

# Ar(*abes*)ques

JUILLET - AOÛT - SEPTEMBRE 2016

DOSSIER

## Créer du lien, faire sens *Un nouveau souffle sur les données*

PLEINS FEUX SUR • Le Cnam, un ensemble « révolutionnaire »

ACTUALITÉS • Retour sur les Journées UKSG 2016 • BiblioDebout



## (Dossier) CRÉER DU LIEN, FAIRE SENS *Un nouveau souffle sur les données*

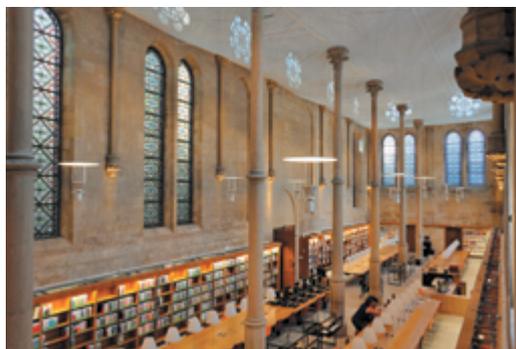
La communauté des bibliothèques, aux côtés d'institutions nationales et internationales, est aujourd'hui amplement engagée dans le web de données ; les standards de description bibliographique, et par là même les catalogues, évoluent vers de nouveaux modèles plus souples permettant de s'inscrire pleinement dans l'environnement du *Linked open data*. À travers quelques exemples, le dossier illustre les nombreux avantages de ces évolutions : facilité d'accès, de publication, de récupération, de mise en relation et d'enrichissement des données... Ce dossier est notamment l'occasion pour l'Abes de revenir sur ses premiers pas dans le web de données mais aussi de présenter ses derniers travaux en la matière : la mise en ligne d'une vaste base de données RDF, « défense et illustration pédagogique d'une approche web sémantique de l'interopérabilité des données de l'IST ». Bonne lecture !

Merci à Michael Jeulin et Yann Olivier, membres de l'équipe Hub de métadonnées à l'Abes, pour leur contribution précieuse à ce dossier.

## 22 (Pleins feux sur...)

### Le Cnam, un ensemble

«révolutionnaire» FRÉDÉRIQUE GAUDIN



© Marie Jarrige, Université de Corse Pascal Paoli

24

(Actualités)

28

(Portrait)

04 Le web de données, de « l'information en réseau »  
THOMAS FRANCAERT

06 L'Abes sur le web de données  
MICHAEL JEULIN

07 Le hub de métadonnées  
YANN OLIVIER

09 Archives et web de données : contextes et perspectives  
JEAN-MARIE FEURTET

10 Bibframe, un nouveau modèle de données pour les bibliothèques  
REINHOLD HEUVELMANN

12 PRESS<sub>00</sub> : décrire les ressources continues dans le web de données  
PATRICK LEBŒUF - CLÉMENT OURY

14 Le web de données à la BNF  
JÉRÔME VILLEMINOZ

16 Le web, renaissance de la publication de recherche  
ROBIN BERJON

18 Triphase : co-construction d'une ressource termino-ontologique  
AGNÈS GIRARD - CLAIRE NÉDELLEC

20 *Linked data*, enjeu(x) et devenir  
MICHAEL JEULIN - ALINE LE PROVOST - YANN OLIVIER

## Ar(abes)ques

REVUE TRIMESTRIELLE DE L'AGENCE BIBLIOGRAPHIQUE DE L'ENSEIGNEMENT SUPÉRIEUR  
227 avenue Professeur Jean-Louis Viala CS 84308 - 34193 Montpellier cedex 5  
T. 04 67 54 84 10 / F. 04 67 54 84 14 - [www.abes.fr](http://www.abes.fr)  
Directeur de la publication : Jérôme Kalfon  
Coordination éditoriale et secrétariat de rédaction :  
Marion Grand-Démery ([grand-demery@abes.fr](mailto:grand-demery@abes.fr))  
Comité de rédaction  
Jean-Marie Feurtet, Christine Fleury, Marion Grand-Démery, Philippe Le Pape, Cyril Leroy,  
Isabelle Martin, Maryse Picard, Marie-Pierre Roux, Bertrand Thomas  
Conception graphique : Anne Ladevie ([anneladevie.com](http://anneladevie.com))  
Impression : Pure Impression  
Couverture : Dandelion [Karahindida, TR] - Phot. Esin Ustün / CC-BY 2.0

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

# Pourquoi le web sémantique ?

**E**n mode de fonctionnement traditionnel, avec les outils utilisés actuellement, les technologies du web sémantique n'apportent que peu de résultats directement tangibles. Et pourtant l'Abes s'est engagée depuis plusieurs années dans l'aventure du web sémantique. Aujourd'hui elle passe un cap important avec la mise en ligne d'un démonstrateur. Il s'agit moins d'offrir un nouveau service que de montrer de façon pédagogique l'interopérabilité des données de l'IST. Ces actions sont le fruit d'un long travail, sinon invisible, du moins discret. Quoi qu'il en soit, ne nous attendons pas à tout voir. Ces services ne sont pas et ne seront pas directement accessibles au large public. C'est aussi l'une des difficultés de ces objets et des ces technologies. Complexes, elles ont du mal à être appréhendées. Souterraines pour la plupart,

préexistantes et de cultures « métier » (bibliothèques, archives, édition, etc.) – chacune s'organisant alors selon sa propre logique. La masse des données héritées, l'effort nécessaire à la convergence, ou plus simplement l'absence d'outils adaptés sont autant de freins aux évolutions.

La qualité des métadonnées est désormais pour l'essentiel entre les mains des producteurs, donc pour l'ensemble des données nativement numériques,



des éditeurs. À terme, les bibliothécaires n'ont vocation à agir qu'en début de chaîne et uniquement pour les documents préexistants. Certes, ils peuvent agir en tant que « réparateurs » pour les documents dont les métadonnées ne sont pas de qualité suffisante, mais pour l'essentiel, ils subiront un mécanisme de désintermédiation.

## Adopter les technologies du web sémantique, c'est favoriser l'ouverture des données.

on n'en perçoit pas directement les bénéfices. Et pourtant elles sont indispensables.

Passer au web sémantique, c'est offrir la capacité de lier des données entre elles. Pour pouvoir les lier, il faut pouvoir y accéder, et pour que l'on puisse y accéder, il faut les exposer. Donc, adopter les technologies du web sémantique, c'est favoriser l'ouverture des données. Un atout de taille pour la Science ouverte. Ces technologies sont aussi au cœur des évolutions de l'Abes, indispensables pour remplir la nouvelle mission confiée à celle-ci : constituer un entrepôt national de métadonnées.

Si l'on voulait résumer les enjeux du web sémantique, on pourrait dire qu'il s'agit de faire converger de très nombreux acteurs, aux pratiques extrêmement diverses. Une convergence qui ne se décrète pas, qui n'est pas imposée, mais construite pragmatiquement. A chacun de s'adapter ou de courir le risque de rester à l'écart.

Et il n'est pas facile de faire converger une multiplicité d'acteurs : chacun a ses exigences légitimes et ses contraintes. Les barrières ne sont pas seulement techniques, elles sont aussi liées à l'héritage de données

Une désintermédiation qui pourrait tout aussi bien toucher les éditeurs. C'est en tout cas ce que nous promet le responsable de la plateforme de publication science.ai, Robin Berjon : grâce à l'automatisation des traitements, les frais de publication se trouveraient réduits à tel point que le débat sur le libre accès deviendrait caduc. Espérons qu'il dise vrai, et que le projet ne se trouve pas justement dans sa phase de « pic des espérances exagérées » propre au « hype cycle » évoqué par Reinold Heuvelmann à propos de Bibframe. Mais il est difficile de prédire à l'avance quand une technologie arrive à maturité. Nous espérons que cette livraison d'Arabesques vous permettra non seulement de constater la diversité des états de l'art, mais aussi de voir à quel point les initiatives foisonnent dans un domaine où les professionnels de l'IST restent dans la course, se préparent à l'exercice de nouveaux métiers et s'investissent dans des services qui seront de plus en plus directement intégrés aux pratiques pédagogiques et de recherche.

JÉRÔME KALFON  
Directeur de l'Abes

# Le web de données, de « l'information en réseau »

**Le web de données (on préférera ce terme à celui plus ambigu de web sémantique), ce n'est pas compliqué; ça marche et c'est utile, en particulier pour les bibliothèques.**

## RETOUR AUX RACINES DU WEB

Le web n'a pas été conçu pour n'être qu'un paquet de documents mis en lien. Il intègre, dès sa conception en 1989 par Tim Berners Lee<sup>1</sup>, plus de sémantique que l'utilisation qui en sera faite ensuite. En particulier par la dualité Identifiant/Représentation :

- Identifiant : ce qui commence par « <http://...> » et que l'on voit dans la barre d'adresse de notre navigateur est une URL, où le « L » est mis pour « Locator ». C'est donc l'adresse d'un document sur le web ; mais ce n'est qu'un cas particulier des URI, où le « I » est mis pour « Identifier », qui sont des **identifiants, dans le contexte du web, de choses du monde réel**. On comprend donc qu'on peut identifier sur le web n'importe quoi à l'aide d'une URI : Victor Hugo, les pizzas *margherita*, le terme de thésaurus « gouvernance », la Loire, la caractéristique « se situe à », etc. On parle d'une façon générale de **ressources**.

- Représentation : si une URI est l'identifiant d'une « ressource », alors quel « document »

obtiendra-t-on en naviguant vers cette URI ? On a l'habitude d'obtenir pour une même adresse toujours le même document, mais d'une façon générale un identifiant peut être associé à **plusieurs représentations** qui varient – de façon transparente – en fonction de préférences de langue, de format, de lieu, etc. C'est ce qu'on appelle la **négociation de contenu**.

Cette capacité des URI à identifier absolument n'importe quoi, indépendamment d'une représentation particulière, est la clé de voûte de l'universalité du web (de données).

Une fois les « choses » identifiées et rendues indépendantes des documents qui les représentent, il devient possible de parler de celles-ci : on peut publier sur le web l'assertion que « La Tour Eiffel se situe à Paris », en utilisant 3 URI pour identifier les 3 composantes de cette assertion : La Tour Eiffel, la notion de « se situer à », et Paris. C'est le standard RDF (*Resource Description Framework*) qui permet d'employer ces assertions en triplets. Notons au passage que, le web étant par nature décentralisé, n'importe quel est libre :

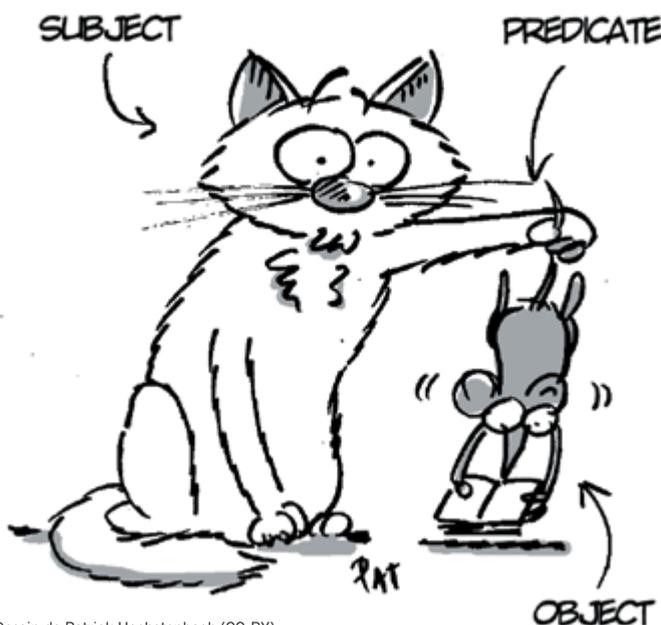
- de créer une nouvelle URI pour identifier Paris ;
- ou de créer une assertion en se référant à une URI déjà existante pour Paris (par exemple celle définie par l'Insee<sup>2</sup>) ;
- ou encore d'exprimer des liens d'équivalence entre identifiants : l'URI que l'on définit pour Paris représente la même « chose » que celle définie par l'Insee.

On voit donc se dessiner ce qui nous occupe : un réseau décentralisé de données liées par des triplets.

Mais il faut aller plus loin pour que l'interopérabilité soit complète – puisque le web de données n'est qu'une solution à la problématique de l'interopérabilité. En effet, pour qu'une autre application puisse décoder mon assertion, il faut que je donne une définition précise des identifiants que j'ai utilisés, qui sont sans doute différents de ceux que comprend cette application. En particulier, il faut que je donne une définition précise de mes « verbes » (« est situé à »)

[1] Voir l'article de référence sur le sujet : Tim Berners-Lee, James Hendler and Ora Lassila, « The Semantic Web », *Scientific American*, Mai 2001.

[2] URI de Paris définie par l'Insee : <http://id.insee.fr/geo/commune/75056>, voir <http://rdf.insee.fr/geo>



Dessin de Patrick Hochstenbach (CC-BY)  
Source : <http://librecat.org/>

et mes « types » (Lieu, personne, etc.). C'est ce que permettent les **ontologies**, dont l'objectif est de donner un sens univoque à ce dont je parle, à l'aide de la logique formelle (on parle également de **vocabulaire** ou de **modèle de données**, un peu par abus de langage). Les ontologies permettent également de déclarer des équivalences entre verbes ou entre types, rendant ainsi interopérables des données hétérogènes. Par exemple, je peux dire que, dans mon contexte, « est situé à » relie quelque chose à un « Lieu » et que cela représente la même notion que l'identifiant « *basedNear* » défini dans une autre ontologie bien connue, FOAF<sup>3</sup>.

Les ontologies font donc émerger de cet océan de liens des structures interopérables, rendant ainsi les données liées plus « sémantiques », c'est-à-dire plus facilement réutilisables.

## QUELS ENJEUX ET QUELLES CONSÉQUENCES ?

Souvenons-nous des fausses promesses entendues au milieu des années 2000 à propos du web de données : les machines allaient bientôt comprendre le sens des textes, on nous parlait de web 3.0, de « Twine » (un site qui n'existe plus maintenant mais qui promettait la révolution des réseaux sociaux), on cherchait quelle serait la « *killer-app* » – une application si attrayante qu'elle aurait justifié la technologie à elle seule, etc. Rien de tout cela n'est arrivé, mais d'autres conséquences ont eu lieu.

D'abord une quantité grandissante de « données ouvertes et liées » publiées par une variété importante de producteurs de données : c'est le fameux « *Linked Open Data* »<sup>4</sup>. Citons-en quelques points notables : DBPedia francophone (une extraction en RDF des données de Wikipedia), data.bnf.fr (notices FRBRisées – voir plus bas –, autorités et thématiques de la Bibliothèque nationale de France), ou encore VIAF (Virtual International Authority File, une mise en commun des fichiers d'autorité d'une quarantaine de bibliothèques et de musées).

Dans cet ensemble de données, il faut en mentionner certaines ayant un statut particulier : les thésaurus. Ceux-ci peuvent être représentés et publiés dans le modèle SKOS. Ce modèle permet d'aligner les thésaurus pour permettre l'interopérabilité de catalogues documentaires utilisant des vocabulaires d'indexation différents (« Désobéissance civile » dans Rameau est ainsi rapprochée de « *Civil disobedience* » dans les sujets de la Bibliothèque du Congrès américain<sup>5</sup>). Quant aux ontologies, on se référera au projet LOV – *Linked Open Vocabularies*<sup>6</sup>. Ensuite, une appropriation de cet enjeu des données structurées et liées par les grands moteurs de recherche : c'est l'initiative schema.org<sup>7</sup>, qui propose un modèle de description de « plein de choses dont on parle sur le web » (blogs, livres, films, produits,

etc.), compréhensible par Google, Yahoo, Bing et consorts. On peut reprocher à schema.org son biais vers le e-commerce, sa vision occidentaliste et son manque de transparence dans la gouvernance, mais si les bibliothèques souhaitent rendre leurs données plus visibles par les moteurs, cela passe par la publication de données compatibles avec schema.org.

D'une façon plus profonde, ces technologies induisent une représentation générale de l'information en **graphe décentralisé**, en réseau. Ce mode de structuration, de pensée, fait suite à celui plutôt tabulaire des bases relationnelles, et plutôt hiérarchique de XML. La conséquence est flagrante sur les notices bibliographiques avec le modèle FRBR. Les *Functionnal Requirement for Bibliographic Records*, successeurs de l'ISBD (*International Standard for Bibliographic Record*) proposent en effet un éclatement de la notice en 4 niveaux conceptuels, eux-mêmes reliés aux personnes ou aux organisations impliquées dans la vie du document (auteur, contributeur, éditeur, possesseur), lesquelles sont elles-mêmes reliées entre elles ou à d'autres données du web.

