

Améliorer et enrichir le signalement

Synthèse établie par Serge Bauin et Raymond Bérard à partir des documents de BSN 3

Le comité de pilotage BSN a retenu cinq recommandations pour le segment du signalement :

- inciter les producteurs publics de métadonnées à en autoriser la réutilisation gratuite et inciter les bibliothèques à les utiliser ;
- construire un référentiel de la production scientifique française en s'appuyant sur des référentiels communs aux universités et organismes (unités, thématiques, auteurs...) ;
- établir un *continuum* entre les outils existants de signalement et les plateformes d'accès au texte intégral ;
- promouvoir au niveau européen des systèmes libres de référencement accompagnés d'une recommandation pour créer une agence internationale de référencement croisé des documents scientifiques ;
- créer un dispositif fédérant le signalement de toute la documentation accessible sur le territoire et faire le lien avec la ressource dans le cadre du contrôle des droits.

Les actions retenues au titre de 2012 sont au nombre de deux :

- étudier un dispositif fédérant le signalement de toute la documentation accessible sur le territoire et faire le lien avec la ressource dans le cadre du contrôle des droits : le levier de ce dispositif sera ISTEEX ;
- spécifier un référentiel de la production scientifique française en s'appuyant sur des référentiels communs aux universités et organismes (unités, thématiques, auteurs...).

Différents chantiers, repris ci-dessous, sont d'ores et déjà en cours d'étude ou de réalisation.

Signalement national des ressources électroniques courantes

Cet objectif figure au projet d'établissement 2012-2015 de l'ABES, dont BSN constitue le cadre structurant. Il s'agit de réaliser une cartographie complète de l'offre documentaire française comprenant l'intégralité des ressources en ligne disponibles sur le territoire, quel que soit leur mode d'acquisition :

- licences nationales : les éditeurs en livreront directement les métadonnées à l'ABES qui en assurera le signalement national. Ce signalement pointera pendant une étape transitoire sur les sites des éditeurs puis sur la plateforme d'agrégation du texte intégral (confiée à l'INIST-CNRS) prévue dans le projet ISTEEX ;
- abonnements courants des établissements.

Le signalement des ressources en ligne se fera au niveau le plus pertinent pour l'utilisateur final (article de revues, chapitre ou titre pour les monographies). L'interface publique permettra à l'utilisateur final d'arriver le plus rapidement possible au texte intégral, en fonction des droits

ouverts, ou lui proposera des alternatives (PEB, *pay per view*, etc.) le cas échéant.

Une étude a été lancée pour estimer la faisabilité et le coût de ce dispositif. Trois scénarios de base, pouvant être amendés, précisés et enrichis par le consultant, sont proposés. Ils n'excluent pas la possibilité de proposer d'autres scénarios.

• **Scénario 1** : la solution s'appuie sur un outil de découverture pour l'ensemble des ressources en ligne. La récupération de toutes les métadonnées se fait par cet intermédiaire. L'accès au texte intégral s'appuie sur un résolveur de lien central articulé avec des résolveurs locaux. L'interface publique est celle de l'outil de découverte ou une surcouche.

• **Scénario 2** : la solution s'appuie sur un développement spécifique. Les métadonnées, concernant autant la description (y compris le texte intégral, pour faire de l'indexation) que les droits ouverts, sont récupérées éditeur par éditeur sur le modèle de l'outil Trove¹. L'interface publique est un outil du type Vu-Find.

• **Scénario 3** : La solution s'appuie sur Google Scholar. L'accès au texte intégral repose sur un résolveur de lien central articulé avec des résolveurs locaux. L'interface publique est Google Scholar. La géolocalisation permet de fournir les liens spécifiques.

L'ABES a lancé un appel d'offres le 23 mai 2012 pour choisir un consultant qui évaluera ces trois scénarios en termes de coût, de *workflow* et d'articulation avec les autres projets en cours : ISTEEX, hub de métadonnées, système de gestion de bibliothèque mutualisé (SGBM), projets de recherche SudocAd (automatisation du lien aux autorités Sudoc) et Qualinca (Qualité et interopérabilité de grands catalogues documentaires).

Le lien entre tous ces projets, c'est la base de connaissance : le SGBM et l'outil de signalement des ressources électroniques courantes pourraient être alimentés par le hub dont les métadonnées seraient enrichies par les traitements automatiques des projets de recherche SudocAd et Qualinca. Des réponses sont attendues de cabinets de consultants ayant déjà été impliqués dans des projets de cette ampleur. La notification du marché est prévue pour début juillet, l'étude finale devant être remise en décembre 2012.

