 2017, à sa nouvelle adresse : 12, rue des Petits-Champs, à Paris. Elle accueille les élèves, étudiants, enseignants-chercheurs et personnels de l'École, les archivistes paléographes, les chercheurs en sciences humaines et sociales à partir du niveau master, les personnels des archives, bibliothèques et musées, et les professionnels de la filière du livre.

Après une fermeture de près de cinq mois, la bibliothèque de l'École a quitté ses anciens locaux en Sorbonne et rouvert ses portes dans les espaces rénovés du site Richelieu. Elle y occupe les bâtiments conçus par l'architecte Henri Labrouste, le long de la rue des Petits-Champs, anciennement dévolus au département des Imprimés de la Bibliothèque nationale, et s'achevant sur la rotonde des donateurs.

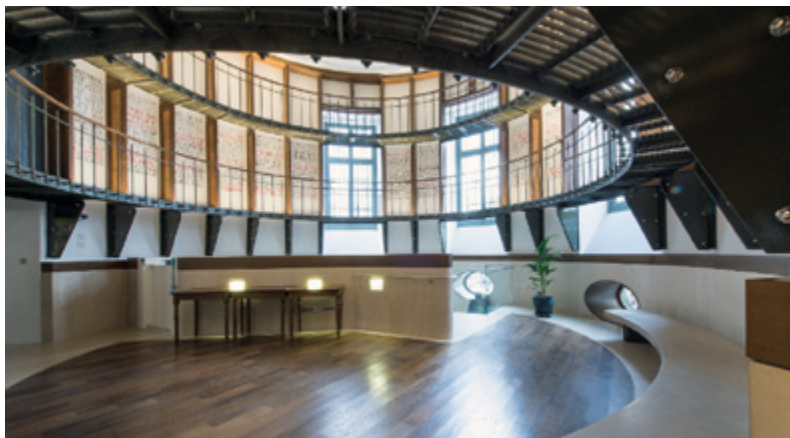
La bibliothèque possède de considérables ressources documentaires et bibliographiques. Ses collections, centrées sur les sciences historiques et philologiques, notamment pour la période médiévale et moderne, se sont formées autour des disciplines enseignées à l'École, en intégrant les évolutions récentes liées au développement des humanités numériques.

### AUX SOURCES DE L'HISTOIRE DE FRANCE

Ses domaines d'excellence sont les sources de l'histoire de France (éditions de sources, instruments de recherche et manuels d'édi-

#### CHIFFRES-CLEFS

- 1 300 m<sup>2</sup> en libre-accès, sur 4 niveaux
- 57 places assises
- 3,7 kilomètres linéaires de collections
- 150 000 volumes
- 1 340 périodiques, dont 474 vivants
- près de 40 000 tirés-à-part
- 60 bases de données, dictionnaires, bouquets de revues et de livres numériques...
- 1 500 livres anciens, dont 12 incunables
- 1 000 estampes



© Jean Christophe Bailot / ENC



↳ Dans la Rotonde des donateurs ainsi que dans les salles réparties dans les différents niveaux, une rénovation en tous points réussie.

complètent cet ensemble : livres anciens, estampes et cartes, ainsi qu'une collection de fac-similés qui fait la singularité de la bibliothèque et de l'École. Pour les ouvrages anciens, mais également pour les domaines d'édition les plus pointus, la bibliothèque s'enrichit au fil des années grâce aux dons des anciens élèves de l'École et d'universitaires, tels Eugène Lefèvre-Pontalis et Léon Pressouyre, dont les collections ont été récemment léguées à l'École.