Cette tendance est à rapprocher du constat que de plus en plus de systèmes informatiques de diffusion des catalogues utilisent une base de graphe RDF (« *triplestore* ») pour centraliser les métadonnées des notices FRBRisées, les fiches d'autorité et les thésaurus. Cette base devient le pivot central des canaux de diffusion (sites web, flux RSS, formats d'échange métier, etc.). Les lois européennes sont notamment diffusées sur ce mode, via la base Cellar et le portail Eur-Lex<sup>8</sup>.

## PROCHAINES PROMESSES ?

Sans retomber dans les promesses hasardeuses évoquées plus haut, on peut néanmoins esquisser les lignes de force du web de données pour les prochaines années : une utilisation grandissante de schema.org par les moteurs de recherche et les projets de diffusion de données ; l'intégration native des fonctions de publication et de récupération des données du web dans les *Content management system* (CMS) et les SIGB ; la publication et l'alignement de plus en plus de données – dont des thésaurus ou des données de la recherche ; la généralisation de FRBR et de ses dérivés pour la description des notices bibliographiques, etc.

Au-delà des aspects technologiques, ce sont des logiques de partage, de réutilisation, de mise en réseau, de collaboration, ou d'insertion dans un écosystème d'acteurs, qui sont favorisés par cet artefact unique qu'est le web de données.

**THOMAS FRANCAERT**  
Consultant chez Sparna  
thomas.francart@sparna.fr

[3] FOAF : <http://xmlns.com/foaf/spec/>

[4] *Linked Open Data* : <http://linkeddata.org/>

[5] En triplet RDF : <http://data.bnf.fr/ark:/12148/cb12049451f>

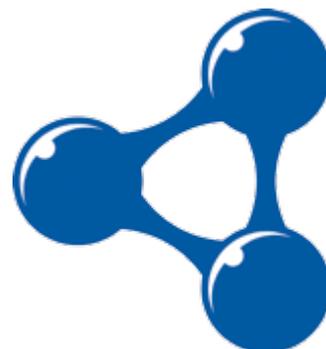
skos:closeMatch  
<http://id.loc.gov/authorities/subjects/sh90000103>

[6] LOV : <http://lov.okfn.org>

[7] <http://schema.org>

[8] EurLex : <http://eurlex.europa.eu>

# L'Abes sur le web de données



C'est dans une démarche très progressive, empirique et pragmatique, qu'à partir de 2007, l'Abes a fait ses premiers pas sur le web de données.

**Au départ, le principal objectif était d'améliorer le référencement des données par les moteurs de recherche**, tout en alimentant une réflexion à plus long terme sur l'interopérabilité et la « réutilisabilité » de ces données, au-delà des formats métiers traditionnels. C'est donc par l'exposition de celles-ci que tout a commencé.

## IDENTIFIER : LA PREMIÈRE BRIQUE

La première étape a consisté à identifier nos ressources en tant que telles. Car si nos catalogues sont riches en identifiants, internes ou normalisés (PPN, ISSN, ISBN, NNT...), ils ne permettent pas, à eux seuls, un référencement par les moteurs de recherche puisque ces derniers parcourent essentiellement des pages web en sautant de lien en lien. Il fallait donc, avant tout et pour chacun d'eux, identifier chaque ressource ou notice par des URL ou URI stables, construites sur les identifiants internes et à partir desquelles une redirection permet d'assurer l'affichage de la page de résultats correspondante. Puis il s'agissait de les lister systématiquement dans des « Sitemaps », pages à l'usage des robots d'indexation, leur permettant ainsi de les référencer comme des pages web.

Après être descendu au niveau des notices, l'étape suivante devait donner lieu à une meilleure indexation du contenu de ces pages. C'est à partir de cette étape que nous sommes donc véritablement entrés dans le web de données, lequel s'appuie, précisément, sur des URI pour identifier et décrire le contenu de façon à pouvoir être lu par une machine. Dès lors, il était naturel de s'intéresser à ce standard émergent du W3C qu'était RDF.

## L'APPRENTISSAGE DE RDF

C'est avec Calames, en 2008, que nous avons commencé à distiller, l'air de rien, des métadonnées en RDF sous sa forme encore la plus répandue : du RDFa, c'est-à-dire des triplets encapsulés dans des balises cachées du code HTML. Ces balises, ignorées par les navigateurs, peuvent être moissonnées par des « parseurs<sup>1</sup> » spécialisés ainsi que par les moteurs de recherche. C'est également une solution de moissonnage des données, alternative à OAI-PMH. C'est d'ailleurs celle retenue par le portail Isidore<sup>2</sup> jusqu'à aujourd'hui.

Avec IdRef, ouvert en octobre 2010, nous avons choisi une exposition distincte de l'interface publique, en exposant des fichiers RDF/XML. Pour les récupérer, nul besoin de parser du code HTML : avec un navigateur, il suffit pour récupérer le fichier d'ajouter à l'URI de la notice l'extension **.rdf**. En voici un exemple : <http://www.idref.fr/033702462.rdf>. Et pour un programme, il est possible avec l'URI seule de demander ce fichier RDF, plutôt que la redirection par défaut vers la page HTML, grâce à la négociation de contenu dans la requête HTTP<sup>3</sup>. Le portail theses.fr, ouvert l'année suivante, a retenu les deux méthodes : RDFa, et RDF/XML. Et finalement, le Sudoc a fait à son tour son apparition sur le web de données. À ceci près que pour ce dernier, nous avons fait une entorse à la standardisation en proposant également dans des pages HTML spécialement destinées aux robots des microdonnées schema.org, promues par les principaux moteurs de recherche.

## RDF INSIDE

Les données exposées collent encore de près aux formats de production sur lesquels elles s'appuient, tout en étant incomplètes et parfois bancales. En effet les vocabulaires les plus répandus sont mal adaptés aux données natives, et les vocabulaires « métiers » (ISBD, RDA, FRBR) ne sont pas toujours bien adaptés au web de données. Surtout, les données sont générées dynamiquement à partir des bases de production, auxquelles elles sont par conséquent étroitement asservies.

Mais ces expériences ont permis de monter en compétences et d'être aujourd'hui plus ambitieux.

Avec le hub de métadonnées, nous avons commencé à interroger et à manipuler les données en RDF dans une base autonome, afin d'explorer plus directement le potentiel des graphes et du langage SPARQL. Enfin, RDF est un bon candidat pour un futur entrepôt de métadonnées ouvertes, synchronisées entre elles, mais aussi avec les bases de production, ainsi qu'avec des référentiels et sources externes.

**MICHAEL JEULIN**

Expert métadonnées, Abes  
jeulin@abes.fr

[1] Programmes permettant de récupérer dans une structure de données – XML, RDF – des contenus (de balises par exemple) et de les rendre accessibles.

[2] Plateforme de recherche et d'accès aux données numériques et numérisées du domaine des sciences humaines et sociales.  
<http://www.rechercheisidore.fr/>

[3] Possibilité, pour une même URI, de proposer plusieurs versions d'un document. Exemple avec l'interface cURL: curl-H «Accept: application/rdf+xml»  
<http://www.sudoc.fr/157385477>

# Le hub de métadonnées

Quand on parle de hub de métadonnées à l'Abes, il faut imaginer un atelier de retraitement en masse des données des éditeurs pour en tirer le maximum : en conserver toute la richesse, les enrichir encore et propager ces données de qualité dans différents environnements.

## RÉCEPTIONNER, DÉBALLER, EXAMINER

Dans le cadre d'Istex<sup>1</sup>, nous recevons des fichiers XML très riches qui contiennent à la fois les métadonnées et le document. Chaque fichier correspond à un article de revue ou à un chapitre de livre : un niveau de granularité plus fin que celui du Sudoc. Ce fichier XML est réputé conforme à un vocabulaire XML, formalisé dans une DTD<sup>2</sup>. Ce vocabulaire est souvent propriétaire, mais, pour les articles, il est de plus en plus souvent conforme à Jats<sup>3</sup>, format initié par la National Library of Medicine (NLM) – Bibliothèque nationale de médecine américaine. Cela fait bien nos affaires.

Pourtant, la DTD ne nous suffit pas. Nous voulons en savoir plus sur le contenu de ces fichiers XML. D'abord, à partir d'un échantillon de 50 000 fichiers, nous listons tous les chemins (Xpath) présents dans les fichiers livrés. En effet, une DTD peut prévoir des chemins qui ne sont jamais suivis. Inutile alors de les modéliser en RDF.

Ensuite, pour chaque chemin, nous listons les valeurs distinctes qu'il contient et leur nombre d'occurrences. Par exemple, pour un champ sujet, ces facettes permettent de voir rapidement s'il s'agit de mots clés ou d'un vocabulaire contrôlé.

Enfin, pour des requêtes complexes, il reste possible d'interroger la totalité du corpus chargé dans une base XML (requête SQL/Xpath ou Xquery).

## MODÉLISER ET CONVERTIR EN RDF

Grâce à cette radiographie précise des données XML livrées par l'éditeur, la modélisation et la conversion en RDF sont facilitées. Sauf exception (normalisation des dates ou des codes langue, par exemple), lors de cette opération, on ne modifie pas les données d'origine. Mais le simple fait de passer les données en RDF oblige à les rentrer dans un cadre conceptuel explicite. Au lieu d'une arborescence XML qui n'a guère de signification en soi, on se retrouve avec un réseau d'entités : œuvres<sup>frbr</sup>, manifestations<sup>frbr</sup>, personnes, organismes, sujets, collections, etc. De surcroît, les différentes DTD convergent vers un modèle unique. Pourtant, face à une information atypique ou spécifique à un éditeur, la souplesse de RDF permet d'inventer des propriétés *ad hoc*, histoire de ne rien perdre des données d'origine.

## ENRICHIR LES DONNÉES RDF

C'est seulement dans la base RDF que vont s'opérer les enrichissements, c'est-à-dire des apports suc-

cessifs de nouvelles couches d'information, sous la forme de triplets RDF qui s'ajoutent et se connectent aux triplets d'origine.

Dans le cas de Springer, les données sur les auteurs sont assez riches (nom, titres, affiliations, mail) mais on aimerait associer l'auteur à un identifiant, IdRef mais pas seulement (ISNI, Vial, Orcid, etc.). Pour ce faire, on emploie deux types de moyens : soit on « dérive » cette information d'une autre source (Sudoc et ses identifiants IdRef, Worldcat et ses identifiants LC<sup>4</sup> –ou Vial, Orcid) ; soit on « calcule » un alignement grâce aux algorithmes de Qualinca<sup>5</sup>. Voilà ce que cela donne :

```
<http://hub.abes.fr/springerB/ebook/3540113185/w> vivo:relatedBy
  <http://hub.abes.fr/springerB/ebook/3540113185/authorship/1> .

<http://hub.abes.fr/springerB/ebook/3540113185/authorship/1> vivo:relates
<http://hub.abes.fr/springerB/ebook/3540113185/kenkrevasudev> .

<http://hub.abes.fr/springerB/ebook/3540113185/kenkrevasudev>
  owl:sameAs      <http://www.idref.fr/11962558X/id> .
```

Toujours chez Springer, tous les chapitres et tous les articles sont indexés au moyen d'un vocabulaire contrôlé maison. Nous avons nous-mêmes établi une correspondance avec les concepts de LCSH (*Library of Congress Subject Headings*, autorités de la Bibliothèque du Congrès) et de Rameau (répertoire d'autorités de la Bibliothèque Nationale de France, donc en français). Cette opération intellectuelle est délicate et coûteuse, mais le jeu en valait la chandelle étant donné les bénéfices : un millier de concepts alignés pour indexer correctement des centaines de milliers de documents. Cette indexation ne vaut pas celle d'un catalogueur, mais elle aurait nécessité une quantité de travail humain sans commune mesure.

```
<http://hub.abes.fr/springerB/ebook/3540113185/w> hub:classification
<http://hub.abes.fr/springer/ebooks/3540113185/concept/P21050> .

<http://hub.abes.fr/springer/ebooks/3540113185/concept/P21050>
  skos:exactMatch<http://hub.abes.fr/springer/concept/P21050> .

<http://hub.abes.fr/springer/concept/P21050> skos:exactMatch
<http://id.loc.gov/authorities/subjects/sh85134783> .

<http://hub.abes.fr/springer/concept/P21050> skos:exactMatch
<http://data.bnf.fr/ark:/12148/cb11933671c> .
```

[1] Le projet Istex (Initiative d'excellence de l'information scientifique et technique) s'inscrit dans le programme « Investissements d'avenir », initié par le ministère de l'Enseignement Supérieur et de la Recherche. Ce projet ANR-10-IDEX-0004-02 est porté par quatre partenaires : le CNRS, l'Abes, Couperin et l'Université de Lorraine agissant en son nom propre et pour le compte de la CPU.

[2] *Document type definition* (DTD), ou définition de type de document, est un document permettant de décrire un modèle de document (XML par exemple).

[3] Format XML : *Journal Article Tag Suite*

[4] *Library of Congress* (Bibliothèque du Congrès).

[5] Projet de recherche qui s'intéresse à la « qualité et l'interopérabilité de grands catalogues documentaires ». Lire l'article du n°77 d'*Arabesques* : « Faire le lien, un besoin vital », Aline Le Provost. <http://www.abes.fr/Arabesques/Arabesques-n-77>

## REDISTRIBUER

Un hub est le contraire d'un cul-de-sac. Ses données, d'origine ou non, doivent en sortir pour alimenter tous les outils qu'utilisent les bibliothèques et surtout leurs usagers.

Il s'agit d'abord du Sudoc : le hub injecte dans celui-ci des métadonnées correspondant au niveau ebook ou revue - mais ni les chapitres, ni les articles. Les notices créées ou enrichies par le hub sont ensuite encore améliorées par les catalogueurs, notamment dans le cadre du dispositif « Cercles »<sup>6</sup>.

A terme, ces métadonnées seront également récupérées dans les outils de découverte. Une expérimentation est actuellement en cours autour des

données acquises dans le cadre du programme Istex. Les enrichissements sont également versés dans le moteur de recherche de l'API Istex, développée par l'Inist.

Enfin ces métadonnées RDF ont naturellement vocation à contribuer à cet immense réservoir de données ouvertes et liées qu'est le web de données... en attendant que tous les éditeurs le fassent, ou quand ils le font, pour y ajouter notre grain de sel, à savoir nos enrichissements.

YANN OLIVIER

Chef de projet Hub de métadonnées, Abes  
nicolas@abes.fr

[6] Correction et Enrichissement par le Réseau de Corpus de l'Enseignement Supérieur <http://www.abes.fr/Sudoc/Produire-dans-le-Sudoc/Le-dispositif-CERCLES.eu/>



## METTRE NOS DONNÉES EN RÉSEAU – UN DÉMONSTRATEUR

Le 16 mai 2016, l'Abes a mis en ligne une base de données RDF, interrogeable en SPARQL. Il ne s'agit pas d'un nouveau service, mais d'un démonstrateur, défense et illustration pédagogique d'une approche « web sémantique » de l'interopérabilité des données de l'IST.

Cette base agrège toutes sortes de données hétérogènes, qui sont exploitées dans des études de cas publiées sur Punktokomo, le blog technique de l'Abes<sup>1</sup> :

- **Nature en VOSTFR** = Interroger les articles de la revue Nature en MeSH et en Français

- **Les revues d'Oxford UP et la classification JEL (économie)** = Interroger les articles d'une revue d'économie au moyen d'un thesaurus spécialisé multilingue

- **Les ebooks Springer, IdRef, RAMEAU, Dewey** = enrichir les métadonnées d'ebooks et de chapitres fournies par l'éditeur Springer

- **Le même auteur dans IdRef, VIAF, HAL, Persée, etc.** = consolider la production scientifique d'un chercheur grâce aux référentiels de personnes

- **Mapping entre structures de recherche de Paris 4 : IdRef/RNSR/HAL** = identifier un laboratoire dans différents référentiels

- **Matrice des fascicules pour conservation partagée** = générer un tableau de tous les fascicules d'une revue, grâce aux métadonnées d'articles

- **Le bouquet des ebooks Dalloz** = se localiser sous un bouquet, pour exemplariser les ebooks qui le composent

- **La fédération a de l'avenir** = lancer une recherche SPARQL qui attaque deux bases différentes.

Chaque étude de cas correspond à un billet de blog, qui peut être commenté.

Le premier billet introduit la série. Le second liste les données agrégées. Le troisième présente le choix de modélisation que nous avons retenus.

