

# ELAG 2010

Tuesday, 8 June 2010 - Friday, 11 June 2010

Metsätalo



## Book of Abstracts



# Contents

Meaningful Relevance Ranking for Heterogeneous Scholarly Materials . . . . .	1
Discovering the library collections - Theory meets practice . . . . .	1
Researching user behaviour on the web –how do we provide library and bibliographic services that appeal to today’s digital natives? . . . . .	2
Open Bibliographic Data . . . . .	3
Linked data as a library data platform . . . . .	4
Making a case for open-source solutions to meet users’ rapidly changing needs . . . . .	5
Users’ relevance behavior and information ecology for digital libraries . . . . .	5
TING.concept - a foundation for change . . . . .	6
Web Services Bootcamp . . . . .	7
Understanding the Needs of Library Users in a Digital Age . . . . .	8
Implementing Auto-suggest . . . . .	8
The Finnish National Union Catalogue . . . . .	9
Aggregation of metadata - Ways to do it! Who should do it? . . . . .	9
Implementing an useful discovery interface for a large scale digital collection . . . . .	10
Working with your users to develop a modern user interface for search . . . . .	11
Keynote speaker: Tales from teaching committee: engaging learners, lecturers and libraries. The presentation is kindly sponsored by Axiell . . . . .	12
Rewrapping you data –Providing a tech savvy interface to your catalogue . . . . .	12



1

## Meaningful Relevance Ranking for Heterogeneous Scholarly Materials

**Author:** Tamar Sadeh<sup>1</sup>

<sup>1</sup> *Ex Libris*

**Corresponding Author:** tamar.sadeh@exlibris.co.il

Relevance ranking of search results has become a de facto standard for information systems' display of result lists. The unparalleled success of Google has demonstrated that a relevance-ranking algorithm that is tailored to the inherent structure of the information—in this case, the Web—indeed enables a search engine to sort results in a manner that is most suitable for end users. As a result of Google's success, the behavior of users has altered, and today people around the globe focus on the first page of any result list regardless of the number of items in the list.

In the past, scholarly information providers were hesitant about relevance ranking because determining relevance, already a difficult task in the world of general Web searches, is even more daunting where academic searches are concerned. The relevance of materials depends on various factors, such as the searcher's discipline, expertise, and goal. For example, academics might seek a basic article on a topic or may look for the most recent information; they might be interested in a specific aspect of an event or a phenomenon; or they might want to avoid information that is too detailed. Nevertheless, to address users' expectations, information providers and library system vendors have been implementing relevance-ranking algorithms for scholarly systems in the last few years.

Scholarly relevance-ranking algorithms take into account the "proximity" of each item on the result list to the query entered by the user, as well as various types of information about each item. Such information may include the item's citation rate, popularity, and availability; for physical items, the number of copies that the library holds may also play a role. Much of the success of such algorithms therefore depends on whether comprehensive information is available for every item in the result list.

The emerging mega-aggregate indexes of global scholarly materials, such as Primo Central from Ex Libris, Summon from Serials Solutions, and EBSCO Discovery Service, pose an even greater challenge. Blending results that originate from a library collection—mostly physical items such as books, journals, videos, CDs, maps, and manuscripts—with results from these indexes, which cover mostly articles and e-books, involves tasks such as comparing items that have only metadata with items that have full text; determining the proportion of items to display from a relatively small collection of a library (typically not more than a few million items) versus the global set of hundreds of millions of scholarly materials; and addressing library policies about how prominently the result list should display materials originating from the library's own collections or specific collections that it considers of great value.

The talk will describe the role and challenges of relevance ranking in the scholarly environment and will highlight several aspects of successful relevance-ranking algorithms.

2

## Discovering the library collections - Theory meets practice

**Authors:** Lukas Koster<sup>1</sup>; Rosemie Callewaert<sup>2</sup>

<sup>1</sup> *Library of the University of Amsterdam*

<sup>2</sup> *BIBNET*

**Corresponding Author:** l.koster@uva.nl

## Presenters:

1. Rosemie Callewaert, Project Manager Metadata Architecture, BIBNET, Belgium, <http://www.bibnet.be>
2. Lukas Koster, Library Systems Coordinator, Library of the University of Amsterdam, The Netherlands, <http://www.uba.uva.nl>

## Format:

Duo presentation, confrontation between theory and practice

## Background:

The Library of the University of Amsterdam (Lukas' employer) has clear and explicit plans to implement new library 2.0 features: uniform discovery interface, mobile services, linked data/RDF, digital content, etc. However at the moment the UvA Library has not implemented any of this yet, mainly because recently much time has been invested in migrating to a new ILS. There are however a couple of web2.0 features in the new Aleph OPAC (<http://opc.uva.nl>). Besides that, the UvA Library offers the metasearch (MetaLib) and OpenUrl service (SFX) portal <http://digitaal.uba.uva.nl>.