### TIRÉS-À-PART ET OUVRAGES ANCIENS NUMÉRISÉS

Si les thèses pour le diplôme d'archiviste paléographe sont conservées aux Archives nationales, l'École diffuse en ligne les positions des thèses depuis 2000 et prépare une opération de numérisation des thèses anciennes, en partenariat avec la Comue Paris sciences et lettres (PSL), dont elle est membre. La bibliothèque propose enfin une collection de ressources électroniques – bases de données, dictionnaires et encyclopédies, plateformes de revues et de livres numériques, cédéroms – accessibles sur tous les postes de la bibliothèque et à distance. Elle met en valeur les documents les plus remarquables en les numérisant, à l'exemple des tirés-à-part, de la Revue numismatique et du fonds Casati, constitué d'ouvrages anciens, légués à l'École dans les années 1880 par Claude-Marie Casati, archiviste paléographe, auteur de nombreux articles sur l'art primitif en France, l'Italie, l'Orient et surtout le droit et l'archéologie étrusques.



tion de textes), la codicologie, la bibliographie et l'histoire du livre, la paléographie, la diplomatie, l'archivistique, la philologie latine et romane, l'histoire du droit et des institutions, l'archéologie médiévale, l'histoire de l'architecture et l'histoire de l'art. La bibliothèque conserve de nombreuses suites de monographies et de périodiques très complètes, notamment pour l'histoire régionale et les sociétés savantes, ainsi qu'une importante collection de tirés-à-part d'articles rares de périodiques. Plusieurs fonds spécifiques





Vue de la salle Labrouste. Le magasin central reste relié à la salle Labrouste, avec des niveaux accessibles au public.



Le magasin central Labrouste, BnF/INHA

# Bibliothèque de l'INHA

## 150 000 documents en salle Labrouste

**H**éritière de la bibliothèque d'art et d'archéologie – créée au début du XX<sup>e</sup> siècle par le couturier, collectionneur et mécène Jacques Doucet (1853-1929) –, la bibliothèque de l'Institut national d'histoire de l'art (INHA) s'est enrichie, le 1<sup>er</sup> janvier 2016, de la Bibliothèque centrale des musées nationaux, hébergée jusqu'à par le musée du Louvre. Elle comprend également les collections des Archives de la critique d'art installées à Rennes.

La réunion de ces collections est l'aboutissement d'un projet voulu de longue date : doter la France d'un outil de recherche en histoire de l'art à la hauteur des grandes institutions existantes dans le monde, en particulier aux États-Unis, en Allemagne, en Italie ou au Royaume-Uni.

Désormais installée dans la salle Labrouste et dans le magasin central adossé, la bibliothèque de l'INHA bénéficie d'un cadre particulièrement propice au travail de son public de chercheurs. Elle accueille, en effet, des étudiants en histoire de l'art à partir du niveau master. Outre les étudiants, elle est également conçue pour recevoir les chercheurs en histoire de l'art du monde universitaire ou des musées. Hormis ce public spécialisé, elle souhaite aussi devenir un lieu d'échanges, de découvertes et de rencontres pour tous les amateurs ou érudits ayant besoin de consulter ses collections, conformément à l'esprit de Jacques Doucet.

Le déménagement a été l'occasion d'une profonde transformation pour la bibliothèque : élargissement des horaires d'ouverture,

multiplication par 4 du nombre de places des lecteurs, création d'espaces de travail en groupe, développement de nouveaux services (réservation de places et de documents, scans et service de reproduction améliorés), carte commune avec la BnF, organisation de visites et d'ateliers à partir des collections, accès sans rendez-vous aux collections patrimoniales...

Les efforts de préparation de cette ouverture ont aussi porté sur la politique documentaire, à la fois dans le cadre du dispositif CollEx, mais également avec les bibliothèques de musées et d'écoles d'art. L'offre à distance a été étoffée grâce au développement d'une bibliothèque numérique.

