Implementing a Discovery Tool:

Options, Experiences, Expectations

Dr. Thomas Mutschler

Head of Acquisitions

Thüringer Universitäts- und Landesbibliothek Jena (Germany)

Email: thomas.mutschler@thulb.uni-jena.de

Inforum 2013, 19th Annual CONFERENCE ON PROFESSIONAL INFORMATION RESOURCES, May 21-22, 2013, Prague

Introduction

When introducing discovery tools, libraries respond to the altered user expectations as well as to the transition of the media market, the growing variety of electronic publications and platforms.[*] The provision of a single search box as a new kind of service aims at converting the library to look somewhat alike Google. This goal is achieved by making the information sources available into a joint index rather than locating each database separately. Whereas Google makes the publicly available content on the web findable, the new discovery tools provide unified indexes of academic content, including the subscriptions of the library as well as the metadata of its print collections. Prior to the discussion of the experiences the Thuringian University- and State Library of Jena (Germany) has made with the implementation of a resource discovery system on the basis of *Summon* (Serials Solutions), the first part of the article shall outline some general questions.

General questions

What is a discovery tool?

A discovery tool is a device which quickly searches through a large amount of data, regardless of being purchased, licensed or free of charge (Open Access content). Generally discovery tools consist of two components: a unified index and a discovery layer which provides different features such as

[*] I would like to thank Angela Hammer (ThULB Jena) for revising the text of this article. The image on page 5 was designed by Marianne Schiller (ThULB Jena).

relevance ranking, an intuitive interface and facets for searching. Various vendors offer a variety of resource discovery tools for subscription – including Ebsco Discovery System, Primo Central (Ex Libris) and *Summon* (Serials Solutions). The alternate way comprises open source discovery tools (e.g. *VuFind*) without providing an index with licensed content.

Why a discovery tool?

Today scientific information derives from many different sources such as databases, library catalogs and institutional repositories. Users expect to discover the world of knowledge à la Google, however they are actually facing a wide range of platforms and many search entries. Patrons and researchers are encouraged to concentrate on topical interdisciplinary research rather than wasting time on finding the proper database. Thus the classical OPAC is barely adequate in Google Age. Libraries today will have to recapture territories which they have lost to the search engines.

What is included?

The discovery search includes both licensed and purchased material (journal articles, e-books…) as well as Open Access content, for example the HathiTrust Digital Library or journal content listed in the DOAJ. Furthermore, discovery systems provide access to locally held digital collections and related metadata including digital collections from other libraries which have been given permission to be integrated into the unified index of the vendor.

How is content included?

Usually publishers provide full text content and related metadata for indexing purposes. The possibility of free access to full text content not only depends on the license between publisher and aggregator but also on the library's purchases and subscriptions. This indicates the libraries do not automatically enjoy free access to full text content but only to their subscriptions. Bibliographic records of libraries may be integrated in two different ways: integration of marc records into the index of the vendor versus ingesting them into a separate index (*Solr*), thus again made be discoverable by additional software tools (e.g. *VuFind*).

How does it work?

To create a unified index, vendors need to ask permission from each individual publisher. Once content from various sources and publishers has been harvested, the metadata is "normalized" into a common schema or type of record, for example by matching and merging. Again, the index is hosted by the vendor and has to be subscribed by the library. Usually the interface is also hosted by the vendor. Some providers however permit local interfaces.

Experiences with implementing *ThULB Search* at the Thuringian University- and State Library of Jena (ThULB)

At the Thüringer Universitäts- und Landesbibliothek Jena (ThULB) we decided to implement a resource discovery system by the end of 2011. The library looks back on a rich history. Established in the first half of the 16th century as a Saxonian electoral library, today it serves as a university library for the Friedrich-Schiller-University Jena as well as a state library for the Federal State of Thuringia. Currently the library's collections comprise more than 3.9 million print copies including rich historical collections. Also, the library provides access to a wide range of electronic resources. Among those there are about 30,000 e-journals, 1,700 databases and different e-book-collections. In addition to this the ThULB Jena displays many digitization projects on its agenda, which are realized in cooperation with other libraries, archives and museums. In 2013 the digital collections will have reached a scope of 4.5 million digital copies (scans).

Why did we choose *Summon*?

From our point of view there were several reasons for selecting *Summon* as a favourite discovery tool. First, the product can be implemented in an "out-of-the-box" style. In this implementation scenario all data is hosted by Serials Solutions which also provides the interface. Second, *Summon* proved being manageable for inhouse developments by providing an application programming interface (API) in order to customize the system. An additional argument for choosing *Summon* was the active user community in Germany (e.g. the university libraries in Konstanz, Heidelberg, Lüneburg and Hamburg are currently using *Summon*). Finally, it has been the adequate cost-benefit ratio which enabled *Summon* to proof as best choice for Jena.