Lukas writes about technological, social and cultural developments in libraries on his blog <http://commonplace.net>, but that is only theory. There has not been an opportunity yet to "practice what he preaches". This will change in the near future.

Bibnet is an organisation founded by the Flemish government for the development of the digital infrastructure and services for local public libraries, in the Flemish (Dutch speaking) part of Belgium. Rosemie is responsible for the Flemish public library union catalogue portal <http://zoeken.bibliotheek.be>, built with Aquabrowser based on the metadata of Open Vlacc, in which lots of library2.0 features have been implemented.

## Content:

The proposal is to highlight a number of specific topics. Lukas will describe these topics briefly in a theoretical manner: "challenges" and solutions, advantages, etc. with supporting slides. After each short introduction, Rosemie will present the actual implementation in <http://zoeken.bibliotheek.be>, with some technical clarifications, practical problems, examples, etc.

There will be a kind of dialogue between two people/two environments, hopefully also with interaction with the audience.

## Topics:

- \* `frbr/rda/marc/dc`
- \* metasearch vs harvesting and indexing
- \* datamashups
- \* `linked data/semantic web/rdf/foaf`
- \* user interfaces: web discovery interfaces/online services, rss/mobile access, services/social n
- \* relevance ranking
- \* search 3.0: blended and universal/vertical search
- \* metadata+classification models: professionals/social tagging/implicit automatic tagging
- \* cooperative models/virtual collections/union catalogues
- \* academic libraries vs public libraries
- \* physical vs virtual libraries
- \* digital content

6

## Researching user behaviour on the web –how do we provide library and bibliographic services that appeal to today's digital natives?

**Authors:** Joy Palmer<sup>None</sup>; Joy Palmer<sup>1</sup>; Lisa Charnock<sup>1</sup>; Rigby Janine<sup>1</sup>

<sup>1</sup> *Mimas, The University of Manchester*

**Corresponding Author:** [lisa.charnock@manchester.ac.uk](mailto:lisa.charnock@manchester.ac.uk)

This paper will summarise the findings of a number of market research studies that have been commissioned by Mimas, the UK National Data Centre based at The University of Manchester(1), in order to gain a better understanding of how students and researchers interact with and search on the Internet. The research revealed a wealth of information about the use of Google, awareness of online resources, use of new technologies, and Internet searching behaviour. The findings are valuable in informing the development of information databases and, specifically for Mimas, have provided significant insights into user expectations and experience as it looks to the next phase of development of its library and bibliographic services.

When beginning work on a number of new projects, Mimas was keen to know more about its users, and initiated the market research to inform new developments of its bibliographic services. The research consisted of a number of focus groups and interviews with a range of participants from a variety of subject disciplines. Librarians, researchers and students, were asked for their thoughts on the design and functionality of online resources, personalisation, aggregation, their use of Web 2.0, and of mobile Internet services.

Key findings of the research revealed, perhaps unsurprisingly, the popularity of Google (thus validating the findings of the CIBER Google Generation report (2)), user expectations that full text would be available online, low levels of awareness of electronic academic information services, and heavy reliance on a small and familiar range of resources. Search practices were largely habitual, and little interest was shown in personalising searches and materials (as also evidenced in the JISC Developing Personalisation in the Information Environment report (3)). Perhaps more surprising, was the lack of interest in new and emerging technologies, with few students as yet using mobile phones to access academic material online, and a dismissive attitude to the use of Web 2.0 in education. As information providers, these insights present us with our fundamental challenge: how do we provide services that appeal to today's digital natives?

This paper will discuss and demonstrate how these findings have been, and will be, translated into Mimas service development, for example in adding personalisation functionality to the Copac(4) aggregation of special and research library catalogues, and in the repurposing of the Internet Detective tutorial for use with mobile technologies. We will also consider the wider applications of our findings for the library community, and we are keen to seek the views and opinions of the participants in order to facilitate debate about the impact of user behaviour on the development of online bibliographic tools.