### CHIFFRES-CLEFS

- Plus de 1,7 million de documents
- 622 780 monographies et 7 887 titres de périodiques
- 146 fonds d'archives d'archéologues, d'historiens et d'acteurs du monde de l'art
- 1 800 manuscrits, dossiers et recueils de correspondances
- Plus de 20 000 livres et volumes anciens, rares et précieux
- Plus de 30 000 dessins et estampes
- Plus de 45 000 lettres autographes
- 96 000 cartons d'invitation aux expositions
- 750 000 photographies

La grande nouveauté est le libre accès à 150 000 documents installés dans la salle Labrouste et le magasin central. Pour la première fois de leur histoire, trois niveaux de magasins sont accessibles aux lecteurs, grâce aux efforts conjugués des architectes, de l'OPPIC et des équipes de la bibliothèque, qui ont dû procéder à la sélection, la recotation et au rondage de cette bibliothèque dans la bibliothèque.

De nombreux chantiers restent à mener, dont la reprise des données des notices de la Bibliothèque centrale des musées nationaux et leur versement dans le Sudoc. Néanmoins, les premiers mois d'ouverture ont d'ores et déjà été l'occasion de constater le succès de cette initiative. La fréquentation de la salle de lecture est en hausse continue, tandis que le grand public est au rendez-vous des propositions de visites des espaces rénovés de la bibliothèque.



### INFORMATIONS PRATIQUES

58, rue de Richelieu, 75002 Paris

#### HORAIRES D'OUVERTURE

Du lundi au vendredi, de 9 h à 19 h 30.  
Le samedi, de 9 h à 18 h 30.

#### CONTACT, INFORMATIONS :

[info-bibliotheque@inha.fr](mailto:info-bibliotheque@inha.fr)  
<http://bibliotheque.inha.fr>  
<http://bibliotheque-numerique.inha.fr>  
<https://blog.bibliotheque.inha.fr>

# Journées Abes 2017 : demandez le programme !



Les 10 et 11 mai 2017, pour la dixième année consécutive, les Journées Abes réuniront les professionnels des réseaux au Corum, Palais des Congrès de Montpellier.

**En 2007, considérées comme essentielles à la bonne cohésion de notre communauté professionnelle, leur format ambitieux avait été porté de ses vœux par Raymond Bérard, à qui l'équipe organisatrice les dédie.**

## En préambule...

« Que les échanges entre professionnels soient riches et sympathiques et que la cause de l'accès libre à une information structurée et de qualité avance grâce à ces Journées ! » (David Aymonin)

Après l'ouverture des Journées par **Alain Abecassis**, chef du Service de la coordination des stratégies de l'ESR au MENESR, les Actualités de l'Abes, coordonnées par **Marianne Giloux**, responsable du département Métadonnées et services aux réseaux, déclineront les temps forts de l'activité de l'Agence et feront le point sur les chantiers en cours, chantiers que le comité de programmation a souhaité mettre en perspective lors des sessions plénières, tables rondes et ateliers.

## SGBm, mode d'emploi

La transition vers de nouveaux systèmes de gestion de bibliothèques – hébergés dans le cloud et associés à des plateformes de services – est en marche. Après les témoignages de **François Renaville**, responsable des systèmes documentaires informatisés à l'Université de Liège et de **Pablo Iriarte**, coordinateur du pôle informatique documentaire de l'Université de Genève, dont les établissements respectifs ont franchi le pas, les sociétés titulaires de l'accord-cadre du marché SGBm présenteront les solutions mises au point lors du dialogue compétitif. Les retours d'expérience de chefs de projets des trois vagues de déploiement du SGBm apporteront ensuite aux participants un éclairage concret, en termes d'organisation, de conduite du changement et de prospective.

## Istex : signalement et accès aux ressources

Au regard des résultats de l'enquête « Signalement des ressources Istex »

menée par le consortium Couperin, les travaux de l'Abes et de l'Inist-CNRS, seront mis en lumière : exposition des données Istex en RDF et rôle du hub au service de leur signalement dans le Sudoc, pour les équipes de l'Abes ; mise à disposition d'une plateforme de diffusion et de ses différentes briques opérationnelles (API, TDM...), pour celles de l'Inist-CNRS.