Qu'est ce que le projet de hub de métadonnées ?

Ce projet vise à redistribuer dans n'importe quel format des métadonnées enrichies récupérées dans n'importe quel format auprès des éditeurs. Le hub permettra de convertir les formats, d'apporter une plus-value aux métadonnées en termes de structuration, de richesse et d'interconnexion.

En entrée, le hub acceptera des notices en provenance des éditeurs, des licences nationales, de l'Open Access, des bases d'organismes publics et du Sudoc lui-même. Les traitements automatiques apportés par les projets Sudoc-Ad et Qualinca, qui devront très vraisemblablement être complétés par des traitements manuels, porteront sur l'identification des données livrées (ex. : états de collection Springer), la normalisation (ex. : dates, noms), les corrections, l'affinage, les enrichissements (ex. : liage à des autorités ou cadres de classement), les conversions, les redistributions.

En sortie, le hub alimentera ISTEEX, le Sudoc, les bases de connaissance commerciales, les SIGB des établissements, le web de données.

Un pilote a démarré en septembre sur une sélection de corpus : ebooks et revues Springer ; Dalloz ; notices de thèse Sudoc avant 1985 ; notices de Revues.org.

Référentiel de la production scientifique française

Des masses d'informations sur ou liées à la production scientifique des chercheurs et unités de l'enseignement supérieur et de la recherche sont produites par différents acteurs aux objectifs divers, mais elles restent dispersées, parfois locales et souvent disparates. On ne dispose donc pas d'une vision claire de la production scientifique française. Le projet de construction d'un référentiel de la production scientifique française vise à améliorer cette situation et à aboutir à un recensement le plus complet possible de cette production.

Un comité de pilotage composé de représentants des universités et des organismes (responsables d'archives, direc-

teurs d'unité, producteurs d'indicateurs scientifiques...) a été constitué en avril 2012 et s'est réuni deux fois.

La première tâche qu'il s'est assigné est de faire un état des lieux des réservoirs existants (archives ouvertes, archives institutionnelles, bases de données...), des référentiels (unités, auteurs...), des pratiques et des difficultés rencontrées par les différents acteurs en matière de recensement de la production scientifique. L'objectif pour fin 2012 est de clarifier le périmètre et le positionnement du projet par rapport aux systèmes existants et aux projets connexes et de dégager des pistes de construction du référentiel de la production scientifique française basées sur un principe de capitalisation et de mutualisation des efforts.

Raymond Bérard
Directeur de l'ABES

Serge Bauin
Directeur de l'Information scientifique et technique,
CNRS

¹ Outil de découverte de la Bibliothèque nationale australienne : <http://trove.nla.gov.au/>



© Lagui / 123 RF

Les échelles de mutualisation pour le numérique : entre signalement local et national / entre signalement généraliste, disciplinaire, institutionnel

Les chercheurs et leurs institutions signalent leur production dans des bases institutionnelles ou thématiques. Selon les besoins, le signalement réparti dans ces différentes bases peut en être rapatrié dans une ou d'autres bases. Pour que cette opération de rapatriement soit la moins contraignante possible, les différents signalements doivent respecter des règles communes.

C'est à ce niveau que **les fichiers d'autorités** jouent un rôle décisif. Un même référentiel doit être utilisé chaque fois que c'est possible afin d'unifier les accès, de limiter les doubles saisies et de permettre la navigation à partir d'une même autorité entre plusieurs applications. Ce référentiel existe : l'application IdRef développée par l'ABES, en permet la réutilisation par tous les établissements de l'ESR. IdRef notamment permet d'enrichir des métadonnées saisies *a minima* par un chercheur dans un dépôt d'archives (HAL ou dépôt institutionnel). IdRef en permet la normalisation et l'enrichissement par un documentaliste, par les communautés, et peut être articulé avec les annuaires existants, en établissant des liens avec les affiliations.

Cette réutilisation n'est qu'un exemple de la complémentarité entre opérateurs qui doit se généraliser, chacun se concentrant sur son domaine d'excellence et mettant ses outils à la disposition des autres afin d'éviter les doublons.

Source : Rapport *Bibliothèque scientifique numérique : stratégie 2012-2015*, mars 2012, p. 27-28.