References:

- (1). <http://www.mimas.ac.uk>
- (2). CIBER, (2008). Information behaviour of the researcher of the future: a CIBER briefing paper. Jan 2008. [http://www.jisc.ac.uk/media/documents/programmes/reppres/gg\\_final\\_keynote\\_11012008.pdf](http://www.jisc.ac.uk/media/documents/programmes/reppres/gg_final_keynote_11012008.pdf) (Retrieved 30 October 2009)
- (3). Curtis+Cartwright, (2008). Developing Personalisation for the Information Environment (2). JISC 2008. [http://www.jisc.ac.uk/media/documents/programmes/amtransition/dpie2\\_personalisation\\_final\\_report.pdf](http://www.jisc.ac.uk/media/documents/programmes/amtransition/dpie2_personalisation_final_report.pdf) (Retrieved 30 October 2009).
- (4). <http://www.copac.ac.uk>

8

## Open Bibliographic Data

**Author:** Patrick Danowski<sup>1</sup>

<sup>1</sup> *CERN Library*

**Corresponding Author:** [patrick.danowski@cern.ch](mailto:patrick.danowski@cern.ch)

End of January CERN Library published all bibliographic book data under Public Domain Dedication License and CC0 [1]. The presentation should highlight the reasons for this step and the chosen license. Already this step even if the data was published in MARCXML, an not very broad used format in the web, created some interest in the semantic web community. The second part of the presentation should show what happened since then and how this step helps to go the road to linked open data.

[1] <http://cern.ch/bookdata>

9

## Linked data as a library data platform

**Author:** Jindřich Mynarz<sup>1</sup>

**Co-author:** Ctibor Škuta<sup>1</sup>

<sup>1</sup> *National Technical Library of Czech Republic*

**Corresponding Author:** [jindrich.mynarz@techlib.cz](mailto:jindrich.mynarz@techlib.cz)

Linked data is a publication model for publishing structured data on the web. It was specifically designed with the web in mind as its target medium, so that it is considered the web-native data model. It serves as a means for achieving the semantic web vision. We argue that it can be an important part of the library vision as well.

Linked data knowledge technologies enable straightforward data integration via globally unique identifiers and self-describing data formats. The data are provided in a standardized manner through a single, easily accessible interface. The URI identification mechanism makes the entities contained in the data addressable, which in turn increases the possibility of their reuse. URI standardizes the manner in which the entities are referred.

Our aim is to prove that libraries can serve an important role of data producers in the linked data ecosystem. We demonstrate the application of linked data as a library data platform on selected databases from the National Technical Library of Czech Republic.

We begin with the process of choosing the set of accessible and machine-readable datasets consisting of the types commonly found in the library setting, such as bibliographic and authority data. The following step lies in choosing the way of describing the datasets with domain ontologies and creating mapping from the selected datasets.

According to the transformation styles and rules, the datasets are then converted to RDF data model. In this phase we employ techniques such as entity recognition or identity resolution based on partial identifiers. Once the entities are defined, they are interconnected inside our dataset and enriched with links to external data.

In the final stage, the data created are stored in an RDF triple store and a lightweight API for accessing the data is built. We mention the design of URI patterns and further options for application access and front-end user interfaces. Once the prototype of linked data platform is created we move on to delve into the benefits it offers for library services.

Since the library's presence is increasingly shifting towards the web, libraries have to adapt to fit in this broader context. The adoption of linked data as a data platform opens up for a wide spectrum of possible applications and reuse. Linked data allows for a plethora of serializations suited for different purposes. This makes possible the distribution of library data by many other channels than just OPACs. The data are crawlable by search engines which may by implication increase the number of incoming links to the library.

Libraries, on the other hand, can translate their vast experience in managing document collections and providing information services to the new environments. In this way, the library's brand of quality and trustworthiness may be associated with the linked data they produce.

10

## **Making a case for open-source solutions to meet users' rapidly changing needs**

**Author:** Michael Pawlus<sup>1</sup>

<sup>1</sup> *University of Sheffield*

**Corresponding Author:** michael.pawlus@gmail.com

The current OPACs being released as proprietary software offer interfaces and functionality that the developers hope will fit all users. However, these systems are often released at long intervals and have been criticised for being too slow to effectively meet user needs in a rapidly changing environment. On the other hand, open-source software has the ability to respond much more quickly to user needs as much of it is often built in pieces as modules that can be compiled by the system administrator to meet the requirements of the community served. However, these systems have yet to be adopted in wide numbers because of concerns that the projects are in a state of perpetual beta that never becomes finalised.