## Une approche renouvelée des collections d'excellence

Le dispositif ColLEX-Persée invite à questionner la notion de « collections d'excellence ». Lors d'une session à visée stratégique, un point d'étape sur les éléments structurants du projet sera proposé par **Sophie Mazens**, cheffe du département de l'IST et des réseaux documentaires au MENESR.

Au cours d'une table ronde, **Jean-Louis Baraggioli**, directeur du CTLe, et **David Aymonin**, directeur de l'Abes, interrogeront le dispositif à l'aune des spécificités de la gestion des ressources continues. Les propos seront étayés par les retours d'expérience de plusieurs acteurs de terrain, membres du réseau Sudoc-PS : qualité des métadonnées, conservation partagée, valorisation des collections, travail en réseaux interinstitutionnels...

## La montée en puissance des référentiels

Au cœur du web de données, les référentiels d'autorité constituent un important capital stratégique. Outre les usages renouvelés d'IdRef qui vient en appui de nombreux services et projets numériques, le projet conjoint ABES-BnF pour mettre en œuvre un Fichier national d'entités témoigne de cette dynamique. **Frédérique Joannic-Seta**, directrice du Département des métadonnées de la BnF, présentera ce projet ambitieux : s'inscrivant dans le mouvement global du renouvellement des modèles de données à l'heure de la Transition bibliographique, il s'agit de renouveler leur mode de production à l'échelle nationale par une mise en œuvre conjointe et coordonnée.

La session « Le signalement augmenté : des algorithmes au service des catalogueurs », coordonnée par **François Mistral**, responsable IdRef, sera le versant concret de ces problématiques complexes. Un tour d'horizon ludique et interactif des outils et process développés en ce sens par l'Abes.

## Les réseaux de catalogage au cœur de la transition

À partir des résultats de deux enquêtes lancées en 2017 – l'une par l'Abes, « Usages professionnels du Sudoc » ; l'autre par l'ADBU, « Stratégie et organisation du signalement » – et de la présentation de chantiers en cours, la session parallèle, animée par **Valérie Neouze**, responsable de la commission ADBU-SSI, et **Laurent Piquemal**, responsable du service Gestion des réseaux, donnera l'occasion aux professionnels du réseau Sudoc d'échanger sur l'impact de ces évolutions sur leurs pratiques.

## Au rendez-vous des convivialités

L'Abes tient à remercier chaleureusement les sociétés sponsors pour leur précieuse contribution à la réussite des Journées. Grâce à leur soutien financier, les moments de pause, au cours desquels les participants ne manqueront pas de visiter leurs stands, sont autant d'occasions d'échanges et de prises de contact fructueuses. Sans oublier le temps réservé à la visite du salon professionnel, où chacun pourra également découvrir l'exposition de posters... et qui précédera le fameux cocktail dinatoire, agrémenté cette année par les notes enjouées du dynamique groupe montpelliérain Carolina Reapers Swing !!!

**CHRISTINE FLEURY**

Chargée de communication et de l'organisation des Journées Abes  
fleury@abes.fr

## Programme complet :

<http://www.abes.fr/Publications-Evenements/Journees-ABES/Journees-ABES-10-11-mai-2017>

## Sara BERNARD

Responsable du système d'information documentaire de l'université de Caen  
Coordinatrice thèses de Normandie Université

### **Parlez-nous de vos fonctions actuelles...**

Je suis conservatrice à l'université de Caen, une université pluridisciplinaire avec un SCD composé de 5 BU et d'une douzaine de bibliothèques associées. J'y assume la direction d'un département transversal, en charge des métadonnées et des applications documentaires : le département du système d'information documentaire. Je suis par ailleurs « coordinateur thèses », pour l'université de Caen et la COMUE Normandie Université.