SPARQL endpoint : <https://lod.abes.fr/sparql>  
Interface de recherche *full text* et de navigation : <https://lod.abes.fr/fct>

[1] <https://punktokomo.abes.fr/2016/05/16/mettre-nos-donnees-en-reseau-un-demonstrateur-1-introduction/>

## ● ● ● ARCHIVES ET WEB DE DONNÉES : CONTEXTES ET PERSPECTIVES <sup>1</sup>



La description archivistique est restée jusqu'à présent **largement marquée par des compromis entre une approche textuelle et documentaire**, induite par la prégnance de l'instrument de recherche dans les modes de signalement, et une orientation progressive vers les données. La version 3 du format EAD<sup>2</sup> ne dément pas ce constat : la révision du standard hésite en effet à sortir de la logique de l'encodage de textes, en l'absence du vocabulaire et des outils nécessaires à la mise en relation d'entités restant parfois à consolider.

« **Records in context** » (Ric) est la première réponse d'ensemble, émanant du monde des archives, aux besoins d'interopérabilité et d'ouverture qui caractérisent le web de données. En 2012, devant la nécessité de mieux articuler les quatre normes de description archivistique (consacrées aux fonds, contextes, fonctions et institutions de conservation), face aussi au manque d'une prise en compte globale de tous les types d'archives (électroniques, ou intermédiaires dans le cadre du *records management*), l'Ica<sup>3</sup> a chargé son Groupe d'experts sur la

description archivistique<sup>4</sup> de créer un modèle conceptuel global (Ric-CM) et de développer une ontologie de domaine (Ric-O) pour exprimer cette modélisation et la rendre exploitable par les machines. Les premières versions en seront présentées à l'occasion du congrès de l'Ica à Séoul<sup>5</sup>, en septembre 2016.

Ric partage des concepts avec les autres modèles du triolet LAM<sup>6</sup> (FRBR<sup>7</sup> et Cidoc CRM<sup>8</sup>), tout en introduisant des entités spécifiques (notamment trois types hiérarchisés de « documents » d'archives : *RecordSet*, *Record* et *RecordComponent*). Au-delà de la singularisation des producteurs que le schéma EAC-CPF (2010)<sup>9</sup> a permis d'établir sous la forme de véritables notices d'autorité archivistiques, des perspectives complémentaires à l'approche par provenances seront offertes, notamment à travers les fonctions et mandats des entités productrices d'archives. Les travaux de l'Egad s'inscrivent dans la continuité des grands principes archivistiques (provenance, respect des fonds) et ne visent pas à substituer Ric aux normes et standards déjà en place,

mais plutôt à offrir à ces derniers un nouveau cadre général pour faciliter leur évolution.

A l'instar des standards XML tels qu'EAD, qui évoluent en tension entre des fonctions d'échange et de catalogage, Ric pose la question de ses impacts, directs ou non, sur les processus de production de données. Il devra offrir les moyens d'une logique moins monolithique ou binaire dans la description archivistique, et constituer le socle d'un réagencement en archipel de composantes plus différenciées.

**JEAN-MARIE FEURTET**

Responsable fonctionnel de Calames, Abes  
feurtet@abes.fr



Phot. FredR (CC-BY-NC-ND 2.0)

[1] Cet article est tiré des présentations d'Anila Angjeli, Florence Clavaud et Stéphanie Roussel à l'occasion du Forum de l'Association des Archivistes Français, organisé à Troyes du 30 mars au 1<sup>er</sup> avril 2016 (site du Forum : <http://forum2016.archivistes.org/>)

[2] *Encoded Archival Description* (site officiel : <https://www.loc.gov/ead/>). Voir le Dictionnaire des balises de l'EAD3 : <http://www.loc.gov/ead/EAD3taglib/>

[3] Conseil international des archives : <http://www.ica.org/>

[4] Egad, groupe présidé par Daniel Pitti : <http://www.ica.org/fr/node/13580>

[5] <http://www.ica2016.com/french/>

[6] Acronyme regroupant les trois institutions productrices de données scientifiques et culturelles que sont les bibliothèques, les services d'archives et les musées (*Libraries, Archives, Museums*).

[7] *Functional Requirements for Bibliographic Records*

[8] *Conceptual Reference Model*, modèle propre au patrimoine culturel. Le Cidoc, Comité International pour la documentation, se consacre d'abord aux collections muséales.

[9] *Encoded Archival Context – Corporate bodies, Persons, Families* (site officiel : <http://eac.staatsbibliothek-berlin.de/>).

# Bibframe, un nouveau modèle de données pour les bibliothèques

**Initiative visant à faire évoluer les standards de description bibliographique vers un modèle de données liées, Bibframe (*Bibliographic Framework Initiative*) doit ainsi faciliter l'accès à l'information bibliographique et maximiser son utilisation par les différentes communautés d'usagers.**

L'initiative Bibframe a été lancée en 2011 par la Bibliothèque du Congrès afin de fournir « une base pour l'avenir de la future description bibliographique, pour le web mais aussi plus largement pour le monde interconnecté »<sup>1</sup>. Le projet vise tout d'abord à remplacer le format Marc 21, qui fut la pierre angulaire de l'informatisation des bibliothèques et de la gestion des métadonnées. Bien que Marc 21 soit un standard très bien structuré et maintenu – offrant un riche panel d'éléments adapté à la complexité de l'information bibliographique –, il montre toutefois aujourd'hui ses limites. Initialement format d'échange, il a été largement implémenté dans les SIGB en tant qu'interface de catalogage. Depuis, Marc est devenu véritablement la *lingua franca* des catalogueurs. Mais les SIGB n'étant malheureusement pas suffisamment adaptés aux nouveaux enjeux du catalogage, Marc est également devenu un format dépassé.

## QUAND MARC 21 RENCONTRE LE LINKED LIBRARY DATA

Pour définir le successeur du format Marc, il était nécessaire de prendre en compte les approches du web de données. Aujourd'hui, le web est la plateforme où se trouvent nos utilisateurs, c'est un outil incontournable. La construction du web sémantique est un effort entrepris par un grand nombre de communautés. Cependant, pendant longtemps, les bibliothèques n'ont pas réussi à prendre part au développement de standards qui constituent pourtant les bases de sa structure<sup>2</sup>. Les « silos de données » stockent des informations accessibles via des interfaces non ouvertes sur le web, donnant ainsi sur Internet l'image d'un trou façonné par les bibliothèques. Pourtant, la valeur ajoutée portée par la richesse de leurs métadonnées les encourage à faire partie intégrante du web afin de contribuer au « *Giant global graph* »<sup>3</sup>.

Avant Bibframe, quelques implémentations allant dans le sens du *Library Linked Data* existaient déjà. Afin de fournir un meilleur service aux « consommateurs » de données, des éléments issus de différents vocabulaires (tels Dublin Core, FOAF, BIBO, RDA, ISBD...) ont été choisis. Si cette sorte de « picorage » reste une approche valide, il semble toutefois nécessaire de fonder un modèle et un vocabulaire spécifiques qui puissent offrir à l'ensemble des éléments une structure cohérente.

## PETIT RETOUR EN ARRIÈRE

Tenant compte des modèles existants et s'appuyant sur des analyses approfondies des potentialités des technologies du web de données, la Bibliothèque du Congrès, en partenariat avec la société Zepheira<sup>4</sup>, a établi un nouveau modèle ainsi qu'un vocabulaire. Quatre classes principales ont ainsi été définies : « Œuvre », « Instance », « Autorité » et « Annotation ». Un premier groupe d'expérimentateurs a ensuite été mis en place. Le modèle et le vocabulaire ont été intensément discutés, des documents de travail et d'étapes ont été rédigés afin de mettre en lumière les cas particuliers, des outils ont été développés pour convertir les données Marc en Bibframe et créer des données Bibframe *ex nihilo*. Enfin, des tests ont été effectués afin de vérifier si Bibframe pouvait convenir autant aux données existantes qu'aux futures implémentations. Au bout d'un an, le cercle des premiers expérimentateurs a été élargi et le « Registre d'implémentation Bibframe » a été mis en place. Il s'agit d'une liste ouverte des institutions utilisant les structures Bibframe dans des situations concrètes.

## DES ACTEURS DIVERS ET VARIÉS

Différents acteurs sont très actifs dans l'environnement Bibframe. On n'évoquera ici que les plus importants.

Tout d'abord, la Bibliothèque du Congrès

a préparé un prototype en vue de tester l'efficacité du modèle lors de la création de métadonnées de documents et de langues variés. Travaillant en parallèle en Marc 21 et en Bibframe, les catalogueurs ont donné leurs avis aux experts du modèle et du vocabulaire. Les outils constituant les blocs structurants du modèle pour le prototype ont été mis à disposition. On note que le Programme pour le catalogage coopératif (PCC) a été largement impliqué dans les activités Bibframe de la Bibliothèque du Congrès. Avec l'initiative LibHub, Zepheira a choisi de rendre visibles les richesses des bibliothèques par la création d'un vocabulaire Bibframe modulable. Les données, collectées principalement à partir de catalogues de bibliothèques de lecture publique, ont été converties en Bibframe. Les relations entre entités ont ensuite été créées. Enfin, les données ont été exposées sur le web afin que les fournisseurs de moteurs de recherche puissent les utiliser. Certains résultats sont déjà visibles.

OCLC s'est concentré sur *schema.org*, vocabulaire déterminé par les quatre plus grands fournisseurs de moteurs de recherche. Quelques éléments pour les données spécifiques aux bibliothèques ont été ajoutés et une couche additionnelle a été définie. L'analyse des résultats a été publiée par les experts de la Bibliothèque du Congrès et d'OCLC dans un article intitulé : « Common Ground »<sup>5</sup>. Les modèles sont maintenant utilisables de façon interactive afin que les données puissent être « mappées » puis réutilisées.

Le projet LD4L (« *Linked Data for Libraries* ») et son successeur LD4P (« *Linked Data for Production* ») – financés par la Fondation Mellon<sup>6</sup> et pilotés par plusieurs bibliothèques de recherche – visent quant à eux à développer une nouvelle infrastructure pour la gestion des ressources et des

métadonnées. Dès le début, Bibframe a joué un rôle significatif. Il reste maintenant à voir comment ces différentes voies vont converger à l'avenir.

## L'IMPLICATION DE LA BIBLIOTHÈQUE NATIONALE ALLEMANDE (DNB)

La Bibliothèque nationale allemande a été parmi les premiers expérimentateurs de Bibframe. En 2013, un projet a été mis en œuvre pour soutenir l'initiative. En plus d'autres activités, la DNB a mis en place un service fournissant des données basées sur un simple *mapping* du format pivot Pica+ vers Bibframe : une première table de conversion a été écrite et implémentée dans le catalogue de la DNB. Concrètement, à partir de l'affichage simple de la plupart des notices, un menu propose d'obtenir une représentation Bibframe.

Le fait d'avoir entrepris ces tests sur des données réelles a permis à la DNB de partager des résultats détaillés avec les autres partenaires Bibframe. Depuis, la DNB suit l'initiative en tant qu'observateur, mais une participation plus active est prévue pour la suite du projet sur 2016-2017. À mi-parcours de son développement, Bibframe deviendra un produit solide du portefeuille de l'offre de métadonnées de la DNB.

## DE NOUVEAUX DÉVELOPPEMENTS

Très récemment, le modèle et le vocabulaire ont subi de profondes modifications. Les résultats issus des analyses du groupe pilote de la Bibliothèque du Congrès et du LD4L ont indiqué qu'un processus de révision s'imposait. En avril 2016, une nouvelle version Bibframe a été publiée. Dorénavant, Bibframe 2.0 structure l'information en trois niveaux fondamentaux d'abstraction : **l'œuvre**, qui reflète l'aspect conceptuel de la ressource cataloguée ; **l'instance**, une représentation « individuelle » de l'œuvre, et **l'item**, nouveau concept pour désigner la copie physique ou numérique d'une instance. Autres concepts clés à partir desquels des relations ont été bâties entre les trois classes fondamentales : les **agents**, i.e. les personnes physiques, les organisations, etc., associés à une œuvre ou une instance à travers des rôles spécifiques, les **sujets** qui décrivent le contenu de l'œuvre, et les **événements**. Les concepts « Autorité » et « Annotation » considérés comme non pertinents ont été abandonnés. Ce nouveau

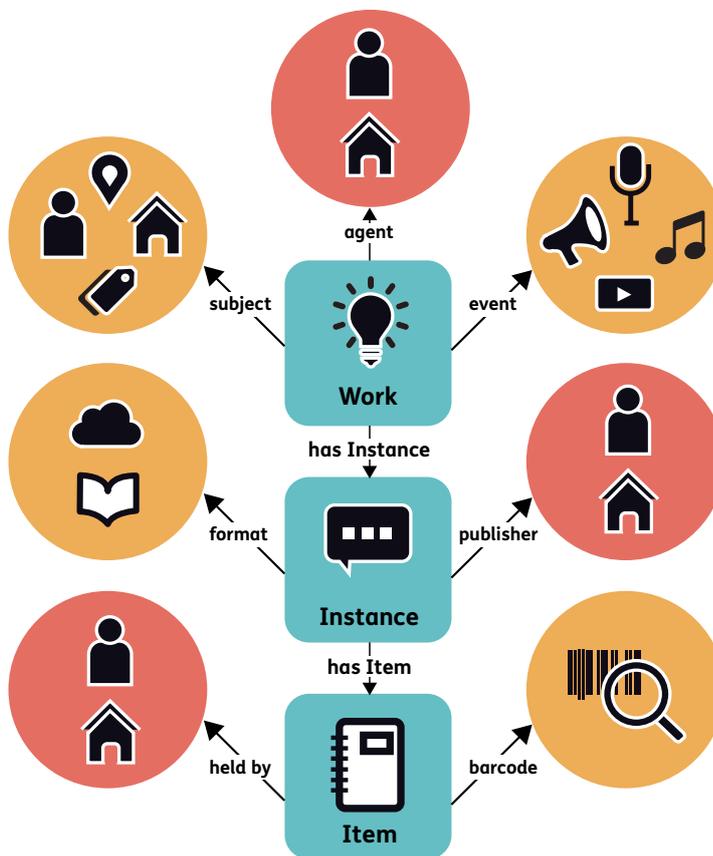


Schéma du modèle et vocabulaire Bibframe  
Source : <http://www.loc.gov/bibframe/docs/bibframe2-model.html>

vocabulaire est publié en RDF et disponible sous une forme qui permet la navigation hypertextuelle et la recherche par différents critères.

## QUESTIONS OUVERTES

Comme tout nouveau produit, Bibframe a parcouru le « *Hype cycle* » : il a débuté par la mise en place d'une nouvelle technologie, avant de suivre l'ascension du « Pic des espérances exagérées », puis la descente vers le « Gouffre des désillusions » et à nouveau l'ascension de la « Pente de l'illumination » pour arriver sur le « Plateau de productivité »<sup>7</sup>.

Quelques interrogations subsistent : Bibframe est-il encore trop fondé sur Marc 21 ? Respecte-t-il les bonnes pratiques du web de données ? Les différentes approches finiront-elles par diviser les communautés ? Comment cette initiative peut-elle devenir vraiment internationale ? Comment intégrer plus largement les secteurs culturels et patrimoniaux ? Combien de temps faudra-t-il pour atteindre un bon niveau de production de métadonnées ? Y aura-t-il un retour sur investissement ? Et dans quelle mesure la mise en œuvre de Bibframe concrétisera-t-elle la nécessité de servir les intérêts des utilisateurs finaux ?

## CONCLUSION

Il est toujours fascinant de voir un nouveau standard in *statu nascendi*.

Bibframe est évolutif et n'en est qu'au début de son développement. Au fil de la transition de Marc vers Bibframe, la communauté des métadonnées se transforme également. Le succès du projet repose sur un équilibre entre les différents objectifs, rassemblant tous les acteurs dans un modèle de maintenance et de gouvernance commune. Les institutions à l'échelle mondiale, y compris les éditeurs de logiciels et de ressources, sont invitées à participer à ce projet en proposant et en « connectant » à Bibframe leurs métadonnées. Intégrer les standards existants tels que Mods, Onix et Unimarc à Bibframe est aussi une voie à explorer. Enfin, toute contribution est bien sûr la bienvenue !

REINHOLD HEUVELMANN

Bibliothèque nationale allemande  
R.Heuvelmann@dnb.de  
Traduction : P.Pons (Abes)

- [1] <https://www.loc.gov/bibframe/>
- [2] « Building blocks » dans le texte original
- [3] « Graphe global géant », expression employée pour la première fois par Tim Berners-Lee
- [4] <https://zepheira.com/>
- [5] <http://www.oclc.org/research/publications/2015/oclcresearch-loc-linked-data-2015.html>
- [6] <https://mellon.org/>
- [7] [https://fr.wikipedia.org/wiki/Cycle\\_du\\_hype](https://fr.wikipedia.org/wiki/Cycle_du_hype)

PRESSoo est une ontologie adaptée à la description des ressources continues, notamment des publications périodiques. Ce modèle, extension de FRBRoo, a été développé par le Centre international de l'ISSN et la Bibliothèque nationale de France. Il est désormais en cours de normalisation à l'Ifla et doit permettre de faciliter la diffusion en *linked data* d'informations sur les ressources continues.