My presentation looks at some major players that have adopted open-source software and compares their functionality with those of other systems. In particular, I will be looking in depth at the National Library of Australia which uses VuFind, an open-source OPAC as well as the New York Public Library which uses Drupal, an open-source Content Management System. These are major libraries that could afford the most expensive solutions and have opted for open-source software. I will look at how libraries of all sizes could benefit from the systems that they use.

In addition, my research collects the feelings of the library system administrators using both models, open-source and proprietary, to show the benefits and drawbacks of each. I also look at sites which have been shown to be popular with users in terms of functionality and ease of use such as Google and Amazon. I then compare different library systems against these sites to see which ones are best able to provide similar user experiences. Lastly, I look at recent discussions on mobile apps and mashups for libraries to see which library systems are most receptive to implementing new code.

The findings of my research show that whilst a case could have been made against open-source software in the past for being unreliable; it now deserves a second look. A number of open-source products have shown that they are reliable and sustainable options. The research also uncovered that a significant amount of money could be saved by utilising open-source technology. In the end, the presentation asks system administrators and other decision makers to investigate the current open-source landscape to see if a better solution exists which can respond faster to user needs whilst demanding less financial resources.

11

## **Users' relevance behavior and information ecology for digital libraries**

**Author:** Jela Steinerova<sup>1</sup>

<sup>1</sup> *Comenius University Bratislava*

**Corresponding Author:** steinerova@fphil.uniba.sk

The paper analyzes origins of the information ecology concept in the sense of balance between users and digital libraries. It is based on results of two completed research projects on information behavior and interaction of man and information environment. Ecological characteristics of digital libraries are determined as mixture of sources, systems and services managed by a social actor.

Basic questions of the paper cover the following issues: Which changes of users' behavior are significant for the electronic environment? How should libraries and designers of digital libraries respond

to the new patterns of users' relevance behavior in the electronic environment? Can we develop relevance 2.0 and ecological relevance with the use of new tools for concept mapping in digital libraries?

Results of the first project suggest that human information behavior in the electronic environment is marked by the emphasis on easy access, quick online reading, cognitive mapping, visualization and social networking (horizontal information processing by the "Google generation"). The method of a series of questionnaire surveys in academic libraries was used.

Results of the second project on information use determine perceptions of relevance, relevance in the electronic environment, sorting of relevant information, and types of relevance. Relevance was studied with the use of phenomenographic methodology. Interviews with 21 students were conducted with the aim to explain behaviors while assessing relevance.

Results indicate that new patterns of relevance emerge. Relevance in the electronic environment is re-defined as driven by contexts and non-linear interactions. The same criteria are used through different contexts. Emerging patterns of relevance in the electronic environment indicate needs of flexibility, interactivity and collaborative use. Hierarchy and associations of knowledge organization support relevance judgments. Based on these results concepts of relevance 2.0 and ecological relevance are defined.

Users' relevance behavior is framed within a new research project devoted to information ecology. Examples of concept maps created as part of a textbook on Information strategies in the electronic environment are mentioned. The maps can help clean information environment and provide ecological tools for efficient information use and for semantic interoperability.

In conclusion, the contribution of information ecology concept to practice of new systems, services, products and digital libraries is summarized.

15

## TING.concept - a foundation for change

**Author:** Mats Hernvall<sup>1</sup>

<sup>1</sup> DBC, Denmark

**Corresponding Author:** mats@hernvall.com

What is the focus of TING.concept?

The ever increasing consumption of digital media in our wireless and always connected world should be good news for the libraries. The interest in and availability of books, film, music and games has never been bigger. Social interaction on internet is booming. The technical possibilities to share content and experiences find new and ever more amazing solutions for each day.

But the libraries are stuck with: old systems on proprietary platforms, new requirements implemented slowly, data and function lock in, information silos, an increasing need to integrate incompatible systems for each new data format and supplier and finally a plethora of strange user interfaces far from common internet standards.