### **Quelles sont les étapes qui vous semblent les plus importantes dans votre parcours professionnel ?**

J'ai toujours travaillé autour des questions de SIGB et d'informatisation : dès le début de ma carrière en 1995, alors que j'étais chargée d'études documentaires au Centre régional de documentation pédagogique de Lorraine, et encore aujourd'hui. J'ai ainsi été adjointe à la directrice de la BU Droit-Lettres à mon arrivée au SCD en 2003, et je suis coordinatrice pour les thèses depuis 2013 : un dossier qui m'a ouvert de fructueuses collaborations avec les autres services de l'établissement, et maintenant avec les autres établissements de la COMUE.

Au-delà des dossiers et des postes occupés, je retiens aussi de belles rencontres professionnelles, en lien direct avec mes fonctions comme dans d'autres cadres. Par exemple, à l'occasion du congrès de l'IFLA en 2014, où j'ai participé à l'encadrement des bénévoles, ou lors de l'échappée cycliste Cyclo-Biblio, la même année.

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

Le SCD de Caen était déjà déployé dans le Sudoc à mon arrivée. Après avoir suivi avec envie cette évolution depuis l'extérieur, j'étais ravie d'en voir le fonctionnement de l'intérieur. Mais mes premiers contacts directs avec l'Abes n'eurent lieu que deux ans après, en 2005, autour de problématiques liées au PEB : Pebnet existait encore mais devait être remplacé à court terme par Supeb, ce qui suscitait des interrogations dans le réseau. Je suis aussi restée marquée en raison d'un violent épisode orageux sur Montpellier, qui avait obligé à écarter une réunion qui se tenait dans les locaux de l'Abes et rendu épiques les conditions du retour...

### **Participez-vous à un groupe de travail spécifique au sein de l'agence ?**

Pas spécifiquement, mais je travaille de manière de plus en plus rapprochée avec l'Abes : depuis 2013 avec l'équipe des thèses, de manière encore plus marquée l'année dernière



en raison de la reprise de compétence par la COMUE. Je me suis également rapprochée de l'équipe SGBM ces derniers mois du fait que Caen et Rouen ont rejoint les sites pilotes à l'automne. Cela nous permet de faire partie du même marché subséquent que l'Université du Havre, avec d'autres sites pilotes, et ainsi de garantir l'installation d'une même solution à l'échelle de la Normandie. Nous espérons ainsi disposer d'un socle commun facilitateur pour la mutualisation de services.

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

De l'anticipation, toujours, mais aussi des choix : pour maintenir son identité et ne pas risquer de voir se diluer son action. Avec une attention particulière sur la cohésion, notamment avec les établissements du réseau dans cette nouvelle ère de relations qu'ouvre le SGBM.

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

Comme beaucoup d'entre nous, j'apprécie avant tout d'être au service du public. En ce qui concerne le poste que j'occupe, c'est la dynamique de projet qui me porte en ce qu'elle est œuvre de construction (d'outils, en général), et que son importance réside tout autant dans le processus de réalisation que dans la production finale : ce sont des projets extrêmement fédérateurs pour les équipes.

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

Je parlerais de perplexité plutôt que d'énervement, face à la différence de tempo entre les évolutions technologiques et le temps d'acculturation dans nos établissements.

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

D'autres ont déjà évoqué une ruche ou une fourmilière. À défaut de comparaison animale originale, je pencherais pour l'image d'un chef d'orchestre : l'Abes rassemble les musiciens que sont les établissements, en guide le mouvement et rehausse la qualité d'ensemble en impulsant sa propre énergie et sa créativité.

### **Votre expression favorite ?**

« Pour voir loin, il faut y regarder de près », Pierre Dac.

sara.bernard@unicaen.fr

10-11 mai 2017  
Montpellier  
Corum



# *Les Journées*

abes ;  
agence bibliographique  
de l'enseignement supérieur

Pack Or  
**SPRINGER  
NATURE**

**WILEY**

**BRILL**



**ACCUCOMS**

**SAGE  
Publishing**

Pack Argent  
**RITME**  
SCIENTIFIC SOLUTIONS

**Science**  
AAAS

**IOP Publishing**

Pack Cristal

**CAMBRIDGE  
UNIVERSITY PRESS**