

# Synthèse de l'étude Découverte pour l'ABES

## Introduction

L'ABES joue un rôle important dans la découverte des ressources nationales grâce à ses services SUDOC (catalogue collectif), Calames (fonds d'archives), Numes (corpus numérisés), et les Signets des Universités (sites web pertinents). Pour pouvoir analyser les possibilités d'étoffer ces services nationaux à l'aide d'un outil de découverte à l'échelle du web destinés aux revues et livres électroniques, l'ABES a mandaté une étude d'exploration des trois scénarios suivants :

- **Un scénario de réalisation interne, élaborant un service de découverte objet d'un développement nouveau** : Ce scénario s'appuie sur Trove, le moteur de recherche en ligne mis au point par la Bibliothèque nationale d'Australie. Dans ce cas, l'outil de découverte français ferait l'objet d'un développement nouveau utilisant des logiciels libres : les métadonnées et l'index du texte intégral seraient extraits au niveau de chaque éditeur. Les utilisateurs membres d'une bibliothèque abonnée à cette ressource bénéficieraient d'un accès immédiat. Pour les autres, un dispositif de distribution devrait être proposé.
- **Un scénario utilisant des outils de découverte présents sur le marché** : Dans ce cas, il s'agirait d'un outil de découverte de ressources national similaire qui reposerait sur une adaptation d'outils existants, comme Summon (Serial Solutions), Primo (Ex Libris), EBSCO Discovery (EBSCO) ou OCLC WorldCat local.
- **Un scénario qui établirait une collaboration avec Google Scholar** : Pour ce scénario, Google Scholar est l'élément central, proposant une interface publique ainsi que le moteur de recherche. L'accès au texte intégral s'effectuerait via un résolveur de liens.

Cette étude a été menée par Pleiade Management and Consultancy en trois phases : la phase 1 a élaboré le cahier des charges de cet outil de découverte national, la phase 2 a étudié les solutions existantes ou en développement dans d'autres pays puis la phase 3 a analysé la faisabilité des trois scénarios décrits plus haut et leur adéquation avec le cahier des charges défini par la première phase. Pour finir, les résultats ont fait l'objet de discussions lors d'un certain nombre d'ateliers avec des membres du personnel de l'ABES, aboutissant ainsi à une feuille de route pour la suite des développements. Cette étude s'interpénètre avec trois autres projets constitutifs intervenant dans la construction de l'infrastructure documentaire nationale : le projet hub de métadonnées, où l'ABES élargit son expertise dans le domaine des métadonnées à l'environnement Internet, un projet sous la direction de l'Agence en collaboration avec un certain nombre de bibliothèques devant élaborer un appel d'offres pour la réalisation d'un système intégré de bibliothèques mutualisé en nuage et le projet ISTEEX qui construit une plate-forme nationale destinée à des contenus disponibles sous licence.

## Principaux résultats des études de scénarios

- **Plate-forme de (méta)données** : Les services de découverte développés intra-muros qui ont produit leurs propres index centralisés dédiés à la littérature savante n'ont pu atteindre qu'une couverture très limitée. Dans ce domaine, l'expérience de l'outil de découverte Trove en Australie et de la *Suchkiste* en Allemagne montre clairement la faisabilité d'un index centralisé sur une *sélection* de la littérature savante mondiale. Cependant, bâtir un index centralisé de la *totalité* de cette littérature au niveau mondial est jugé comme un effort important, mobilisant beaucoup de main-d'œuvre et, pour cela, n'est pas considéré comme faisable. À l'inverse, les outils de découverte à l'échelle du web prétendent couvrir largement la littérature savante dans le monde et indexer une bonne part du texte intégral. Concernant la qualité des métadonnées, tous les systèmes de découvertes sont dotés de dispositifs exploitant les métadonnées d'autres

sources comme les bases de données d'analyse et de signalement pour enrichir celles fournies par les éditeurs primaires grâce à des mécanismes de mise en correspondance et de fusion.

- **Services de localisation** : Le résolveur de liens allemand EZB (qui s'appuie sur les données du catalogue collectif de revues électroniques EZB) offre un très bon exemple de niveau national. Il peut être utilisé par des bibliothèques qui ne possèdent pas de résolveur pour servir leurs utilisateurs finals. De plus, le résolveur EZB peut être utilisé par des bibliothèques ayant le leur soit en employant EZB comme cible de leur résolveur, soit en exploitant les données de la base de connaissances d'EZB dans leur propre base de connaissance sous-jacente à leur résolveur de liens en local. De plus, toujours en Allemagne, *Journal Online & Print* (JOP) est un cas (probablement unique) de webservice mis à la disposition des utilisateurs finals portant sur les fonds de revues papier de leur bibliothèque. Ce service leur indique la disponibilité d'une revue dans le fonds papier de leur bibliothèque, à partir des catalogues collectifs ZDB et EZB. Associés au résolveur EZB, ces services aident les utilisateurs à localiser le texte intégral dans leur bibliothèque. Les outils de découverte à l'échelle du web nécessiteraient que soit développé un service de localisation de niveau national. S'il est vrai que l'interopérabilité entre les différents résolveurs de liens et les diverses solutions de découverte ne pose aucun problème, une base de connaissances devra servir de socle à un résolveur de liens national. Les expériences actuelles en matière d'échanges entre les bases de connaissances des différents fournisseurs montrent qu'un pourcentage important de sources posera des problèmes de correspondances. À relativement long terme, il est probable que les échanges de données issues de bases de connaissances seront facilités, étant donné que tous les fournisseurs de systèmes pour bibliothèques prennent part à l'élaboration des standards KBART concernant les données des bases de connaissances. Sur ce point, Google Scholar poursuit une autre démarche en créant sa propre base de connaissances en demandant aux éditeurs de lui procurer les données sur les fonds de leurs bibliothèques clientes et en demandant à des consortiums de bibliothèques qu'ils autorisent cette communication.
- **Portail** : Les portails utilisés par tous les outils de découvertes étudiés (y compris VuFind tel qu'il est utilisé par plusieurs outils de découverte développés en propre) correspondent en grande partie au cahier des charges défini dans la première phase de l'étude (V. Annexe C).
- **Connecteurs** : Le Sudoc et WorldCat sont intégrés dans la plupart des outils de découverte sur le web, ou le seront, et de la sorte pourront donner aux utilisateurs finals une idée des fonds papiers présents dans les bibliothèques.