# PRESSoo : décrire les ressources continues dans le web de données

## D'OÙ VIENT PRESSoo ?

Depuis 1997, le monde des bibliothèques dispose d'un modèle conceptuel pour l'ensemble des documents constituant les collections qu'il a pour mission de gérer. Ce modèle, appelé FRBR<sup>1</sup>, a été développé par l'Ifla, et a donc une valeur internationale. Le nouveau code de catalogage RDA<sup>2</sup>, qui s'est substitué en 2010 aux anciennes règles de catalogage anglo-américaines AACR, a une structure entièrement calquée sur le modèle FRBR. RDA n'émane pas de l'Ifla, mais c'est néanmoins un code supranational, dont la sphère d'influence s'étend bien au-delà des seuls pays anglo-saxons. Depuis 2013, c'est le cadre de référence officiel de tout le catalogage produit par la Bibliothèque du Congrès et la British Library.

Le Centre international de l'ISSN, dont la fonction est de centraliser l'ensemble des notices de ressources continues produites par les Centres ISSN nationaux, reçoit donc régulièrement des notices conformes au code RDA pour les publications en série.

Or, le modèle FRBR n'entre pas dans tous les détails de la notion de « sérialité », comme il le reconnaît lui-même : « Certains aspects du modèle mériteraient un examen plus approfondi. [...] L'analyse de la notion de « sérialité » [...] demande notamment à être approfondie<sup>3</sup> ». C'est pourquoi le Centre international de l'ISSN s'est mis en quête, dès 2012, d'un modèle conceptuel qui lui permettrait d'interpréter les notices RDA qui lui seraient fournies et de les faire dialoguer dans une même base de données avec des notices n'ayant pas la même structure. Il se trouve qu'à côté du modèle FRBR tel qu'il fut approuvé en 1997, il existe aussi le modèle FRBRoo, qui constitue une synthèse entre le modèle conceptuel des bibliothèques et celui des musées, le Cidoc CRM<sup>4</sup>. L'élaboration de ce modèle a été l'occasion d'entrer un peu plus dans les détails des publications en série, sans toutefois encore aller au fond des choses. Cette amorce de modélisation des publications en série a paru intéressante au

Centre international de l'ISSN, qui s'est tourné vers la Bibliothèque nationale de France, en raison de sa participation aux travaux autour du modèle FRBRoo, pour l'aider à construire un modèle susceptible d'être accepté au niveau international<sup>5</sup>.

Les travaux ont eu lieu tout au long de l'année 2013, et ont débouché sur la diffusion d'un nouveau modèle, intitulé PRESSoo, en 2014. Ce modèle, dont le nom n'est pas un acronyme mais cherche simplement à évoquer le monde de la presse, plus familier aux non-spécialistes que les expressions « publications en série » ou « ressources continues », s'appuie sur les modèles FRBRoo et Cidoc CRM,

## Le Centre international de l'ISSN s'est tourné vers la BNF pour l'aider à construire un modèle susceptible d'être accepté au niveau international.

dont il est indissociable. Approuvé par le Groupe d'intérêt spécial pour le Cidoc CRM et le Groupe de révision des FRBR en 2014, il fut soumis à enquête internationale en 2015. La version 1.2 du modèle a été publiée en janvier 2016 et est en cours d'adoption officielle par l'Ifla<sup>6</sup>.

## À QUOI RESSEMBLE PRESSoo ?

PRESSoo reprend à son compte la classe consacrée par le modèle FRBRoo aux ressources continues, F18 *Serial Work*. Les ressources continues se plient difficilement aux distinctions fondamentales dans le modèle FRBR entre « Œuvre », « Expression », « Manifestation » et « Item » : quand on essaie de leur appliquer mécaniquement ce schéma, on se heurte très vite à des situations inextricables. Aussi a-t-il été décidé, tant dans le modèle FRBRoo que dans le modèle PRESSoo, de considérer les ressources continues uniquement comme des « œuvres », sans chercher à en détailler ni les « expressions » ni les « manifestations ». Il en résulte que les relations entre les différentes versions linguistiques ou les

[1] Pour : *Functional Requirements for Bibliographic Records*. Voir : Groupe de travail Ifla sur les Fonctionnalités requises des notices bibliographiques, *Fonctionnalités requises des notices bibliographiques : rapport final*, 2<sup>e</sup> édition française, Paris : Bibliothèque nationale de France, 2012. [http://www.bnf.fr/documents/frbr\\_rapport\\_final.pdf](http://www.bnf.fr/documents/frbr_rapport_final.pdf).

[2] Pour : *Resource Description and Access*. Voir : *RDA Toolkit*, <http://www.rdatoolkit.org/>.

[3] Groupe de travail Ifla sur les Fonctionnalités requises des notices bibliographiques, *Fonctionnalités requises...*, p. 10. Plusieurs commentateurs ont souligné les défaillances du modèle FRBR dès lors qu'il est question d'analyser périodiques et collections. Voir par exemple Ed Jones, « The FRBR model as applied to continuing resources », *Library Resources & Technical Services*, 2005, 49(4), 227-242 ; Steve Shadle, « FRBR and serials », *The Serials Librarian*, 2006, 50(1-2), 83-103 ; Adolfo R. Tarango, « FRBR for serials : rounding the square to fit the peg » : présentation donnée en 2008 lors de la conférence annuelle de l'American Library Association (ALA).

[4] Voir le site <http://www.cidoc-crm.org/>, où les deux modèles Cidoc CRM et FRBRoo sont disponibles.

différents formats de publication d'un périodique sont analysées comme des relations entre des œuvres, et non des relations entre des expressions ou des manifestations, comme c'est le cas pour les monographies.

Le modèle PRESSoo introduit en revanche la notion de « règle éditoriale » (Z12 *Issuing Rule*), qui peut être n'importe lequel des éléments de la politique éditoriale globale définissant la « sérialité » spécifique d'une ressource continue : sa fréquence prévue, le formatage prévu de sa numérotation, les dimensions prévues et le titre prévu pour chacun de ses fascicules, etc. L'adjectif « prévu » n'est pas répété ici par hasard : il constitue un aspect fondamental de la description bibliographique des ressources continues. Une notice de périodique décrit à la fois ce que l'on a constaté sur les fascicules publiés de ce périodique et le comportement éditorial que l'on s'attend à constater dans l'avenir pour les fascicules à paraître. Tous les éléments de description présents dans le Manuel de l'ISSN<sup>7</sup> ont été transcrits dans PRESSoo et des classes et des propriétés spécifiques ont été définies en fonction des besoins. Ce modèle permet ainsi de représenter l'ensemble de la richesse sémantique propre à la description des ressources continues.

## UTILISER ET IMPLÉMENTER PRESSoo

Quelles sont les applications du modèle PRESSoo? Notons d'abord que comme FRBR ou FRBRoo, ce n'est pas une norme de catalogage et il n'a pas nécessairement vocation à être implémenté tel quel. Comme modèle conceptuel, il est ainsi étudié par le groupe de travail sur les Agrégats de RDA<sup>8</sup>, qui étudie le catalogage des anthologies, des œuvres complètes ou encore des ressources continues. Cependant, comme modèle orienté objet, il est également tout à fait adapté à la publication de données dans le *Linked Open Data*. Il a fait ainsi l'objet d'une implémentation partielle dans le cadre d'un projet mené avec le soutien de l'Unesco : le projet Road, le Répertoire des ressources scientifiques en accès ouvert<sup>9</sup>. Road donne accès librement aux notices de l'ensemble des publications en *open access* décrites par le Registre ISSN et offre une possibilité de téléchargement en *dump*. C'est le modèle PRESSoo – accompagné d'une ontologie *ad hoc*, l'ontologie Road – qui est utilisé dans ce cadre, car lui seul sait transcrire toute la complexité de la description des ressources continues et des relations entre ces ressources.



Phot. Rodney Campbell, The Long and Windy Road / CC-BY NC ND 2.0 (Flickr)

➔ Road : une route de plus vers l'*open access* !

Outre un objectif de promotion de l'*open access*, Road devait permettre au Centre international de l'ISSN d'acquérir une expérience pratique en matière de *linked data*. Fort de cette expérience probante, le Centre international ISSN a conçu et validé sa future stratégie d'enrichissement et de diffusion des données qui s'articulera autour d'une mise à disposition de l'ensemble des notices du Registre ISSN dans des formats et des technologies du web sémantique selon des modalités économiques qui restent à préciser. Plusieurs modèles, s'appuyant notamment sur l'ontologie PRESSoo, seront donc élaborés pour répondre aux besoins des divers utilisateurs de l'ISSN.

**PATRICK LE BŒUF**

Bibliothèque nationale de France  
patrick.le-boeuf@bnf.fr

**CLÉMENT OURY**

Centre international ISSN  
clement.oury@issn.org

[5] Pour une présentation détaillée de la genèse et des caractéristiques de PRESSoo, voir Patrick Le Bœuf et François-Xavier Pelegrin, « FRBR et publications en série : le modèle PRESSoo », *Actes de la 80<sup>e</sup> conférence de l'Ifla*, Lyon, août 2014, <http://library.ifla.org/838/7/086-leboeuf-fr.pdf>.

[6] Le texte est disponible à l'adresse : <http://www.ifla.org/node/10410>.

[7] Le Manuel de l'ISSN, règle de catalogage des ressources continues appliquée par les centres ISSN nationaux, est harmonisé avec l'ISBD, et tient compte des évolutions apportées par RDA. Disponible à l'adresse <http://www.issn.org/fr/comprendre-issn/regles-d-attribution/manuel-issn-en-ligne/>.

[8] Voir la présentation de l'ensemble des groupes de travail de RDA à l'adresse <http://www.rda-rsc.org/workinggroups>.

[9] Voir <http://road.issn.org>.

# Le Web de données à la BNF

Dans quel sens faut-il le dire ? Est-ce la BNF qui s'intègre dans le web de données, ou le web de données qui gagne la BNF ? Il est certain en tout cas que le web de données fait penser et agir différemment, et que pour l'information bibliographique en général, il apparaît désormais comme un horizon naturel.

À la Bibliothèque nationale de France (BNF), l'intention est nette et affichée. Dans son *Contrat de performance pour les années 2014 à 2016*<sup>[1]</sup>, elle inscrit clairement son action bibliographique « dans le mouvement d'ouverture des données publiques (*open data*) comme dans le développement du web sémantique » (p.8), ainsi que dans « la politique d'ouverture des données publiques culturelles poursuivie par le ministère de la Culture et de la Communication. » (p.10). Une priorité est énoncée en appui de cette direction : le développement de [data.bnf.fr](http://data.bnf.fr)<sup>[2]</sup>.

Le *Schéma numérique* publié en mars 2016<sup>[3]</sup> cite lui aussi le web de données, conjointement au mouvement d'ouverture des données publiques, comme « une réelle opportunité technologique pour le développement des missions de service public qui incombent à la BNF » : l'entrée dans le web de données est à la fois l'occasion et le fruit d'un décloisonnement pour les bibliothèques ; elle rend nécessaire le dialogue avec d'autres communautés

professionnelles et ouvre la voie à de nouvelles coopérations, avec des opérateurs culturels comme des producteurs de données de tout type : statistiques, géographiques, scientifiques, etc. Au-delà de l'expression des données en RDF et de l'ouverture juridique, le *Schéma numérique* insiste particulièrement sur deux axes, dont l'association est présentée comme le socle de ce qui pourrait être un nouveau contrôle bibliographique universel : « la normalisation et la dissémination d'identifiants de confiance », et la mise à disposition des données d'autorité sous la forme de vocabulaires ou de référentiels.

Data.bnf.fr et l'exposition en RDF, les identifiants, les référentiels d'autorité : on retrouve là les trois modalités principales d'investissement de la BNF dans le web de données.

## DATA.BNF.FR

À son lancement en 2011, [data.bnf.fr](http://data.bnf.fr) était un projet expérimental, l'occasion pour la BNF de se frotter au web de données. Il était fondé sur l'intuition qu'avec les données d'autorité, les catalogues de bibliothèques disposent de référentiels pivots cor-

respondant parfaitement à la logique et à la structure du web de données. Ses objectifs étaient multiples : améliorer la visibilité des ressources de la BNF depuis les moteurs de recherche du web, favoriser la réutilisation des données et décloisonner les catalogues de la BNF, en les articulant entre eux et en liant leurs données avec des données externes. [Data.bnf.fr](http://data.bnf.fr) se présentait sous une double forme : des pages web visibles par les internautes, et des jeux de données en RDF destinés aux machines et à la réutilisation. Les premiers résultats se sont avérés encourageants, en termes d'audience, d'appropriation par les équipes et d'image institutionnelle. La réutilisation a quant à elle avancé notablement lorsque l'intégralité du répertoire d'autorité matière Rameau a été exposée et mise à disposition en Skos<sup>[4]</sup> : la complétude de ce jeu de données et les perspectives d'utilisation

**Data.bnf.fr et l'exposition en RDF, les identifiants, les référentiels d'autorité : on a là les trois modalités principales d'investissement de la BNF dans le web de données.**

multilingue qu'offrait son alignement avec des référentiels anglo-saxon (LCSH<sup>[5]</sup>) et allemand (SWD<sup>[6]</sup>) ont sans nul doute joué en sa faveur.

Fort de ces bilans convaincants, [data.bnf.fr](http://data.bnf.fr) a pu se développer jusqu'à exposer la totalité des données d'autorité validées de la BNF, et l'ensemble des données bibliographiques qui leur sont liées (soit 70 % du Catalogue général). Aujourd'hui, les 2,8 millions de pages de [data.bnf.fr](http://data.bnf.fr) attirent 12 à 14 000 visiteurs uniques par jour, dont les deux tiers se redirigent vers les autres applications de la BNF. Désormais, après une phase marquée par un objectif principal de croissance, l'enjeu pour le site est de faire jouer à plein les ressorts du web de données, en développant les liens et les alignements de données par une politique active de partenariats. Il reste également à inventer autour des jeux de données un véritable service aux utilisateurs, au-delà de la seule mise à disposition d'une infrastructure de récupération et d'interrogation : comme toute autre collection proposée par la bibliothèque, ces jeux de données méritent une médiation. Cette médiation passera d'ailleurs par l'évolution de l'interface de

[1] [http://www.bnf.fr/documents/contrat\\_performances\\_2014.pdf](http://www.bnf.fr/documents/contrat_performances_2014.pdf)

[2] <http://data.bnf.fr>

[3] [http://www.bnf.fr/documents/bnf\\_schema\\_numerique.pdf](http://www.bnf.fr/documents/bnf_schema_numerique.pdf)

[4] *Simple Knowledge Organization System* : recommandation du W3C pour représenter du vocabulaire contrôlé

[5] *Library of Congress Subject Headings, autorités de la Bibliothèque du Congrès*

[6] *Schlagwortnormdatei*, thésaurus de la Bibliothèque nationale allemande (DNB)

recherche et de visualisation, car on ne consulte pas des jeux de données comme on consulte les documents d'une bibliothèque numérique ou même les notices d'un catalogue ; les logiques d'accès et de navigation sont spécifiques, et il reviendra à [data.bnf.fr](http://data.bnf.fr) d'expérimenter des techniques de *data visualisation*.

## RELIURES, BP16 ET DOREMUS

En complément de [data.bnf.fr](http://data.bnf.fr), la BNF a expérimenté la structuration en RDF et la publication dans le web de données pour deux sites spécialisés, *Reliures* et *BP16*. Le premier ([reliures.bnf.fr](http://reliures.bnf.fr)<sup>[7]</sup>) propose une sélection de reliures numérisées accompagnées de descriptions détaillées, et le second ([bp16.bnf.fr](http://bp16.bnf.fr)<sup>[8]</sup>) est une édition sous une nouvelle forme de la *Bibliographie des éditions parisiennes du 16<sup>e</sup> siècle*. Ici encore, les résultats sont concluants, car le RDF offre des facilités de navigation interne et permet que ces sites, dont la spécialisation aurait pu être une cause d'isolement, soient intégrés à un écosystème dans lequel ils sont reliés à d'autres ressources. Cependant, cette expérience a aussi montré qu'elle serait dans l'immédiat difficile à rééditer, en raison du coût actuel de la maintenance et de la mise à jour de telles applications.