The *Summon* discovery service provides all key features of discovery search: it integrates a large amount of pre harvested full text content and metadata from different sources into a unified index. *Summon* harmonizes the collected metadata by matching, merging and normalizing it into a standard format. The current scope of content made accessible via *Summon* index covers about 1,2 billion records. Catalog records, contents and metadata deriving from subscribed e-resources are integrated as well as digital collections and bibliographic records from A&I Services. Additionally, *Summon* provides a single search box including facets for topic, era, language and media type in order to encourage users to screen large clusters of results. The *Summon* service includes relevance ranking and automated query expansions. A special highlight out of the many characteristic features of *Summon* is its database recommender as to guide users to subject specific databases.

Implementation Process

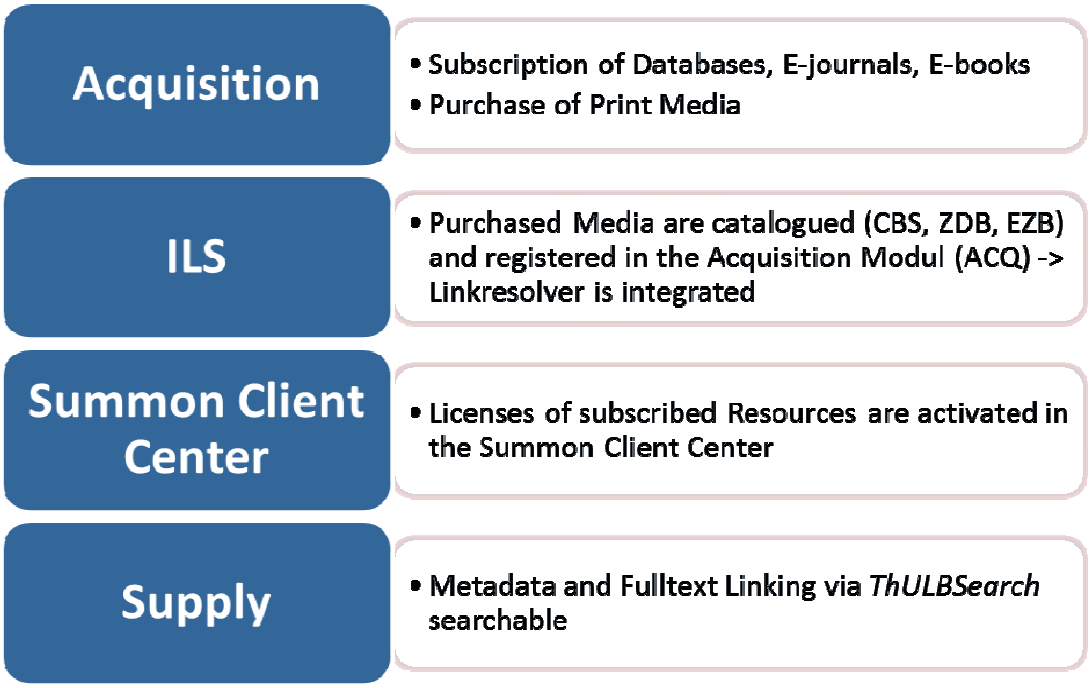
The implementation process has been initiated in winter 2011/12 by evaluating the products of various vendors with the decision for subscribing *Summon*. January 2012 marked the beginning of the start-up phase comprising the individual tasks as follows:

- Activating licensed sources via *Summon* knowledge base

- Uploading records from the union catalog into *Summon*
- Harvesting contents from institutional repositories and ingestion into index
- Integrating the *Summon* index into *VuFind* and adaptation of the interface
- Integrating link resolver (360 Link)
- Preparing concepts for staff teaching

In December we decided to launch a beta version focusing on the subscribed e-resources (e-journals, e-books, databases…).