## Feuille de route de l'ABES

Les résultats décrits plus haut ont conduit à la conclusion que la faisabilité de construire un nouvel outil de découverte national appliqué à des contenus savants en réalisant un index centralisé était à écarter. Toutefois, l'expérience qu'ont les utilisateurs finals français des solutions de découverte pourrait être grandement améliorée en introduisant deux services qui s'intégreraient dans des outils existants de découverte à l'échelle du web, à savoir : (1) un service de localisation national et (2) une amélioration des métadonnées par le hub de métadonnées permettant une sélection pertinente de la littérature savante. Ces deux services seraient utiles pour toutes les bibliothèques de l'enseignement supérieur en France, qu'elles aient des outils de découverte sur le web, des résolveurs de liens ou pas. Par conséquent, les éléments suivants figurent dans la feuille de route de l'ABES :

- **Le hub de métadonnées comme composante d'une solution de découverte sur le web** : Le hub de métadonnées créera une plate-forme française de métadonnées permettant une sélection de la littérature savante, comme c'est le cas des métadonnées des licences nationales et de contenu des éditeurs français. Les métadonnées de la plateforme française seront ensuite enrichies par celles du hub de métadonnées. Les métadonnées de la plate-forme seront en Libre accès et pourront être utilisées par d'autres services pour enrichir leurs propres métadonnées via des mécanismes de mise en correspondance et de fusion.

Les études de scénarios ont montré clairement que tous les services de découverte utilisaient ces mécanismes pour l'enrichissement de leurs métadonnées. Ces métadonnées amélioreront dès lors la découverte au sein de la communauté de l'enseignement supérieur via des systèmes existants.

- **Un service de localisation pour des fonctions découverte à l'échelle du web reposant sur une base de connaissances nationale** : Un résultat important de l'étude du scénario de réalisation interne a montré la possibilité de mettre en place un service de localisation suivant la logique du résolveur EZB et du webservice JOP. Cependant, développer un service de localisation national implique d'utiliser un résolveur de liens exploitant des données de bases de connaissances. Pour cela, avec Couperin et les bibliothèques françaises, l'ABES va explorer les options possibles pour se raccorder au projet GOKb dans la démarche du projet britannique Knowledgebase+ :
  - GOKb (ou *global open knowledgebase*, base de connaissances globale) vise à devenir une base de connaissances utilisant une architecture normalisée et une licence Creative Commons.
  - Knowledge Base + est un service mutualisé qu'a récemment mis au point JISC Collections, dont l'objectif est de permettre aux bibliothèques britanniques de gérer leurs ressources électroniques avec plus d'efficacité.
- **Un plan de développement complet et cohérent** : Un prototype du hub de métadonnées sera disponible à la mi-2013. Au cours de cette même année, l'ABES analysera en étroite collaboration avec Couperin et l'ADBU les possibilités de collaborer à GOKb et Knowledgebase+ pour mettre en place une collecte de données en France pour une base de connaissances d'échelle nationale. Vers la fin 2013, un prototype de localisateur national sera développé à partir des données du SUDOC concernant les fonds papier et des données GOKb pour certains fonds électroniques. Les développements ultérieurs de la collecte de données pour la base de connaissances nationale se concentreront sur l'élaboration de procédures opératoires collaboratives au niveau national avec Couperin pour ce qui concerne les licences consortiales, avec l'ABES pour les licences nationales et les bibliothèques de l'enseignement supérieur françaises pour les licences de niveau établissement. Ces procédures collaboratives permettront de s'assurer que les données des bases de connaissances n'alimenteront pas seulement le service de localisation national mais permettront de faciliter les travaux des bibliothèques universitaires visant à maintenir et mettre à jour leurs bases de connaissances pour leurs systèmes d'ERM. Par conséquent, on peut considérer la collecte mutualisée de données de base de connaissances comme une première étape vers un système d'information documentaire mutualisé « dans les nuages ». Il est prévu que la plate-forme ISTEEX avec ses contenus de licences nationales débutera sa mise en place à partir de 2015, alors qu'un premier groupe (pilote) de bibliothèques universitaires migrera vers un système documentaire mutualisé dans les nuages pendant la période 2015-2017. Entre 2016 et 2017, on peut conjecturer qu'une expérience suffisante aura été amassée avec ces éléments de l'infrastructure nationale française pour envisager une autre migration vers l'informatique en nuage.