TING.concept is a project initiated by the two largest public libraries in Denmark –Århus Public Libraries and Copenhagen Public libraries in a consortium with DBC, a Danish supplier of national and local library infrastructure, products and services. The projects vision has been to share the results and collaborate with everybody that are interested and was formulated as: "To release all information in the libraries, make the user and staff knowledge visible, create relations between them that facilitates interaction and place it in a context where it gives meaning to the users". In short, are we trying to answer questions like:

- How do we ensure that libraries are relevant in the digital society?
- How do we take part in the conversations where the users are?
- How do we mediate huge amounts of content and knowledge in the right context?
- How do we harness and share the collective innovation power of the library community?
- How do we create win-win-win situations between non-profit organizations, commercial partners and the library users?

- How do we create cost effective solutions that scales –big and small?
- And so on.

What is TING.concept?

TING.concept is fundamentally three things:

1. Our conceptual solution to the project vision and how to ensure that libraries are relevant in the digital society. In order to offer library services in the multifaceted digital society we need to standardize on data, data structures and data access. (For more info look at - <http://gnit.dk/dokumentation/diagram-over-ting-konceptet2>) The brønd.TING. (Data dwell) is capable of harvest, store and index large amounts of data in different formats. It builds and preserves relations between objects to keep API's and web services as lightweight as possible. Middle.TING delivers API's and web services for interaction with the content through search, faceted results, auto-completion, spelling, "others that have borrowed", and so on. The transaction.TING layer handles integration with back-end Library Systems and other back-end digital content systems. User interaction on different platforms- the web, mobile phones, information screens, touch screens, photo-frames and so on is handled by presentation.TING projects. New content types and media types will be added to brønd.TING through analysis of new sources and harvesting schemas through content.TING.
2. An infrastructure platform that implements the TING.concept. The platform is based on Open Source technology with an SOA architecture approach. The back end uses de-facto standards as Apache Lucene, Solr, Fedora Commons and PostgreSQL. Front end technologies used today are Drupal for web interfaces. (More to come).
3. TING.concept Community is an open eco-system for cultural innovation, collaboration, and shared results in the digital society. The Open Source community handles everything from publication of the TING.concept software and support to it's users to strategic decisions for future development and partnerships. The TING.concept Community consists of several quite self governed projects (brønd.TING, mobile.TING, drupal.TING, content.TING, partner.TING) and a governing Community Council. The TING.concept Community will officially be launched at the end of Q1 2010.

The current state of the TING.concept is public beta for Copenhagen and Århus. There are several other libraries and projects committed to the TING.concept. Read more on the project site [www.gnit.dk](http://www.gnit.dk) (gnit are ting spelled backwards).

16

## Web Services Bootcamp

**Author:** Karen Coombs<sup>1</sup>

<sup>1</sup> OCLC

**Corresponding Author:** [librarywebchic@gmail.com](mailto:librarywebchic@gmail.com)

Today there are a variety of web services which can provide data which can be used to enhance library interfaces and services. From bibliographic data to author information to subject headings, the data available to libraries is growing exponentially. Open Library contains bibliographic data, covers, and author information. LibraryThing offers covers, ratings, awards and honors and a variety of other metadata. The New York Times offers an API for their best seller lists, movie reviews, and articles. Library system service providers too are offering more web services. Serial Solutions offers web services for their OpenURL resolver and Summon discovery platform. In addition, OCLC Web Services such as xISSN, WorldCat Search API, WorldCat Identities, WorldCat Registry, and QuestionPoint web service provide a variety of data which can be used to enhance and improve the experience of library users. Additionally, an increasing number of services which offer Linked Data are also available such as Dewey.info, LCSH (<http://id.loc.gov/authorities>), OpenLibrary, and VIAF (<http://viaf.org/>).

This workshop will provide hands on experience working with a variety of web services. Key library web services will be described and demonstrated. Participants will learn the basics of key search protocol; how to work with data in a variety of formats: XML, JSON, HTML and manipulate data from web services using both client and server side scripting languages. Attendees will learn about existing web services utilizing Linked Data and skills needed to utilize Linked Data. Then attendees will participate in a discussion of ways to connect library information with other Linked Data projects as well as how OCLC can expose bibliographic data in Linked Data format(s). Come and learn how to effectively use web services in your library applications.