***Summon* in the acquisition process**

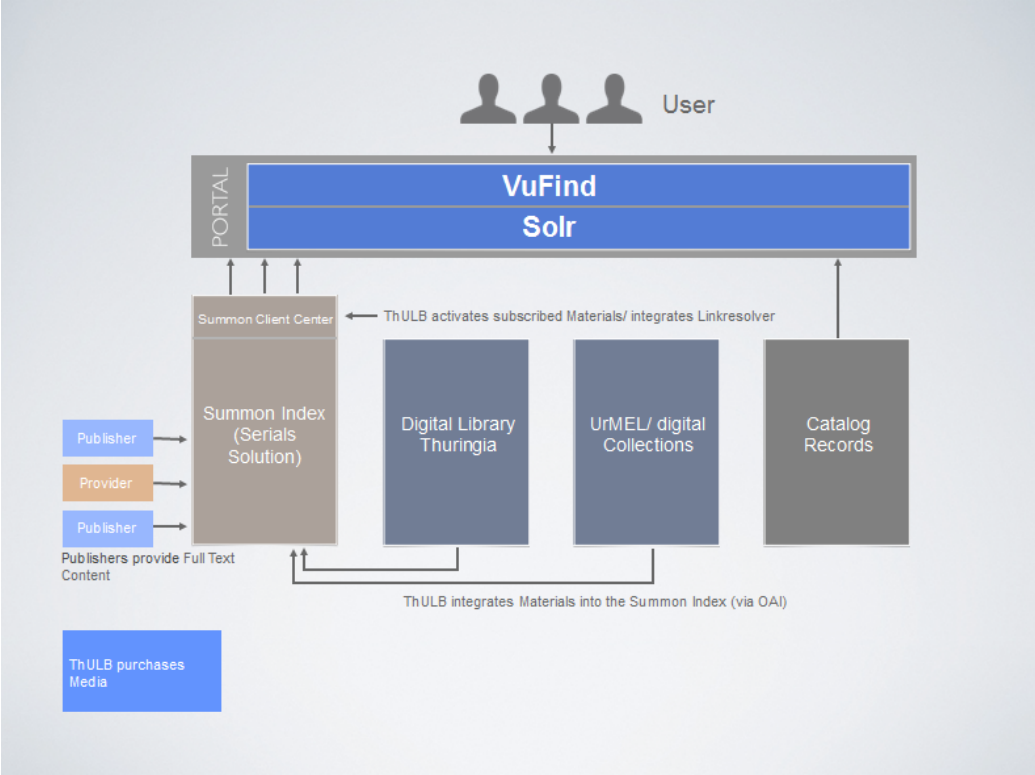


Why did we decide to integrate *VuFind* during our start-up phase?

Soon after the beginning of the start-up phase we decided to integrate the open source software *VuFind* as a search interface in combination with Solr Index which includes our catalog records. Why did we do this? First of all we could rely on the expertise of our library union, the Common Library Network (GBV), which offers support for libraries concerning *VuFind*. Second, we were able to activate existing know-how in our library deriving from building up a discovery tool especially designed for digitized historical collections (Digitales Thüringen). Additionally, certain features meet the local requirements more favourable than the vendor's interface (e.g. realtime status or direct access to the user account). One of the main reasons for choosing an additional open source software was to become more independent from the vendor.

The other side of the coin when choosing an open source software is the greater expenditure in maintenance, hardware, software and staff. Also it has to be taken into account that the start-up phase appears to be considerably longer.

System Architecture



Some issues of special concern to us

After having launched the beta version for a period of almost half a year, our users are enabled to place their search queries for local e-resources via a single search box embedded directly on the library's website. Meanwhile more than 200.000 records (about 56 million articles) correlating to the current subscription of our library are searchable by means of *ThULB Search*. The integration of our library's catalog records will be completed soon. This will enable our users to directly search via a two column search box (*articles & more* for subscribed and free full text content / *books & more* for catalog records).

However some challenges have not been resolved yet. For example: The *Summon* index, although integrating tons of scholarly content, still does not cover our subscribed e-resources to the 100% desired. An additional challenge may arise from the enormous flood of information. Especially the known item search is sometimes not satisfying. This very factor leads us to plan the two column box, where catalog records related to the rich print collections may be separately displayed and searched. Another open question comprises dealing with the classic catalog, the database list (DBIS) or the Electronic Journal Library (EZB) – all of them serving as the familiar access points for our users. Where should these features be located in future? Further aspects discussed focus the role of discovery search in the library's information literacy courses: Where does discovery search actually end? And where does the need for consulting specialized subject databases begin? And how shall we manage to finance the new service confronting tighten budgets?

Conclusion

After a period of six month following the launch of the beta version our conclusion is that we do not regret having implemented a discovery tool in our library. We are looking ahead: The acceptance of *ThULB Search* among our users is steadily increasing and our technical staff has learned much about *VuFind*. And last but not least: library staff has learned a lot from each other.

Further Readings

Judy LUTHER & Maureen C. KELLY: *The Next Generation of Discovery*, in: Library Journal, Vol. 136 (2011), Iss.: 5, pp. 66-71, URL: http://www.libraryjournal.com/lj/home/889250-264/the_next_generation_of_discovery.html.csp

Jason VAUGHAN & Tamera HANKEN: *Evaluating and implementing Web Scale Discovery Services in Your Library*, ALA TechSource Workshops, July 13 & 20, 2011, URL: http://de.scribd.com/doc/59958617/Evaluating-and-Implementing-Web-Scale-Discovery-Services-Part-1

Sharon Q. YANG & Kurt WAGNER: *Evaluating and comparing discovery tools: how close are we towards next generation catalog*, in: Library Hi Tech, Vol. 28 (2010), Iss.: 4, pp. 690-709, DOI: 10.1108/07378831011096312