Enfin, la BNF s'est engagée dans le projet Doremus, aux côtés de partenaires culturels et scientifiques, afin d'explorer des modèles de données plus fins que ceux qui ont été utilisés pour [data.bnf.fr](http://data.bnf.fr). De fait les modèles choisis pour [data.bnf.fr](http://data.bnf.fr) ont été volontairement simples, de façon à favoriser la réutilisation par des communautés du web. Le projet Doremus a des besoins de plus grande précision : il a pour objectif de doter les institutions culturelles musicales, les éditeurs, les distributeurs et les communautés de passionnés de modèles et de référentiels communs pour publier dans le web de données, partager et enrichir des catalogues d'œuvres et d'événements musicaux. C'est donc une extension du modèle FRBRoo (*Functional Requirements for Bibliographic Records – object oriented*) qui a été retenue ; sa souplesse permettra de gérer différents niveaux de granularité, jusqu'à restituer dans toute leur richesse les données de catalogues spécialisés.

## ISNI

Du côté des identifiants, on peut indiquer l'investissement de la BNF sur l'ISNI, qui présente de nombreux atouts. Identifiant attaché aux identités publiques de créateurs, personnes ou collectivités, sans limitation de discipline ou de domaine, et dont la qualité est garantie par l'assignation au sein d'une base centralisée, selon des règles d'attribution

contrôlées : autant d'aspects qui font de l'ISNI un puissant levier d'interopérabilité pour le web de données. La BNF est donc devenue agence ISNI, de façon à pouvoir attribuer des ISNI conformément à ses besoins, et travaille pour le moment à des scénarios d'échanges vertueux de données avec le monde de l'édition dans le cadre du dépôt légal.



Phot. R2hox / CC-BY-SA 2.0 (Flickr)

 **Data path**  
« Exposition de données »

## WEB DE DONNÉES ET FICHER NATIONAL D'AUTORITÉS

Sur le front des autorités enfin, comme l'Abes l'a annoncé lors de ses journées 2016, le temps est venu d'explorer l'opportunité et la faisabilité de constituer un fichier national. Le principe en est acquis, les tutelles en conviennent ; il reste cependant beaucoup à préciser : le périmètre d'un tel fichier, sa gouvernance, son modèle, ses formats, son architecture...

La route est longue, mais il est clair qu'elle mérite d'être empruntée : si les stratégies des deux agences bibliographiques passent par le web de données, leurs référentiels seront d'autant plus des référentiels qu'ils auront fusionné. Le web de données aura alors procuré, en plus de l'opportunité de développer les missions de signalement, les conditions d'une substantielle économie de production.

**JÉRÔME VILLEMINOZ**

Directeur adjoint du Département  
des métadonnées à la BNF  
[jerome.villeminoz@bnf.fr](mailto:jerome.villeminoz@bnf.fr)

[7] <http://reliures.bnf.fr>

[8] <http://bp16.bnf.fr/>

De nouveaux modèles de plateformes reposant sur les technologies du web sémantique et l'automatisation des traitements ont fait leur apparition, à l'instar de la plateforme de publication science.ai qui propose des contenus scientifiques aux données riches et ré-exploitable. Décryptage par Robin Berjon, expert en technologies du web et membre actif des groupes de standardisation au sein du W3C.

# Le Web, renaissance de la publication de recherche

**La publication de recherche est dans un sombre état.**

Si le volume d'articles publiés ne cesse de croître, les modalités de leur production semblent incapables de dépasser leurs archaïsmes. Ce volume pose lui-même problème, particulièrement dans les domaines les plus prolixes comme la recherche biomédicale, car la pauvreté des métadonnées et de l'indexation des articles publiés rend difficile voire impossible l'exploitation efficace du contenu existant. Il apparaît d'ailleurs qu'une écrasante majorité des citations pointe vers des publications récentes, suggérant qu'une part non négligeable du savoir produit est en tout état de cause oublié, et ce rapidement<sup>1</sup>. Les formats employés pour la mise en ligne de ces publications font typiquement montre d'une qualité de production déplorable. Quand le PDF est employé, il ne fonctionne au mieux que très péniblement sur des terminaux mobiles et présente de sérieuses barrières d'accès pour certaines personnes en situation de handicap ; si c'est du HTML il est de piètre facture, dénué de sémantique, et d'ergonomie inexistante.

S'ajoute à ces éléments la question du coût : les frais de publication augmentent de façon constante, au point que certains laboratoires limitent les publications de leurs chercheurs, et les tarifs de consultation ont suffi à rendre « Sci-Hub », fournisseur de plus de 50 millions d'articles piratés, mondialement célèbre.

Ce constat n'est pas nouveau, ces problèmes sont connus de longue date. Pourtant, la situation semble n'avoir fait qu'empirer et plusieurs tentatives d'amélioration se sont soldées par un échec. Qu'est-ce qui permet de penser qu'un changement peut désormais survenir ?

## DES MÉTADONNÉES VALORISÉES PAR LES TECHNOLOGIES DU WEB

Tout d'abord, la technologie est enfin mûre. L'avènement de HTML5 a rendu la plateforme Web suffisamment solide pour une publication de qualité, et la réussite du projet schema.org<sup>2</sup>, un système de métadonnées soutenu par les principaux moteurs de recherche du Web, a banalisé l'usage de l'enrichissement sémantique à plus de 10 millions de sites.

À titre d'exemple, le site du *New York Times* a récemment commencé à utiliser ces métadonnées pour ses recettes de cuisine. Celles-ci sont interprétées correctement par les moteurs de recherche généralistes, permettant ainsi aux utilisateurs de trouver bien plus facilement le gâteau au chocolat qui leur sied. Résultat : leur trafic a rapidement augmenté de plus de 50 %<sup>3</sup>.

Par ailleurs, le projet Sci-Hub et ses articles illégalement gratuits, dont il est établi que le monde entier l'utilise<sup>4</sup>, met les éditeurs sous pression. En effet, outre la mise à disposition gratuite du contenu il

[1] Della Briotta Parolo, Pietro, Raj Kumar Pan, Rumi Ghosh, Bernardo A. Huberman, Kimmo Kaski, and Santo Fortunato. 2015. "Attention decay in science." <http://arxiv.org/abs/1503.01881>

[2] <http://schema.org/>

[3] New York Times. 2014. "The Full New York Times Innovation Report." *New York Times*. <https://www.scribd.com/doc/224608514/The-Full-New-York-Times-Innovation-Report>

[4] Bohannon, John. 2016. "Who's downloading pirated papers? Everyone." *Science*. 28 April. <http://www.sciencemag.org/news/2016/04/whos-downloading-pirated-papers-everyone>



Page d'accueil de la plateforme Pirate Sci-Hub

apparaît qu'un nombre croissant de chercheurs ont recouru à Sci-Hub même quand ils ont légalement accès aux articles, tout simplement parce qu'ils trouvent l'interface plus ergonomique que celle fournie par les éditeurs.

Sci-Hub n'est pas la solution, cependant. Outre son illégalité, c'est pour l'essentiel une source de documents sous format PDF, pauvres en sémantique et largement inaccessibles. Mais c'est un clairon qui réveille les acteurs de l'édition de recherche.

Le moment est donc propice au changement, et c'est précisément l'ensemble de ces problèmes que la plateforme de publication science.ai<sup>5</sup> s'applique à résoudre.

## SCIENCE.AI : UNE CHAÎNE DE PUBLICATION OPTIMISÉE

L'approche choisie est résolument centrée sur les technologies du Web moderne et sur l'automatisation du traitement. En termes de coûts, celle-ci permet de tirer les frais de publication tellement bas que le débat sur le libre accès devient caduc : il ne subsiste tout simplement plus de raison financière de faire payer la consultation.

Ceci se fait sans compromission au niveau de la qualité de production, le Web moderne permettant précisément de bien meilleures publications et utilisations que précédemment. Dès l'obtention du

trouvent que le Web ne présente pas le côté pratique du PDF, simple à télécharger. C'est sans compter sur les nouvelles fonctionnalités hors-ligne du Web et sur la convergence entre le Web et les formats de livres électroniques. Il est aujourd'hui possible d'utiliser du contenu Web de façon locale et déconnectée.

Finalement, comment obtient-on des métadonnées pertinentes ? Les auteurs sont en effet souvent peu coopératifs lorsqu'il s'agit de remplir de longs formulaires décrivant leurs travaux... Ce problème est résolu par le truchement de deux méthodes. D'une part, des techniques issues de l'intelligence artificielle fournissent là où c'est possible une classification du contenu dont la précision est comparable à celle obtenue par des intervenants humains. D'autre part, l'interface de mise en ligne du manuscrit met l'enrichissement naturel de ce document par les métadonnées dans un contexte compréhensible, ce qui encourage subrepticement les auteurs à fournir plus d'informations qu'ils ne le feraient face à un formulaire rebutant.

Si ce système doit bien entendu faire ses preuves, un partenariat commercial avec John Wiley & Sons, un des plus grands éditeurs de recherche, est déjà en cours. Afin d'en juger par soi-même, chacun peut s'inscrire à un système de démonstration, limité mais fonctionnel, sur le site <https://science.ai/>. Cette plateforme est conçue pour fonctionner à plusieurs

## Une renaissance de la publication de recherche, par l'entremise des technologies ouvertes qui sont aujourd'hui disponibles, est à portée de main.

manuscrit, le contenu est converti en un ensemble utilisant à la fois le format de données sémantiques RDF et le format de texte structuré HTML. Ceci permet de publier un contenu non seulement riche en métadonnées (lesquelles sont utilisables par le projet « science.ai » lui-même comme par toute personne tierce car celles-ci sont imbriquées directement dans la page) mais aussi pleinement accessible indépendamment des terminaux ou des handicaps, le tout sur la base de standards ouverts. Bien entendu, cette transition connaît ses obstacles. Premièrement, les manuscrits fournis par les auteurs restent dans des formats dénués de sémantique et péniblement exploitables comme Word ou LaTeX. À cet effet nous avons établi un guide d'utilisation de Word (outil de plus de 85 % des manuscrits) qui, sans installation d'outil ou de feuille de style spécifique, permet d'extraire toutes les informations essentielles d'un document. C'est le "DOCX Standard Scientific Style" (ou DS3), disponible librement en ligne à <http://docx.science.ai/>.

Un autre obstacle est que beaucoup de chercheurs

échelles. À une extrémité du spectre un éditeur de grande taille peut s'en servir pour gérer de nombreuses publications, mais l'outil reste tout à fait accessible à une petite société savante, voire à un auteur seul, qui désire mettre en place une

revue. Un de ses objectifs est d'ailleurs d'offrir une interface suffisamment simple et accessible pour libérer l'innovation en matière de ligne éditoriale, de types de contenus, ou d'organisation de comité de lecture. C'est une plateforme permettant au plus grand nombre d'accéder potentiellement à la publication de recherche.

Une renaissance de la publication de recherche, par l'entremise des technologies ouvertes qui sont aujourd'hui disponibles, est à portée de main. Car si le bon sens est la chose du monde la mieux partagée, il s'agirait que la connaissance le soit aussi. Des articles apportant plus de précisions techniques sur certains aspects sont disponibles sur <https://research.science.ai/> (en anglais), et bien entendu nous attendons vos retours sur notre système lorsque vous l'aurez essayé!

**ROBIN BERJON**  
Directeur technique,  
science.ai – New York, USA  
robin@berjon.com

[5] <https://science.ai/>

# TriPhase : co-construction d'une ressource termino-ontologique<sup>1</sup>

La construction de *TriPhase* ou «Terminologie pour la Recherche d'Information du Département Phase», a fait l'objet d'un projet initié par le chef du Département scientifique Physiologie Animale et Systèmes d'Élevage (Phase) de l'Inra (Institut National de la Recherche Agronomique) en 2013. Retour sur les différentes étapes de ce projet collaboratif inédit.

[1] Une ressource termino-ontologique est une ontologie légère qui vise à répondre à un objectif précis et est orientée par un point de vue. Elle se base sur trois sources de connaissances : l'expertise humaine, les ressources existantes, les documents d'un corpus – Voir le poster de Tissaoui, A. « Typologie de changements et leurs effets sur l'évolution de ressources termino-ontologiques » présenté au 20e Journées francophones d'ingénierie des connaissances (IC 2009), Hammamet (TN). [http://ic2009.inria.fr/docs/posters/Tissaoui\\_Poster\\_IC2009.pdf](http://ic2009.inria.fr/docs/posters/Tissaoui_Poster_IC2009.pdf)

[2] Golik W., Dameron O., Bugeon J., Fatet A., Hue I., Hurtaud C., Reichstadt M., Salaün M.-C., Vernet J., Joret L., Papazian F., Nédellec C. et Le Bail P.-Y., « ATOL: the multi-species livestock trait ontology » in *Proceedings of The 6th Metadata and Semantics Research Conference (MTSR 2012)*, pages 289-300. Springer Verlag Communications in Computer and Information Science Serie. Cadiz, Espagne, 28 au 30 novembre 2012. DOI: 10.1007/978-3-642-35233-1\_28

[3] Nédellec C., Golik W., Aubin S., Bossy R., « Building Large Lexicalized Ontologies from Text: a Use Case in Indexing Biotechnology Patents », *International Conference on Knowledge Engineering and Knowledge Management (EKAW 2010)*, Volume 6317 of the series *Lecture Notes in Computer Science* pp 514-523, Springer Verlag, Lisbon, Portugal, 11<sup>th</sup> 15<sup>th</sup> octobre, 2010.

[4] [www.atol-ontology.com/index.php/fr/](http://www.atol-ontology.com/index.php/fr/)

[5] <http://prodinra.inra.fr/?locale=fr>

Le domaine de recherche du Département Phase est par nature transversal en termes d'objets étudiés, de niveau d'organisation (de la cellule au système d'élevage en passant par l'organisme animal) et de disciplines mobilisées (biologie, physiologie, endocrinologie, neurosciences, comportement...). En 2013, le chef de Département, Benoît Malpaux, pose le problème du classement des publications des chercheurs par thème et sous-thème et l'évolution de ce classement au cours du temps, en l'absence de thésaurus structuré. Il s'agit d'un besoin récurrent, à des fins d'analyse stratégique, d'évaluation ou de pilotage de la recherche. L'indexation manuelle exigerait une quantité de travail humain très importante compte tenu du volume et de la récurrence du besoin. La solution retenue est alors celle d'une indexation sémantique automatique en texte plein des références des publications, titres et résumés. Pour cela les ingrédients nécessaires sont des outils logiciels d'indexation sémantiques et une ressource terminologie-ontologique structurée, laquelle représente avec précision la structuration thématique du Département. Une indexation structurée hiérarchique permet de visualiser l'évolution et la distribution des thèmes au cours du temps à différents niveaux de détail. La ressource étant intégralement à construire dans un délai court pour respecter l'échéance de l'évaluation du Département, il a été décidé d'utiliser une technologie qui avait fait ses preuves dans un projet précédent d'indexation de journal par une ontologie<sup>2</sup> : extraction automatique des termes candidats des références et organisation de ceux-ci en thésaurus à l'aide du logiciel collaboratif, TyDI (Terminology Design Interface)<sup>3</sup>. Les compétences diverses et complémentaires de l'équipe projet sont à la mesure de l'ambition. Il s'agit de compétences en indexation de documents et en construction de thésaurus apportées par le réseau des six documentalistes du Département, des compétences en ingénierie des connaissances et en développement informatique de système d'information, d'indexation sémantique et d'outils d'aide à la conception d'ontologie apportés par cinq personnes de l'équipe de recherche Bibliome de l'Unité Inra MaLAGE (Mathématiques et Informatique Appliquées du Génome à l'Environnement).

## CRÉATION D'UNE ARBORESCENCE DE THÉMATIQUES

La première étape du projet a démarré par la construction manuelle du premier niveau de l'arborescence représentant l'ensemble des thématiques de recherche du Département Phase, à partir du document d'orientation à 5 ans. Cette structure a été importée dans l'outil TyDI, ainsi que les ontologies pertinentes pour le domaine, développées par les scientifiques du Département Phase sur les caractères et les phénotypes des animaux d'élevage dans leur environnement : Atol<sup>4</sup> (*Animal Trait Ontology for livestock*) et Eol (*Environnement Ontology for Livestock*).

## ENRICHISSEMENT VIA LES PUBLICATIONS DES CHERCHEURS

La deuxième étape a consisté à enrichir la ressource termino-ontologique *TriPhase* par le vocabulaire utilisé dans les publications des chercheurs du Département. Ces publications ont été exportées depuis ProDInra, la base de données institutionnelle de l'Inra<sup>5</sup>. ProDInra archive les publications des chercheurs de l'Inra et le Département Phase y intègre toute sa production. Le vocabulaire associé aux publications et chargé dans TyDI provient de trois origines : les mots-clés des auteurs des publications, les mots-clés choisis par les documentalistes pour l'indexation dans ProDInra et les termes, mots simples et mots composés, extraits automatiquement par l'outil BioYaTeA des titres et résumés (Golik et al., 2013)<sup>6</sup>.