17

## Understanding the Needs of Library Users in a Digital Age

**Author:** Jasmine de Gaia<sup>1</sup>

**Co-author:** Arnold Arcolio<sup>1</sup>

<sup>1</sup> OCLC

**Corresponding Author:** degaij@oclc.org

Some of the habits of the “Google generation” may be new, but many of the core behaviors of this group - of searching for information, collaborating with others and expecting rapid fulfillment –are not. In fact, they have been around for a long time, and are at the heart of what libraries offer. What IS new is a digital arena which makes the discovery of content and interaction with other people faster, easier and more accessible to a wider population in places and ways that did not exist before. The challenge before us is to recognize and understand our various users’ needs in a digital age to enhance and extend libraries.

OCLC works in collaboration with libraries to develop, assess, and improve access to information. In this session, we’ll explore recent findings about users of academic and public libraries. We’ll talk about the methods and research that produced this data, and the changes that resulted.

We’ll share findings from library users about:

- Relevance ranking
- Expectations about local, group, and global scope
- Works, editions, and FRBR
- Collaborative and social features
- Expectations about digital delivery

We’ll talk about some of the methods we’ve used:

- Creation of personas to capture and communicate knowledge about various kinds of users
- Formative testing with prototypes to inform designs
- Summative testing to confirm designs

At the end of this session, we hope to leave you with some useful information about users like these, and a discussion of what has been fruitful and what has been challenging in our efforts to gather and use this information.

18

## Implementing Auto-suggest

**Author:** Thomas Hickey<sup>1</sup>

**Co-author:** Ralph LeVan<sup>1</sup>

<sup>1</sup> OCLC Research

**Corresponding Author:** hickey@oclc.org

One aspect of meeting users' needs 'instantaneously' is immediate feedback, offering suggestions, or even returning results, while the user is typing.

We have experimented with both instant results and search suggestions in several environments. Recently we have implemented auto-suggest servers for searches of the FAST subject heading vocabulary and names in the Virtual International Authority File. As in searching, ranking is a crucial component of auto-suggest and we have explored a number of different approaches.

The auto-suggest service is derived from information collected during database indexing and is designed for the very fast response times (less than a tenth of a second) needed to support useful auto-suggest. We believe the techniques developed to be of general use and plan to make the code open source and port it into other search environments, such as Lucene.

Currently the auto-suggest responses are returned in JSON format. The contents of the response are driven by XSLT transformations of the underlying XML records, and therefore quite flexible. In addition to 'traditional' user search suggestions, we expect this approach to be extremely useful in metadata creation environments.

Thomas B Hickey & Ralph LeVan  
OCLC Research

19

## The Finnish National Union Catalogue

**Author:** Nina Hyvönen<sup>1</sup>

<sup>1</sup> *The National Library of Finland*

**Corresponding Author:** nina.hyvonen@helsinki.fi

Traditionally a union catalogue has been mainly seen as a bibliographic database, a sort of "inventory" including access information. A next generation union catalogue will also include additional information like book covers and social metadata. It will no longer be just a tool for data retrieval, but a way of participation and creativity.

When completed as planned, the Finnish national union catalogue will be freely available to everyone with an unrestricted access to information. It contributes to the new system architecture of Finnish libraries as a one of the back end systems. It can also be seen as a step towards the next generation library system. The data itself can be shown in various front end systems, e.g. National Digital Library and library's local web services.

The Finnish national union catalogue will have various advantages compared to local library databases. It makes library processes more effective, e.g. cataloguing and description. Bibliographic data can be stored in one place only, which is effective and saves a lot of cost. In the meantime local library systems can be simplified.

The national union catalogue contributes to various upcoming changes in cataloguing and description rules, e.g. FRBR, RDA, ontologies and social metadata. It offers possibilities to develop totally new web services with new partners by using web 2.0-technologies and to make the most of metadata.

20

## Aggregation of metadata - Ways to do it! Who should do it?

**Author:** Marja Haapalainen<sup>1</sup>

<sup>1</sup> *National Library of Sweden*

**Corresponding Author:** marja.haapalainen@kb.se

Which came first, the libraries' need to enhance opportunities for integrated search or the commercial vendors launching product after product, meeting this need or maybe just creating this need? Today there are quite a few central indexes to choose from; Serials Solutions came up with Summon, followed by Exlibris Primo Central, there are OCLC and EBSCO with their respective products. The libraries' need has of course its origin in the increasing amount of electronic resources they have to administer and provide to users in an easy way. The demand for a single search interface would be fulfilled by the collection of metadata and pre-processed indexes on top of this data. But the collection and aggregation of data from different sources/publishers is a challenge, both technically and legally.

