

Summary of the ABES Discovery study

Introduction

ABES plays an important part in the discovery of national resources by its services SUDOC (the union catalogue), Calames (archival collections), Numes (digitised corpora) and Signets des Universités (relevant websites). In order to look into the possibilities to expand these national services with a webscale discovery tool for e-journals and e-books, ABES assigned a study to explore the following three scenarios:

- **A do-it-yourself scenario, building a newly developed discovery service:** This scenario is based on Trove, the online search engine developed by the National Library of Australia. In this scenario, the national discovery tool for France will be newly developed using open source software: the metadata and an index of the full text will be retrieved from each publisher. Users that are a member of the library with a subscription to the resource will be given immediate access. For others, a delivery mechanism should be provided.
- **A scenario using discovery tools available on the market:** In this scenario, a similar national resource discovery tool is envisaged based on an adaptation of existing discovery tools, such as Summon (Serial Solutions), Primo (Ex Libris), EBSCO Discovery (EBSCO) or OCLC WorldCat local.
- **A scenario that develops collaboration with Google Scholar:** In this scenario, Google Scholar plays a central role, providing the public interface and the search engine. Access to full text is provided using a link resolver mechanism.

The study was carried out by Pleiade Management and Consultancy in three phases: in phase 1 the requirements for such a national discovery tool were developed, in phase 2 existing and/or developing solutions in other countries were studied and in phase 3 the feasibility of the three above-mentioned scenarios was studied and matched with the requirements as defined in phase 1. Finally, the results were discussed in a number of workshops with the ABES staff members, leading to a roadmap for further development. This study intertwines with three other projects building elements of the French national library infrastructure: the Metadata Hub project, in which ABES extends its expertise on metadata into the internet environment, a project led by ABES with a number of libraries to set-up a tender for a shared integrated library system in the cloud and the ISTEX project that develops a national platform for national licensed content.

Main results of the scenario studies

- **(Meta)data platform:** The self-built discovery services that produced their own centralised indexes for scholarly literature only have achieved a very limited coverage. The experiences in this respect of the Australian discovery tool Trove and the German Suchkiste make it clear that building a centralised index of a *selection* of the worldwide scholarly literature is feasible. However, to build a centralised index of the *entire* world wide scholarly literature is seen as a major effort, requiring a lot of manpower and therefore not seen as feasible. In contrast, existing webscale discovery tools claim to cover the worldwide scholarly literature extensively and to index the full text of it for a large part of it. With regard to the metadata quality, all discovery

systems have mechanisms in place to use metadata from other sources such as A&I databases to enrich the metadata delivered by the primary publishers via match & merge mechanisms.

- **Locator services:** The German EZB link resolver - based on the data of the EZB union catalogue for e-journals - provides a very good example of a national link resolver. The EZB link resolver can be used by the libraries without link resolver of their own to support the discovery by their end-users. In addition, the EZB link resolver can be used by libraries with a link resolver of their own: either by using the EZB link resolver as a target for their own link resolver, or by using the knowledgebase data of the EZB for their local knowledgebase that supports their local link resolver. In addition, the German Journal Online & Print webservice (JOP) provides an (probably unique) service for end-users with regard to the print journals holdings of their library. This webservice indicates the availability of a journal article in the print holdings of the library of the end-user, based on the union catalogues ZDB and EZB. In combination with the EZB link resolver these two services support the end-users in the location of the full text in their library. The existing web scale discovery tools would need to develop such a national locator service. While interoperability between the various link resolvers and the various discovery services pose no problem, for a national link resolver an underlying a national knowledgebase will be needed. Present experiences with exchanges between knowledgebases of the various providers show that an important percentage of sources will not match. In the longer-term, it can be expected that the exchange of knowledgebase data will be facilitated as all library system providers participate in the development of the KBART standards for knowledgebase data. In this respect, Google Scholar follows a different approach by creating a knowledge base of its own by asking publishers to provide the holding data of the libraries that are their customers and by asking library consortia to permit such a delivery.
- **Portal:** The portals used by all discovery tools studied (including VuFind as used by several self-built discovery tools) meet most requirements as defined in the first phase of the study (see Appendix C).
- **Connectors:** Sudoc and WorldCat are already integrated in most webscale discovery tools and via this way can provide end-users with insight in the print collections of the libraries.

ABES roadmap

The above-mentioned results led to the conclusion that building a new national discovery tool for scholarly content was not feasible with regard to building a centralised index. However, the discovery experiences of the French end-users could be very much improved by introducing two services that would be integrated in existing webscale discovery tools: (1) a national locator service and (2) improving the metadata by the Metadata Hub for a relevant selection of the scholarly literature. These two services would serve all Higher Education libraries in France: libraries with webscale discovery tools, with link resolvers and without link resolvers. The ABES roadmap therefore consists of the following elements:

- **Metadata Hub as a component in webscale discovery:** Metadata Hub will create a French metadata platform for a selection of the scholarly literature, such as the metadata of the national licences and content from French publishers. The metadata for the French platform will then be enriched by the Metadata Hub. The metadata at the French platform will be Open Access and can be used by other services to enrich their own metadata via match & merge mechanisms. From the scenario studies it has become clear that all discovery services used these mechanisms

to enrich their metadata. The enriched metadata will thus improve the discovery by the French HE community via existing discovery systems.

- **A national locator service in webscale discovery based on a national knowledgebase:** A major outcome of the do-it-yourself scenario study was the possibility to set up a national locator service along the lines of the EZB link resolver and the JOP web service. However, the development of a national locator service implies a link resolver using national knowledgebase data. For this, together with Couperin and the French library, ABES will explore the options to connect to the GOKb project along the lines of the British Knowledgebase +:
 - GOKb (the global open knowledgebase) aims to become an open knowledgebase using standards-based architecture and with a CCO license.
 - Knowledge Base+ is a recently developed shared service from JISC Collections that aims to help UK libraries manage their e-resources more efficiently.
- **A cohesive development plan:** A prototype of the Metadata Hub will be delivered mid-2013. During 2013, ABES will explore in close cooperation with Couperin and ADBU the options to collaborate with GOKb and Knowledgebase + in order to set up a national knowledgebase data collection for France. Towards the end of 2013, a prototype for a national locator service will be developed based on SUDOC data for print holdings and on GOKb data for some digital holdings. Subsequent development of the national knowledgebase data collection will focus on developing collaborative working procedures at a national level with Couperin for the consortia licences, with ABES for the national licences and the French HE libraries for the individual licences. These collaborative working procedures will ensure that the knowledgebase data not only will feed into the national locator service but also will facilitate the efforts by the French HE libraries to maintain/update their knowledge bases as part of their ERM systems. As a result, the shared knowledgebase data collection can be seen as a first step towards a shared integrated library system in the cloud. It is foreseen that the ISTEEX platform with the national licensed content will be implemented from 2015 onwards, while the first (pilot) group of HE libraries will migrate to a shared ILS in the cloud in the period 2014-2016. In the period 2016-2017, one expects that enough experiences will be gathered with these new elements of the French national library infrastructure to consider a further migration to the cloud.