Les documentalistes ont structuré les classes sémantiques de *TriPhase*, sachant qu'une classe sémantique est définie par un concept avec un label (le terme favori) et les termes qui lui sont associés, (synonymes, variations typographiques et traductions). Chaque concept est relié à ses parents (concepts plus généraux) et éventuellement à ses enfants (concepts plus spécifiques). Lors de cette étape, des ressources terminologiques extérieures pertinentes pour le domaine comme le MeSH<sup>7</sup>, Agrovoc<sup>8</sup> et Cab Thesaurus<sup>9</sup> ont été consultées régulièrement par les documentalistes afin de structurer, justifier ou arbitrer les choix de peuplement.

## ÉVALUATION DE LA RESSOURCE PAR L'USAGE

Enfin, la dernière étape du projet s'est concentrée sur l'adéquation de la ressource termino-ontologique TriPhase avec le besoin du Département sous la forme d'un travail itératif d'évaluation par l'usage et d'enrichissement de *TriPhase*. Concrètement, la version courante de *TriPhase* est utilisée pour indexer automatiquement le texte plein des publications des chercheurs du Département. Les termes extraits des documents non indexés sont examinés pour être ajoutés prioritairement à *TriPhase*. La pertinence de l'indexation est également évaluée à travers les deux outils destinés au chef de Département et développés par MaIAGE : un moteur de recherche sémantique « Alvis IR-TriPhase » (Figure 1) et un outil d'analyse de corpus permettant une représentation graphique « ANStrat » (Figure 2).

## UN TRAVAIL COLLABORATIF FÉCOND

Ce projet a été conduit sur une période de 8 mois. Au terme de celui-ci étaient disponibles une première version de la ressource termino-ontologique *TriPhase* distribuée publiquement par le portail AgroPortal, l'outil d'analyse stratégique et le moteur de recherche sémantique. Ces résultats sont les fruits d'une collaboration étroite entre les documentalistes, les chercheurs en informatique et une ingénieure en ingénierie des connaissances. Le projet a répondu aux attentes du chef de Département lui permettant la visualisation des thèmes de recherche à différentes échelles au cours du temps. Il est à noter que la période de 5 ans sur laquelle a été menée l'étude n'est pas une durée suffisante pour observer une évolution significative des concepts liés aux thématiques scientifiques du Département. Des améliorations sont encore à apporter à la ressource *TriPhase*. Celle-ci reste en effet à compléter et à homogénéiser, puis à faire valider par les experts des différents domaines. Cette collaboration a en tout cas été enrichissante pour l'ensemble des acteurs du projet.

Du point de vue des chercheurs en informatique, il a apporté des éléments nouveaux dans l'approfondissement des méthodologies d'utilisation et d'évaluation des outils qu'ils développent, en particulier pour le travail collaboratif et à distance.

Du point de vue des documentalistes, les apports de ce travail collectif ont été nombreux : découverte de l'ingénierie de la connaissance, appropriation de nouvelles technologies... Les documentalistes ont en effet acquis de nouvelles compétences dans l'usage des technologies du web sémantique et ont appris à utiliser de nouveaux outils.

Les documentalistes - chacune spécialisée dans un domaine - ont également partagé et élargi la connaissance de leur Département de recherche. Elles ont aujourd'hui une meilleure compréhension des enjeux

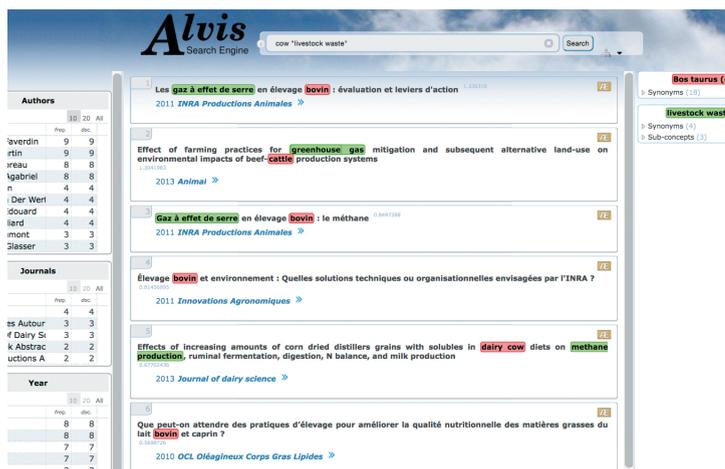


FIGURE 1. Exemple de requête sur moteur de recherche AlvisIR TriPhase.

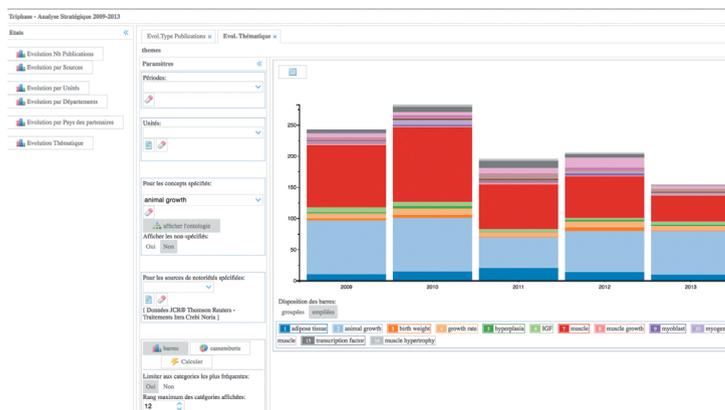


FIGURE 2. Exemple de l'évolution du thème « croissance » au cours du temps dans AnStrat.

scientifiques du Département. Elles ont aussi appris à modéliser les concepts et les sous-concepts relatifs au domaine de recherche du Département ; les outils informatiques utilisés produisant des termes extraits des publications, mais pas l'arborescence de la ressource termino-ontologique.

Au niveau humain, ce projet a renforcé le réseau des documentalistes dans une réelle énergie collective et créé une nouvelle dynamique de collaboration avec l'équipe de recherche en analyse de corpus et ontologie. Il a aussi permis de nombreux échanges avec les chercheurs du Département sur les différentes thématiques de recherche. Enfin, cette expérience a renforcé notre conviction que les professionnels de l'IST ont un rôle à jouer dans la formalisation des démarches et des connaissances, ceci dans la continuité des savoir-faire documentaires comme la sélection, la structuration, la qualification ou encore la normalisation de l'information.

**AGNÈS GIRARD**

Pour le réseau des documentalistes du Département Phase, Inra Rennes  
 agirard@rennes.inra.fr

**CLAIRE NÉDELLEC**

Pour l'équipe de recherche Bibliome de l'unité MaIAGE, Inra-Université Paris-Saclay, Jouy-en-Josas  
 claire.nedellec@jouy.inra.fr

- [6] Golik W., Bossy R., Ratkovic Z., Nédellec C. «Improving term extraction with linguistic analysis in the biomedical domain» in *Proceedings of the 14th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing'13)*, Special Issue of the journal *Research in Computing Science*, vol 70 ISSN 1870-4069, [http://rcs.cic.ipn.mx/2013\\_70/RCS\\_70\\_2013.pdf](http://rcs.cic.ipn.mx/2013_70/RCS_70_2013.pdf), 24-30 mars, Samos, Grèce, 2013.
- [7] <http://www.ncbi.nlm.nih.gov/mesh>
- [8] <http://aims.fao.org/standards/agrovoc/about>
- [9] <http://www.cabi.org/cabthesaurus/>

# Linked data, enjeu(x) et devenir

Les technologies du *Linked data* offrent de nombreuses possibilités à tous les producteurs et utilisateurs de données. Michael Jeulin, Aline Le Provost et Yann Olivier, qui connaissent bien la question à l'Abes, font le point sur ces perspectives élargies.

Les bibliothèques pratiquent depuis longtemps l'échange de données catalographiques. Elles savent que toute collaboration implique l'adoption de normes communes faisant en sorte que ce qui est échangé soit utile à tous : un socle commun bibliographique transportable et relativement fiable. Bien que robuste, ce système présente néanmoins un inconvénient majeur, celui de supporter difficilement les extensions et les spécialisations.

Le web de données permet, au contraire, d'exprimer et de partager des données produites de façon hétérogène. L'enjeu descriptif s'attachant autant au vocabulaire utilisé qu'à la chose décrite, le point de vue sur la ressource décrite devient aussi important et accessible que la ressource elle-même. Grâce aux identifiants uniques d'une part et aux ontologies d'autre part, l'interopérabilité est devenue une réalité prometteuse. Chaque métier, chaque organisme, chaque communauté peut produire et partager un référentiel (IdRef pour les chercheurs français, RNSR pour les structures de recherche, MeSH pour l'indexation matière médicale...) ou un vocabulaire (Bibframe, PRESSoo...), qui, dès lors que les expertises sont reconnues, feront foi et seront réutilisés.

L'autre grand intérêt du web de données est de faciliter l'accès aux contenus. L'ajout de métadonnées descriptives dans les pages HTML est un premier pas que beaucoup de sites ont aujourd'hui franchi. Les moteurs de recherches généralistes indexent ces données et les exploitent pour fournir à leurs utilisateurs de plus en plus d'information structurée – qui n'a jamais rencontré le *knowledge graph* de Google ?

Les possibilités sont pourtant loin d'être atteintes, en particulier dans le domaine de la publication scientifique. Si une partie des éditeurs fournit des notices en XML, ces dernières restent difficiles à exploiter et doivent être retraitées ; c'est ce que le hub de métadonnées s'attache à faire. Et s'il faut saluer les efforts de certains groupes commerciaux comme Nature, qui a publié en *Linked data* l'ensemble des titres de sa plateforme, on peut espérer la généralisation de ces efforts, notamment du côté des acteurs de l'*open access*.

L'exposition native en RDF d'un nombre croissant de sources permettrait de faire émerger de nouveaux outils de découverte, qui sachent vraiment intégrer des données hétérogènes, au lieu de juxtaposer des notices XML.

## DE L'EXPOSITION AU « TRIPLESTORE »

Mais le web de données est bien plus qu'un moyen d'exposer ou d'exporter les données. Les travaux de la BNF, du Centre International de l'ISSN et de l'Abes donnent à voir des cheminements similaires : la modélisation et l'exposition ont été d'abord l'occasion d'apprendre en marchant, de se familiariser avec la structure en réseau des graphes, très différente de celle des bases de données relationnelles. Chemin faisant, on a commencé (et ce n'est pas rien pour des bibliothécaires !) à se défaire du concept de *notice*, pour mieux constater que celle-ci est elle-même un document, juste une vue parmi d'autres sur des triplets décrivant une ou plusieurs ressources représentées par des URI.

Mais s'arrêter là aurait un goût de trop peu. Après avoir proposé du RDF pour les autres, la suite logique est d'essayer d'en exploiter soi-même toutes les potentialités : pour cela, il faut les charger dans une base RDF interne, et se mesurer à la puissance et à la souplesse du langage SPARQL. Celui-ci est un outil d'administration avec lequel les bibliothécaires ou documentalistes peuvent explorer et manipuler très finement leurs métadonnées, effectuer des croisements et des inférences complexes, interroger des sources hétérogènes, internes ou externes, qui peuvent y coexister à volonté. Et tout cela, en s'affranchissant, dans une certaine mesure, du recours aux informaticiens.

Il permet ensuite d'utiliser au maximum la logique des alignements pour connecter de façon cohérente toutes ces sources, et *in fine* de s'appuyer sur eux pour enrichir ou compléter ses propres contenus.

Le SPARQL Endpoint de démonstration élaboré par l'Abes donne un aperçu, partiel, de ce genre d'« atelier ».

Cette articulation ne va cependant pas de soi, car les formats actuels de production de métadonnées (Marc, EAD...) obéissent à des logiques et des structures qui n'ont pas été pensées pour le web de données. Le « *triplestore* » est l'environnement où l'on peut construire et tester de nouveaux modèles hors des contraintes des bases natives, en particulier pour accompagner la transition bibliographique vers les modèles FRBR, ou FRBRoo, et pour livrer un cadre commun à la documentation, aux archives et aux musées. C'est là que des initiatives comme Bibframe, PRESSoo ou le projet « Records in context » du Conseil international des archives donneront toute

[1] <http://bibframe.org/tools/editor/#>

leur mesure, ou devront évoluer encore. Enfin, le web de données ouvre de nouveaux horizons aux professionnels de l'IST, en leur donnant l'opportunité de valoriser leurs compétences traditionnelles, comme la création de classifications et de thésaurus, la normalisation et la structuration des informations. Bref, à ce stade, la base RDF tend à devenir une usine de retraitements de métadonnées provenant de multiples bases de production et de référentiels, pour y retourner ensuite, enrichies. C'est déjà beaucoup, mais ce n'est pas forcément tout. Plusieurs questions se posent : comment ces données sont-elles produites par les utilisateurs professionnels en amont ? A quoi ressembleront les interfaces qui permettront à l'utilisateur final d'explorer et d'interroger ces données ?

## QUELLES INTERFACES POUR LES DONNÉES RDF ?

Le catalogueur de demain n'aura pas pris RDF en LV1. Il serait absurde de l'imaginer « cracher » du RDF brut... aussi absurde que d'imaginer le catalogueur d'aujourd'hui saisir du Marc au kilomètre... ce qu'il fait pourtant...

Si on ne veut pas saisir du code, RDF ou Marc, il faut bien concevoir des interfaces permettant de saisir les données sans pour autant en maîtriser la syntaxe qui sert à les échanger ou les stocker. Bref, il faut un formulaire. Mais il y a formulaire et formulaire. Le défi est de concevoir des formulaires orientés « clavier » et non « souris » : le recours à la souris ralentit et exaspère ; il faut pouvoir saisir au kilomètre dans un formulaire... sinon, on préférera saisir du code. C'est un défi qui n'a rien de propre aux données RDF. Par contre, grâce au caractère auto-descriptif des données RDF, le formulaire peut être généré automatiquement : un schéma RDF sait que la propriété « marcrel:aut » s'affiche avec le libellé « auteur » en français et prend pour valeur l'identifiant d'une personne, que l'on peut ensuite interroger par son « foaf:name » dans un référentiel de personnes. Ce genre de mécanismes permet de donner plus d'autonomie aux bibliothécaires par rapport aux développeurs et favorise la collaboration entre experts données et « bibliothécaires système ». Le projet Bibframe propose ainsi un éditeur web qui peut servir à produire du RDF/Bibframe, ou n'importe quel autre vocabulaire<sup>1</sup>.

Comme pour les interfaces de saisie, les interfaces qui permettent d'interroger ou d'explorer les données RDF sont censées cacher la syntaxe. Mais comment faire pour rendre le RDF invisible tout en mettant entre les mains de l'utilisateur toute la puissance de SPARQL ? Le Graal serait de guider l'utilisateur de sorte que, pas à pas, sans le savoir, intuitivement mais rigoureusement, il construise une requête de plus en plus complexe. C'est exactement l'ambition de SPARKLIS, prototype conçu et développé par Sébastien Ferré, chercheur à Rennes. SPARKLIS est

une interface web qui permet de faire du SPARQL sans le savoir. Elle est la rencontre entre une interface de recherche à facettes, un constructeur de requêtes interactif et une interface en langue naturelle. Voici un exemple de construction pas à pas d'une recherche sophistiquée qui correspond à une requête SPARQL :

| step | query  |
|------|--|
| 1    | Give me something  |
| 2    | Give me a <b>Writer</b>  |
| 3    | Give me a Writer <b>that has a nationality</b>   |
| 4    | Give me a Writer <b>that has nationality Russians</b>  |
| 5    | Give me a Writer that has nationality Russians <b>and that has a birthDate</b>   |
| 6    | Give me a Writer that has nationality Russians and whose birthDate <b>is after 1800</b>  |
| 7    | Give me a Writer that has nationality Russians and whose birthDate is after 1800 <b>and that is the author of something</b>  |
| 8    | Give me a Writer that has nationality Russians and whose birthDate is after 1800 and that is the author of <b>a Book</b>   |
| 9    | Give me a Writer that has nationality Russians and whose birthDate is after 1800 and that is the author of <b>a number of Book</b>                                       |
| 10   | Give me a Writer that has nationality Russians and whose birthDate is after 1800 and that is the author of <b>the highest-to-lowest</b> number of Book                   |
| 11   | Give me a Writer that has nationality Russians <b>or something</b> and whose birthDate is after 1800 and that is the author of the highest-to-lowest number of Book      |
| 12   | Give me a Writer that has nationality Russians <b>or Russian_Empire</b> and whose birthDate is after 1800 and that is the author of the highest-to-lowest number of Book |

**Sparklis - Complex queries and exploratory search on SPARQL endpoints at your fingertips!**  
 NEW in 2016: nested aggregations, and computable/filtering expressions !!