The easy way is maybe to rely on the commercial vendors and buy an out-of-the-box solution. The vendors have already done the work of collecting the metadata, but then again, your library has already done the necessary work of license agreements with the information providers to get the right to access this data. Furthermore the access to the vendors' metadata indexes seems to be solely provided through API:s that do not allow for the flexibility and control over the search service the library may want to provide.

The hard way is to do it yourself. If you store and index the data by yourself it will give you more control over the search services you develop and it would also mean giving libraries more control of how metadata is produced, used and re-used. One huge part of the work is the metadata aggregation workflow and normalization process, which is quite time consuming and resource intensive. Another part of the work is the negotiation with information providers. It becomes necessary to secure not only the rights to metadata but also the delivery of sufficient and relevant metadata.

Several questions arise. Is it meaningful to aggregate metadata? Whose responsibility is it really? Should the individual library or a consortium of libraries do this? Is it possible to cooperate across nations? What are the benefits?

22

## Implementing an useful discovery interface for a large scale digital collection

**Author:** Till Kinstler<sup>1</sup>

<sup>1</sup> *Verbundzentrale des GBV*

**Corresponding Author:** kinstler@gbv.de

How to sort search results?  
 Who cares what's down the list, if the best results are on top?  
 When do users notice, we don't have anything matching their information needs?  
 Why not let users decide about the size of their result set?  
 [...]  
 And how do we find out?

We address these questions in a project called Suchkiste. We are building a discovery interface for a large and growing collection of digital scientific content licensed under the DFG Nationallizenzen funding programme[1]. One major project goal is providing good usefulness.

We drafted user experience requirements based on current literature, best practice examples and findings of similar projects (eg. Beluga[2]). Results of that process are four basic concepts well known by web users:

- one single, simple to use, but powerful search box as primary interaction tool
- "relevance ranking" of results
- tools that support browsing and filtering
- linkability: every page displayed must have a simple, persistent URL in the browser address bar

The actual user interface design was done in cooperation with consulting HCI experts of HTW Chur. We used focus groups to get qualified user feedback on the ideas we developed.

At project midterm in January 2010 we put the first public release of Suchkiste on the net. It holds about 22 million digital objects in a unified search index (consisting of metadata and full text) so far (growing as we get more data) and features the basic implementation of the user interface.

The user interface is based on a table showing a framed view of the data space in Suchkiste. By default the frame is filled with the 20 “best hits” matching a user’s query (sorted by “relevance”). There is no pagination in the table. Record details are zoomable. So far, we provide only basic tools to filter and navigate results.

Depending on results of usability tests scheduled for spring 2010 we will decide how to improve the user experience. We are thinking about additional filtering tools (instead of traditional faceted navigation), personalization functionality based on our framed views concept, individualization of result ranking and ways to link up our discovery service with the web even better.

In my presentation at ELAG I will explain how and why our user interface concept differs from traditional OPACs, talk about limitations met while implementing it (for example imposed by data quality and web technology) and elaborate on ideas we consider for implementation.

[1] About Suchkiste:

Suchkiste (<http://finden.nationallizenzen.de/>) is developed by Verbundzentrale des GBV (VZG) in a three year low budget project that started in Summer 2008. Verbundzentrale des GBV is the technical service center of Gemeinsamer Bibliotheksverbund (GBV), the largest library consortium in Germany.

Suchkiste aims to provide easy access to the Nationallizenzen collections hosted across lots of different publisher web sites. It should serve as a single user friendly starting point for exploring this rich variety of scientific content. So far (January 2010) license agreements covering more than 150 million digital resources have been sponsored. Some information in English on DFG Nationallizenzen is available at [http://www.dfg.de/en/research\\_funding/programmes/infrastructure/lis/digital\\_information/library\\_licenses/index](http://www.dfg.de/en/research_funding/programmes/infrastructure/lis/digital_information/library_licenses/index)

Participating publishers are required to provide complete metadata about purchased content by Nationallizenzen license agreements. In a cumbersome workflow we analyze, convert, enrich and store that data in the central metadata management system of our library consortium (a OCLC CBS system). Some publishers deliver fulltext content as well. We provide access to that aggregated data pool through interfaces (Z39.50, SRU) and data download for entitled users (participants of Nationallizenzen). Suchkiste is one application using that data in a unified search index.

We evaluated different technical solutions and finally decided to use open source software to build Suchkiste, because the commercial options available do not meet our requirements in terms of customizability. We use Solr as index engine and VuFind as easily adoptable user frontend framework.

[2] Beluga is a project developing a next generation library catalogue that strongly uses methods of user centered design, <http://beluga-blog.sub.uni-hamburg.de/>

26

## Working with your users to develop a modern user interface for search

**Author:** Charles Lowell<sup>1</sup>

<sup>1</sup> *Serials Solutions Summon*

We all take feedback from our users very seriously, but how do we most effectively loop it back into our production systems? The real challenge is not merely finding the most effective channels through which to solicit feedback, but how to best integrate it with the expert knowledge of your team and to reconcile it with the overall product vision before it finally merges with and becomes part of the finished product.

In this talk we will discuss the approach used by the Summon User Interface Team to tackle this

problem in terms of how we collect user feedback data, how that data is digested into discreet features, and, not least important, the implementation process through which those features must pass before they can ultimately be deployed to production.

User feedback is collected through the normal channels which include among others things user surveys, live usability sessions and usage statistics. We will not focus a great deal of time on the methods of collecting user feedback, but instead give a brief overview of which ones we find important, what data is interesting in each technique.

Once feedback has been collected, the next step is to use our knowledge of the domain and our knowledge of the users to produce a coherent feature or set of features which answer that feedback. We will detail these steps and explore the apparent (and sometimes actual) tensions between the opinions of our users, the opinions of our team experts and the overall product vision. We will also discuss why, in our experience, this process functions best when its participants represent a diverse set of opinions, especially those that lie outside the domain of library science.

Finally, we'll talk about why the agile methodologies we use are so integral to the total user experience. Far from a mechanical process for implementing a set of pre-ordained features thus derived from the steps above, we will show why the actual software development process must be an equal partner in the conversation, and how it can positively interact with the other parts of the system.

27

## **Keynote speaker: Tales from teaching committee: engaging learners, lecturers and libraries. The presentation is kindly sponsored by Axiell**

**Author:** Andrew Cox<sup>1</sup>

<sup>1</sup> *The University of Sheffield*

Lecturers and academic libraries share the joint challenge of enhancing student engagement in learning. This paper explores some of the current opportunities and dilemmas arising from such factors as social and technical change, internationalisation, changing educational theory, and employability agendas. The paper synthesises results from a number of research studies motivated by the desire to explore student engagement, including explorations of Inquiry Based Learning techniques; experiments in changing assessment methods; a study of student use of communication technologies; investigations of students' communication and social networks and use of space (Sheffield opened an award winning Information Commons building in 2007). Through an in-depth exploration of factors shaping staff/student engagement in learning, debates around the changing role of the library service are opened up.

28

## **Rewrapping you data –Providing a tech savvy interface to your catalogue**

**Author:** Markus SKöLD<sup>1</sup>

<sup>1</sup> *National Library of Sweden*

Svensk Mediedatabas (SMDB) is the national catalogue for audiovisual material in Sweden. It has gone through multiple stages of automation and digitization and as of today it makes up a complete system platform for the audiovisual department of the national library of Sweden.

Each day more than 1.5TB of audiovisual material is ingested together with corresponding metadata in a fully automated manner. More than 2.500.000 hours of material is directly viewable in the system and another 5 million hours are available via on-demand digitization.

The first version of the catalogue interface was based on technical and biographical requirements rather than actual user needs; even if it provided all the functions needed it was complex and not very intuitive to use. Nor was the support for external search services such as Google satisfactory.

Late 2008 a decision was made to split all middleware business logic into a bus architecture and keep the web frontend clean and simple. The idea was to create a web application which felt clean and modern rather than your traditional OPAC, and also to make sure the data was accessible for others to use.

Today, all data is served from the bus in a native XML format which can easily be transformed to many different formats accessible via a single API. This effectively publishes our metadata and makes it available to third party applications. By utilizing Cool URIs all catalogue entries has their own unique URI, something that has eased search engine optimization.

During the session we will explain what we have done so far and tell you our plans for the future and why we need to be where our users are by making sure our data is accessible in a wide variety of ways; e.g. semantic web, linked data.