SPARQL endpoint:

Your query and its current focus: **person:uk**

give me every writer  
 whose nationality is  
 Russians or Russia  
 and whose birth date is after 1800  
 and that is the author of the highest-to-lowest book  
 and for each result  
 give me the number of books

Sparklis suggestions to refine your query  
 The current focus is **person:uk** (click on different parts of the query to change it)

matches all of:  matches all of:   matches all of:

for each result    
 for each writer    
 for each birth date

Results of your query  
 results 1 - 10 of 13 Show 20 results

|   | writer                          | birth date | number of book |
|---|---------------------------------|------------|----------------|
| 1 | Shirvan Vozdek                  | 1976-11-01 | 1              |
| 2 | Fyodor Dostoyevsky              | 1821-11-11 | 11             |
| 3 | Vinogeny Zamyatin               | 1884-02-01 | 1              |
| 4 | Nikolai Ostrovsky               | 1904-05-29 | 2              |
| 5 | Joseph Brodsky                  | 1940-05-24 | 1              |
| 6 | Aleksey Konstantinovich Tolstoy | 1817-09-05 | 4              |
| 7 | Vasily Akayorov                 | 1912-08-20 | 1              |

Pour faire joujou avec SPARKLIS et les données du démonstrateur Abes, c'est ici : [http://tiny.cc/sparklis\\_abes](http://tiny.cc/sparklis_abes)

## CONCLUSION

Les données des bibliothèques sont des données comme les autres. Avec le web sémantique, les bibliothécaires échappent à une logique de niche, adoptent une approche et des techniques universelles. C'est bon pour nos ressources humaines : nous capitalisons sur nos compétences professionnelles pour leur donner une portée plus large et oser l'expérimentation. C'est bon pour nos marchés : de nouveaux prestataires peuvent nous proposer des produits, des services et du conseil. C'est bon pour le moral : créolisons les données !

**MICHAEL JEULIN**  
 jeulin@abes.fr

**ALINE LE PROVOST**  
 Le-provost@abes.fr

**YANN OLIVIER**  
 nicolas@abes.fr

Experts métadonnées à l'Abes

(Pleins feux sur...)

# Le Conservatoire National des Arts et Métiers, un ensemble « révolutionnaire »

**Avec des fonds à la fois bibliographiques et muséaux, le Conservatoire National des Arts et Métiers (Cnam) tient une double fonction de conservation et d'enseignement issue des grandes idées des Lumières et des Encyclopédistes.**

Le Conservatoire a été créé par la Convention en 1794 sur proposition de l'abbé Henri Grégoire en tant que centre d'enseignement pour « perfectionner l'industrie nationale » et en tant que Musée pour « protéger le patrimoine artistique de l'Ancien Régime afin d'en faire bénéficier le peuple ».

Avec un statut de Grand Établissement, membre de la Comue HeSam<sup>1</sup>, le Cnam Paris pilote aujourd'hui un réseau de 29 centres régionaux et de 158 centres d'enseignement, y compris à l'étranger. La vocation « étendard » du Cnam est la formation tout au long de la vie pour tous, à tout moment de son parcours professionnel. Parmi les réalisations intéressantes, on peut citer celle centrée sur le patrimoine, à savoir le Conservatoire Numérique des Arts et Métiers (Cnum)<sup>2</sup>, mais aussi celle tournée vers le réseau actuel et futur : c'est le guichet unique d'accès aux ressources documentaires.

## D'UNE BIBLIOTHÈQUE PATRIMONIALE...

Outre deux ouvrages « natifs », un anti-phonaire du XVI<sup>e</sup> siècle et un exemplaire de l'édition illustrée du Sacre de Louis XV relié aux armes du roi, échappés aux confiscations révolutionnaires, le fonds de la bibliothèque centrale a été constitué en puisant dans ces mêmes confiscations révolutionnaires. Le fonds ancien (80 000 volumes) est constitué de 6 incunables (dont la *Cosmographia* de Ptolémée dans une édition de Jean Reger de 1486), de 600 ouvrages du XVI<sup>e</sup> siècle, d'environ 2 000 du XVII<sup>e</sup> siècle et de plus de 4 000 du XVIII<sup>e</sup> siècle, tous liés aux sciences.

Avec la création des chaires d'enseignement du Conservatoire tout au long du XIX<sup>e</sup> siècle, la bibliothèque s'est enrichie d'une production éditoriale scientifique et technique de qualité : collections de mathématiques, de physique, de chimie, de mécanique et de

constructions civiles, avec de nombreux ouvrages sur les chemins de fer, l'électricité, la photographie ou l'automobile<sup>3</sup>. Outre ces fonds spécialisés, la bibliothèque centrale possède 450 volumes des manuels Roret, des années 1830 au début du XX<sup>e</sup> siècle, manuels traitant de l'histoire et de la pratique des métiers les plus divers (un par sujet). À cet ensemble s'ajoute une collection d'environ 1 000 titres de périodiques extrêmement variés : on peut ainsi consulter les *Mémoires de la Société des ingénieurs civils depuis*

1868 comme le *Bulletin de l'Institut du verre* (1947) ou le *Répertoire de législation et de jurisprudence forestières* (de 1862 à 1896). Il faut signaler aussi le fonds Godin, – Jean-Baptiste-André Godin, disciple de Charles Fourier, qui fonda le Familistère de Guise –, un fonds des catalogues des grands magasins, un fonds sur l'histoire des recherches sur l'électricité et ses applications, ou encore le fonds des expositions universelles. Depuis 2000, ces fonds entrent peu à peu dans le vaste corpus du Cnum, projet



➔ Bibliothèque centrale du Cnam

© Phot. Christophe Le Toquin / CC-BY-NC 2.0

commun de la Bibliothèque centrale, et du Cedric (Centre d'études et de recherche en informatique et communications du Cnam). 1365 ouvrages ont d'ores et déjà été numérisés en version intégrale, avec des planches en haute définition, et 7 manuscrits dont ceux de Vauban comme le *Traité de la défense des places* (sic). Les manuscrits et les archives, répartis entre le Musée et la Bibliothèque Centrale, sont signalés dans Calames depuis 2014.

### ... À UN GUICHET UNIQUE D'ACCÈS AUX RESSOURCES DOCUMENTAIRES

À cet inventaire patrimonial, s'ajoutent toutes les sources pédagogiques agrégées au fil du temps à la Bibliothèque Centrale mais aussi dans les chaires, à Paris ou en région. Cela concerne des ouvrages et revues dans l'ensemble des sciences couvertes par le Cnam mais aussi une quantité impressionnante de mémoires d'ingénieur, dont certains domaines présentent des correspondances d'une chaire à l'autre.

Ainsi, il existe au Mans un fonds important en géomatique, sciences de la Terre, géologie, mais aussi en expertise foncière et immobilière (École Supérieure des Géomètres et Topographes) que l'on peut mettre en résonance avec le fonds de la chaire de géologie appliquée au BTP et géotechnique à Paris, ou avec l'ICH (Institut d'études économiques et juridiques appliquées à l'immobilier, la construction et l'habitat) à Paris également. Il peut en être de même avec un fonds extrêmement intéressant sur le travail et la formation, au CDFT (Centre de documentation sur la formation et le travail) que l'on peut mettre en correspondance avec la bibliothèque de l'INETOP (Institut national d'études du travail et d'orientation professionnelle) rue Gay-Lussac à Paris, spécialisée en ergonomie, orientation scolaire et professionnelle, psychologie du travail et psychologie de l'enfant et de l'adolescent. Il ne faut pas oublier le fonds d'une richesse incomparable du CDHTE (Centre d'Histoire des Techniques et de l'Environnement) qui a rejoint depuis peu la Bibliothèque Centrale. Côté Musée, une base documentaire recense les objets muséaux – environ 80 000 objets et 15 000 dessins techniques formant le « Portefeuille industriel ». Par ailleurs, 500 films de conférences et d'expositions composent une base accessible directement sur les pages du Musée, en sus d'un catalogue recensant, lui, 18 000 monographies

et périodiques spécialisés. Des collections de catalogues de constructeurs y figurent également, fonds originaux en partie numérisés et accessibles via le Cnum<sup>4</sup>. Un fonds photographique diffusé par la photothèque du Musée vient compléter le tout. Mémoire et témoin du développement des sciences et des techniques du XVI<sup>e</sup> siècle à nos jours, il est constitué d'environ 25 000 clichés, et bénéficie d'un taux d'accroissement de plusieurs milliers d'images par an.

Les différents centres, que ce soit sur le site parisien ou dans les sites régionaux, ont ainsi traité des fonds spécialisés souvent très originaux dans des domaines d'excellence du Cnam parfois connexes, voire semblables. La multitude de ces fonds est gérée par des outils tout aussi multiples.

La pléthore de catalogues, tous très différents, ne représente cependant pas la masse totale des sources recensées au Cnam. Les mémoires d'ingénieurs, ainsi que les annales d'examens et les thèses sous embargo sont sur un serveur interne. Cependant, les publications des chercheurs, les revues éditées par le Cnam – entre autres, la *Revue française de métrologie*, les *Cahiers du Lise* et les *Cahiers d'histoire* – sont publiées sur les sites d'archives ouvertes comme revues.org ou OpenEdition, et surtout sur deux portails Hal dédiés (pour Lise et le CDFT). Ces publications, absentes des catalogues, sont signalées sur les pages des chaires elles-mêmes.

Cette profusion éparse d'informations a amené le Cnam à lancer le « guichet d'accès unique aux ressources documentaires » : il s'agit de l'un des trois projets phares du schéma directeur numérique national. L'idée est simple : interroger toutes les ressources acquises ou produites par le Cnam, en une seule fois et à l'aide d'une interface unique (catalogues, revues en ligne, archives, documents sonores ou vidéo, photographies...). Ce projet est destiné à tous les publics du Cnam, élèves et professeurs, ou lecteurs externes, selon des droits liés à des profils.

Après une phase de recensement de l'ensemble des ressources possibles, un cahier des charges a été établi et des actions à mener en interne ont été identifiées : mise en place d'une chaîne de numérisation et de stockage des mémoires d'ingénieurs, de cours en ligne, d'archives institutionnelles et harmonisation *a minima* de référentiels, en accord avec les référentiels actuels nationaux et internationaux.



➔ Catalogue des microscopes, Cnum.

La conduite du changement va impliquer de nombreux services, qui doivent passer au « tout numérique ». De nouvelles habitudes de travail doivent être instaurées en se fondant sur une vraie collaboration entre les services et l'instauration de nouveaux « workflows ». Ce projet absolument transverse est un défi intéressant avec, à la clé, un repositionnement des unités documentaires dans l'institution.

#### FRÉDÉRIQUE GAUDIN

Responsable Pôle Informatique Documentaire  
Cheffe de projet du Guichet unique – Direction  
Nationale du Numérique  
SCD du Cnam  
frederique.gaudin@lecnam.net

[1] HeSam Université (Hautes études Sorbonne arts et métiers) est une Comue (Communauté d'Universités et d'Établissements) qui fédère 12 établissements français d'enseignement supérieur, de formation et de recherche.

[2] Bibliothèque numérique en histoire des sciences et des techniques : <http://cnum.cnam.fr/>

[3] Exemples de manuels : *Construction des escaliers en bois, ou manipulation et posage des escaliers ayant une ou plusieurs rampes, des paliers ou des marches dans leurs encoignures...*; *Nouveau manuel complet des constructions agricoles, traitant des matériaux et de leur emploi dans les constructions destinées au logement des cultivateurs, des animaux et des produits agricoles, dans les petites, les moyennes et les grandes exploitations*

[4] [http://cnum.cnam.fr/thematiques/fr/9.catalogue\\_de\\_constructeurs/cata\\_auteurs.php](http://cnum.cnam.fr/thematiques/fr/9.catalogue_de_constructeurs/cata_auteurs.php)

# Retour sur les journées UKSG 2016



C'est au sud de l'Angleterre, dans l'agréable ville côtière de Bournemouth, que s'est tenu du 11 au 13 avril 2016 le congrès annuel de UKSG, occasion pour l'équipe Bacon de l'Abes de rencontrer ses homologues britanniques.

Si le sigle « UKSG » avait été originellement pensé pour « *United Kingdom Serials Group* », l'association a depuis largement dépassé son cadre initial. Par son champ d'application tout d'abord, étant donné que la plupart des ressources électroniques se trouvent désormais dans son giron. Mais également par l'espace territorial dans lequel elle s'inscrit, puisque l'avènement de cette manifestation de très grande envergure est aussi l'occasion, pour plus de mille interlocuteurs principalement venus de pays anglophones, de se rencontrer et d'échanger, tout en arpentant un salon professionnel composé d'une centaine de stands. UKSG réunit ainsi l'ensemble des acteurs de l'information scientifique anglophone, des bibliothécaires aux éditeurs en passant par les fournisseurs d'outils documentaires (outils de découverte, résolveur de liens, etc.). C'est dire la richesse des discussions auxquelles nous avons participé<sup>[1]</sup>. Nous nous concentrerons ici sur les interventions et les débats traitant de l'*open access*, thématique qui occupa une place de choix dans ce congrès, et sur nos actions effectuées dans le cadre de notre travail pour Bacon, la Base de Connaissance Nationale de l'Abes.

## L'*open access* à l'honneur

Par la qualité de leur propos et leur portée, les deux interventions inaugurales ont particulièrement retenu notre attention. Directrice exécutive de la SCONUL<sup>[2]</sup>, Ann Rossiter a tout d'abord axé son discours sur la gestion des relations entre bibliothèques et éditeurs dans le but de favoriser leur objectif commun : l'accès au Savoir mondial. Après avoir rappelé les difficultés budgétaires et les baisses d'effectifs que connaissent actuellement les bibliothèques britanniques (mais aussi, plus largement, l'ensemble des bibliothèques universitaires à l'échelle internationale) ainsi

que ses conséquences sur les relations avec les éditeurs scientifiques, Ann Rossiter a exposé un argumentaire pertinent sur la nécessité de tendre davantage vers le mouvement de l'*open access*. Elle a, dans cette optique, enjoint au monde éditorial d'embrasser rapidement cette transition et d'ajuster ses pratiques en concertation avec les bibliothécaires. C'est-à-dire comprendre et tenir compte de la réalité budgétaire des bibliothèques universitaires pour adapter en conséquence leurs modèles économiques et améliorer la qualité des services et ce, dès maintenant. En effet, quelle que soit la position adoptée par les éditeurs pour y faire face, ce mouvement prend inéluctablement, et chaque année, de l'ampleur. Continuer à refuser ce « nouveau » modèle est donc une attitude inadéquate, voire contre-productive, qui pourrait de surcroît conduire certains éditeurs à disparaître, purement et simplement.

En complément de cette intervention, Michael Jubb, ancien directeur du *Research Information Network*<sup>[3]</sup>, a dressé un panorama extrêmement précis de l'implantation de l'*open access* en Grande-Bretagne. Les fondements de la discussion ont reposé sur le constat que, depuis le Rapport Finch de 2012<sup>[4]</sup>, le Royaume-Uni est devenu l'une des figures de proue de l'*open access*, notamment grâce à quatre sources de financement majeures outre-Manche (Wellcome Trust, Research Councils UK, HEFCE<sup>[5]</sup> et Horizon 2020) qui en promeuvent l'utilisation, et ce malgré l'immense diversité de leurs politiques en la matière : 447 politiques différentes cohabitent actuellement en Europe. Cette position novatrice de l'information scientifique britannique sur l'*open access* s'affirme également dans les statistiques de Scopus présentées par Michael Jubb. En effet, l'ensemble des revues scientifiques britanniques présentes dans la base d'Elsevier et inscrites dans une politique d'accès ouvert entre 2012 et

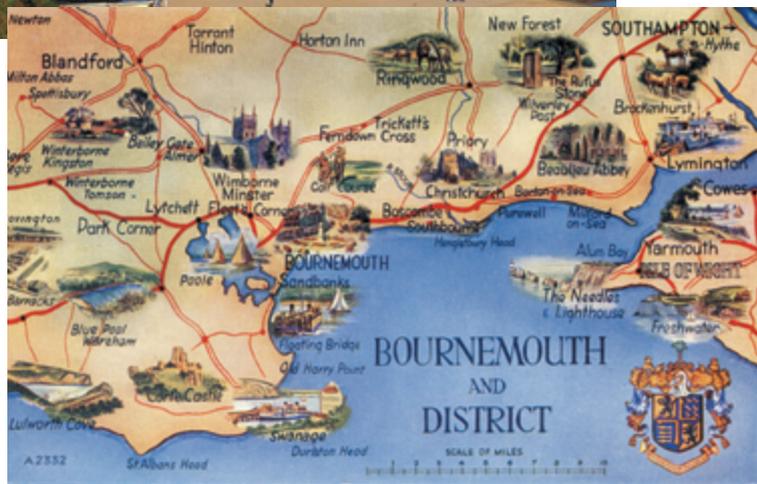
[1] Pour voir les enregistrements des interventions : <https://tv.theiet.org/?event=3583p-f>

[2] *Society of College, National and University Libraries* (Royaume-Uni et Irlande).

[3] <http://www.rin.ac.uk/>

[4] <http://www.researchinfonet.org/wp-content/uploads/2012/06/Finch-Group-report-FINAL-VERSION.pdf>

[5] *Higher Education Funding Council for England*



« Bournemouth and District »  
Carte postale de  
JValentine & Sons, Ltd,  
Dundee and London.

2014, toutes catégories confondues (Gold avec ou sans *Article processing charges* – APC -, dépôts institutionnels, revues hybrides et d'*open access* avec embargo) est passé de 64,5% à 68,2% de la totalité des périodiques de cet éditeur. De même, la proportion d'auteurs britanniques publiant dans une revue scientifique en *open access* a bondi de 4% entre 2012 et 2014, pour atteindre 80,5%. Consécutivement, le nombre de titres dans la base d'Elsevier uniquement accessibles via abonnement et les revues payantes dans lesquelles les chercheurs anglais publient, ont marqué une baisse de 5% durant ce même laps de temps.

Si le monde de l'information scientifique anglophone paraît donc s'engager véritablement dans cette voie, certaines nuances ont néanmoins été apportées à la fois par les éditeurs scientifiques, les institutions et les chercheurs.

Du point de vue des éditeurs, le manque de clarté des politiques d'*open access* a été pointé du doigt. Côté institutions en revanche, c'est surtout l'extrême variété et complexité des modalités de la Voie Verte –majoritairement choisie dans le cadre des dépôts institutionnels- qui a été mise en exergue.

Quant aux universitaires britanniques, ils semblent particulièrement enclins, pour près de 64% d'entre eux, à publier de préférence dans des revues hybrides (notamment pour des raisons d'*impact factor*), malgré les implications financières préjudiciables de ce modèle, puisqu'une très forte augmentation globale du coût des APC est observée. Ces derniers ont été globalement multipliés par six entre 2012 et 2014, pour atteindre 33 millions de Livres sur l'ensemble des universités du Royaume-Uni. Malgré ce dernier constat, nous pouvons bien qualifier de succès l'adoption de l'*open access* par le milieu scientifique britannique, même s'il reste encore beaucoup de travail à poursuivre dans cette voie.

La qualité initiale des interventions constituant les sessions plénières a été enrichie par d'abondantes sessions parallèles abordant les thèmes les plus variés : le phénomène des éditeurs-prédateurs, l'interopérabilité des plateformes et des services, le développement de maisons d'édition au sein de bibliothèques notamment... Ces sessions ont permis des échanges opportuns et particulièrement riches entre les différents participants, et ce, malgré la frustration certaine, pour tout auditeur, de ne pouvoir assister à l'ensemble des présentations souhaitées.

### **Bacon – KB+ : une collaboration renforcée**

Plus pragmatiquement, et du point de vue de Bacon, un cran supplémentaire a été franchi dans le partenariat avec KB+<sup>6</sup> puisque le congrès a également été l'occasion d'échanges pertinents et particulièrement constructifs avec nos homologues britanniques. Cette rencontre a notamment permis de mieux comprendre les objectifs, workflows et usages respectifs que nous faisons de nos fichiers Kbart (de KB+ vers Bacon mais aussi, plus récemment, de Bacon vers KB+). Elle a également offert l'opportunité de dégager une ébauche de

[6] Knowledge Base+, service du Joint Information Systems Committee (Jisc)

visées communes, tant dans les exigences que nous pourrions conjointement attendre des fournisseurs d'outils documentaires, que dans nos relations avec les éditeurs et les développements envisagés. La communication et la mutualisation des deux services vont donc continuer à s'intensifier.

Dans l'optique d'inscrire nos activités dans cette dynamique et d'apporter notre contribution à ce processus, l'équipe de Bacon postule d'ailleurs actuellement à l'intégration du groupe de travail NISO/UKSG sur la recommandation Kbart, groupe co-présidé par Magaly Bascones, directrice de KB+. Notre présence nous permettrait non seulement d'affiner l'interprétation des différents items de la recommandation, mais aussi, et surtout, de faire valoir les positions que nous avons adoptées jusqu'à maintenant sur celles-ci.

## Conclusions

C'est la seconde fois qu'une délégation de l'Abes se rend à ce rassemblement de très grande ampleur, puisque les deux responsables du département Adele étaient déjà présents à Harrogate en 2014. Cette année, seules deux institutions françaises ont fait le déplacement jusqu'à Bournemouth (l'Abes et l'Inist-CNRS). Nous sommes persuadés qu'il est nécessaire que les bibliothécaires français participent davantage à cette manifestation internationale. Le congrès UKSG permet en effet aux professionnels des bibliothèques de France d'obtenir une visibilité accrue chez leurs voisins. Par-delà les frontières étatiques et les différences indéniables dans le fonctionnement des diverses institutions, les bibliothécaires du monde entier se trouvent confrontés aux mêmes évolutions professionnelles, aux mêmes contraintes budgétaires, aux mêmes interrogations philosophiques et prospectives. *Quid* du mouvement général de l'*open access*, de ses avantages, limites et implications diverses, de l'amélioration du signalement de la documentation électronique, de l'administration et de la promotion des dépôts institutionnels... ? La participation des bibliothécaires français à ces réflexions est indispensable.

Concernant l'Abes, la pertinence de sa présence à ce *symposium* transnational vient d'être confirmée par la récente invitation de Magaly Bascones à participer au prochain congrès UKSG de 2017, où nous nous joindrions à KB+ et CUFTS afin de présenter nos trois bases de connaissances non commerciales.

Nous espérons en tout cas avoir convaincu nos collègues français de nous y rejoindre : le déplacement en vaut la peine.

**BERTRAND THOMAS**

Gestionnaire de métadonnées,  
membre de l'équipe Bacon à l'Abes  
thomas@abes.fr

## PROGRAMME DES SESSIONS PLÉNIÈRES DU CONGRÈS UKSG

### LUNDI 11 AVRIL

#### Plenary Session 1 IMPACT

Tregonwell Hall  
Chair : Kate Price,  
King's College London

11:00

*How far has the UK got towards open access and what have been the costs and implications ?*  
Michael Jubb, Consultant

10:30

*Managing relationships between libraries and publishers for greater impact*  
Ann Rossiter, SCONUL

11:30

*Data diving : understanding cause and effect in reputation management*  
Charlie Rapple, Kudos

### MARDI 12 AVRIL

#### Plenary Session 2 USER EXPERIENCE

Tregonwell Hall  
Chair : Jill Taylor-Roe,  
Newcastle University

9:30

*Open Music Library : a community-curated open index of the world's scholarly music resources*  
André Avorio, Alexander Street

9:00

*Ethnographic approaches to the practices of scholarly communication : tackling the mess of academia*  
Donna Lanclos, University of North Carolina, Charlotte

10:00

*Engaging students, shaping services : the changing face of student engagement at the Hive*  
Sarah Pittaway,  
University of Worcester

### MERCREDI 13 AVRIL

#### Plenary Session 3 INNOVATION AND INSPIRATION

Tregonwell Hall  
Chair : Incoming Chair, UKSG

12:45

*Investing in scholarly futures : communities, funding and the reimagining of research Communications*  
Cameron Neylon, Centre for Culture and Technology,  
Curtin University

12:15

*Google – Digital Child*  
Emma Mulqueeny, Elbi Digital

# #BiblioDebout

Initiative du collectif SavoirsCom1<sup>1</sup>, la BiblioDebout est une bibliothèque participative en plein air dont le fonds se constitue grâce aux dons des publics. Ses deux principes fondamentaux sont le partage et la circulation du savoir.

L'idée n'est pas nouvelle et les influences sont nombreuses – on peut citer par exemple la Bibliothèque du peuple d'Occupy Wall Street, celle du Parc Gezy à Istanbul, ou encore le concept des *Bücherschränke* en Allemagne... – mais la BiblioDebout est la première bibliothèque participative à faire un lien direct avec la notion de « Commun » : une communauté se rassemble pour prendre en charge une ressource, la mettre en partage et en assurer sa pérennité<sup>2</sup>.

La première BiblioDebout est apparue début avril, dès la deuxième semaine du mouvement NuitDebout, Place de la République à Paris. Le succès est immédiat : à la fin du premier weekend, on comptabilise déjà près d'un millier de dons.

Très vite aussi, le groupe de « gestionnaires » initial s'élargit pour accueillir des personnes de tout âge et de tout horizon : étudiants, journalistes, chômeurs, libraires...

L'initiative est aujourd'hui complètement autonome et la BiblioDebout s'est développée, au-delà de la capitale, dans bien d'autres villes en France : Toulouse, Lyon, Montpellier, Marseille, Rennes... Les points communs avec la gestion d'une bibliothèque « classique » existent, bien sûr : étiquetage et « marquage » des ouvrages (ce qui permet de laisser une trace de leur passage dans la BiblioDebout), recueil des suggestions des usagers (on trouve par exemple des demandes de livres pour apprendre le français, notamment de la part des migrants), médiation et conseils aux lecteurs... La BiblioDebout toulousaine a même mis en place un système de catalogage pour signaler ses « collections », quand bien même la cohérence et la sta-

bilité du fonds ne peuvent être garanties et le système de prêt est inexistant. Il s'agit ainsi de gérer un flux plutôt qu'un stock. En outre, à Paris, les membres de BiblioDebout ont choisi de ne pas effectuer de sélection sur les ouvrages en acceptant et en mettant à disposition tous les dons. L'idée est d'éviter tout tri politicié.

Depuis sa création, la BiblioDebout semble bien fonctionner. La majorité de ses usagers joue le jeu du partage et de la réciprocité, et l'initiative est fortement soutenue par les membres de NuitDebout : elle a ainsi rejoint le groupe des « Commissions structurelles » du mouvement<sup>3</sup>.

Finalement, les seules vraies contraintes proviennent des pouvoirs publics locaux. En effet, les municipalités interdisent la mise en place de structures pérennes, ce qui implique de reconstruire la Bibliothèque à chaque nouvelle mobilisation. Plus fâcheux : après la manifestation du mardi 14 juin contre la Loi Travail, la BiblioDebout lyonnaise s'est vue obligée de plier boutique pour « occupation du domaine public sans autorisation ». Cette interdiction policière est une première depuis le début de l'initiative et dans le climat national actuel, il faut sans doute s'attendre à d'autres prohibitions de ce genre. Mais nos ingénieux BiblioDebouticiens ne tomberont pas si facilement : les astuces sont encore nombreuses pour continuer de partager le savoir en public en toute liberté !



Phot. Charlotte Henard (CC-BY SA 2.0)

Merci à Lionel Maurel, juriste et bibliothécaire, membre du collectif SavoirsCom1, pour l'éclairage apporté. Contact : [calimaq@gmail.com](mailto:calimaq@gmail.com)

**MARION GRAND-DÉMERY**

*AbesDebout*  
grand-demery@abes.fr

[1] SavoirsCom1 est un collectif engagé pour le développement de politiques et d'initiatives liées aux communs de la connaissance : <http://www.savoirscom1.info/>

[2] A ce propos, lire le billet de Lionel Maurel sur son blog S.I.Lex : <https://scinfolex.com/2016/05/11/en-quoi-la-bibliodebout-constitue-t-elle-un-commun/>

[3] Les Commissions structurelles sont l'ensemble des services gravitant autour du mouvement NuitDebout ; on y trouve par exemple la Commission Restauration et la Commission Infirmerie.

## Au revoir et merci !

Après trois années passées à l'Abes – et quatre numéros d'*Arabesques* –, je suis appelée vers d'autres horizons. Grand merci aux collègues du comité de rédaction, à Anne Ladevie et à l'ensemble des contributeurs avec qui j'ai eu la chance et le plaisir de travailler au cours de ces quatre numéros !

**Marion Grand-Démery**

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

(Portrait)

## Marie-Line GUILLAUMÉE

Coordinatrice SU  
Correspondante catalogage  
à la Bibliothèque interuniversitaire  
de la Sorbonne (BIS)



### **Quelles sont vos fonctions à la BIS ?**

Je suis coordinatrice Sudoc, correspondante catalogage depuis 2003 et co-responsable de la Bibliothèque professionnelle. La BIS est rattachée pour sa gestion à l'Université de Paris I. Ses collections très riches, spécialisées en lettres et sciences humaines, s'adressent à un public de chercheurs (2 millions de volumes, 22 700 lecteurs inscrits).

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

Entrée à la Sorbonne en 1991 comme magasinier, j'ai réussi le concours de BAS en 1998 et réintégré la BIS en 2000 où j'ai été promue bibliothécaire en 2014. Mon histoire avec la Sorbonne a commencé par un coup de foudre qui ne s'est jamais démenti. J'ai aussi eu la chance de rencontrer des professionnels passionnés qui m'ont transmis leur amour du métier et du fonds de la Sorbonne. Un moment important : 2001 et mes débuts dans WinIBW, outil de catalogage partagé que j'apprécie depuis ses origines. Depuis 2004, ma fonction de coordinatrice m'a permis de toucher du doigt le cœur du système : le travail en réseau. Les mutations en cours dans le domaine des données nous rappellent les services que celles-ci peuvent rendre aux chercheurs lorsqu'elles sont propres et bien exploitées. Une nouvelle ère s'ouvre pour nos catalogues.

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

Mon premier vrai souvenir remonte à mes premières Journées Abes. Avec mes collègues, nous parlions une langue commune : le catalogage. Une communauté d'esprit préoccupée par la même interrogation : comment faire de son métier un outil utile pour les lecteurs ? Il ne s'agit plus de parler de « ; » (même si « couv. ill. en coul. » n'est pas sans poésie) mais bien de comprendre que nous sommes à une étape cruciale où bibliothèques, musées et archives vont pouvoir partager leurs données pour être enfin visibles sur le Web.

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

Étant très impliquée dans le réseau, l'Abes m'a sollicitée pour devenir formatrice relais en 2008 : avec 12 collègues, nous élaborons un support et assurons des formations sur l'outil WinIBW. Je participe aussi à plusieurs groupes de travail du programme « Transition bibliographique » où nous rédigeons les futures normes et réfléchissons à leurs impacts pour les catalogueurs.

### **Quelles en sont les répercussions dans l'exercice de vos fonctions ?**

J'ai toujours eu à cœur de partager mes connaissances et d'aider les collègues en rédigeant des mémos utiles à la compréhension des changements actuels. Deux lignes directrices sous-tendent mon parcours professionnel : la gestion des métadonnées et la formation.

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

L'Abes doit rester connectée et accompagner les gestionnaires de métadonnées dans la Transition bibliographique.

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

Travailler dans une grande bibliothèque me permet d'être très active dans le réseau. J'ai pu ainsi réaliser, en collaboration avec des collègues de la BIS, plusieurs chantiers d'envergure, utiles au réseau (traitement des *Monumenta Germaniae Historica*, de la Pléiade, participation à un chantier « Cercles » avec l'École française d'Athènes...). Ces activités m'ont permis de me confronter à tous types de problèmes et de réfléchir aux moyens de les résoudre.

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

La dichotomie entre le travail du catalogueur qui essaye de suivre les consignes Abes, les normes de catalogage et la réalité économique qui nécessite une adaptation à des contingences pouvant remettre en cause la qualité des données. Celle-ci est pourtant indispensable à une bonne exploitation par une machine.

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

Un colibri « *allant chercher quelques gouttes d'eau dans son bec pour les jeter sur le feu...* ». Il fait sa part, comme chaque catalogueur du réseau. A l'heure où le savoir-faire des catalogueurs va permettre aux bibliothèques d'être présentes sur la toile, nous allons enfin sortir de ce cliché de coupeur de cheveux en quatre (à lunettes et à chignon, bien sûr).

### **Votre expression favorite ?**

L'expression tronquée : « *errare humanum est* » que je complétera par une citation de Confucius : « L'homme sage apprend de ses erreurs, l'homme plus sage encore apprend des erreurs des autres. »