Abstracts and presentations from the ELAG 2013 Conference *The Inside-Out Library*
May 28 – May 31 2013, Ghent, Belgium

For more information, please contact Peter van Boheemen,
Chair of the ELAG 2013 Programming Committee
(peter.vanboheemen@wur.nl)
PREFACE

Dear colleagues,

Elag 2013, wonderfully hosted by the University of Ghent, was a great success. This year’s theme, ‘The Inside-Out Library’, turned out to be a very relevant one. Libraries are struggling with their future role and lots of people presented interesting thoughts on that. In this booklet, abstracts of all presentations, boot camps and workshops are gathered. Links will give you access to presentations on Slideshare and videos of presentations on You Tube.

We will hope that this enables participants to return to some interesting moments and non participants to also learn of the things that were discussed in Ghent. Most of all, I hope that this booklet may inspire lots of people to take part in Elag 2014 and make next year’s event in Bath, U.K., just as successful.

Have fun reading.

Sincerely yours,

Peter van Boheemen
Elag chair

Cartoon by Patrick Hochstenbach
# TABLE OF CONTENTS

## KEYNOTE
- A Clean Slate ................................................................. 6

## PRESENTATIONS
- The Inside-Out Research Library: Managing Research Output Beneficially .......... 8
- Magic Mirror in My Hand, Who is the Fairest in the Land? How to Re-structure Information Services .... 9
- An Open Source Infrastructure for Preserving Large Collections of Digital Objects .................. 10
- Linked Data Enhanced Publishing for Special Collections .................................................. 11
- Facing the Music: Are Information Professionals and Researchers Dancing to Different Tunes? ...... 12
- Partners in Research. Outside the Library, Inside the Infrastructure ...................... 13
- The Library Happens Elsewhere: Using and Adapting Library Workflows as Infrastructure for Stockholm University Press .................... 14
- Making Future-Proof Library Content for the Web: Metadata-Driven Workflows and Doing Things the “Right” Way ............................................. 15
- Collecting Twitter Data with Social Feed Manager ......................................................... 16
- Reinventing Discovery: An Analysis of Big Data .............................................................. 17
- “The Good, The Bad and The Ugly!” Using APIs to Develop Reading List Software at the University of Huddersfield .................................................. 18
- Triple Bypass - Open MARC Surgery ................................................................. 19
- Who is the Authority? Authority Control in Linked Data Environment .................. 20
- MOOCs and Libraries: An Overview of the Landscape, and how Libraries can serve the “Inside Out” Classroom .......................................................... 21
- Annotated Books Online: The Public as a Contributor .................................................. 22
- Improving the Discoverability of Learning Materials with Linked Data ...................... 23
- The GLAM-wiki Toolset Project ........................................................................ 24

## WORKSHOPS
- Creating Better User Interfaces for Library Catalogs: How to Present and Interact With (Frbr-Based) Bibliographic Data? ........................................ 26
- Very Gentle Linked Data ....................................................................................... 27
- An Open Access Workflow to Extract, Transform, Map, and Publish Dynamic Metadata .................. 28
- Scoremodel.org: Audit Your Digital Preservation .................................................. 29
- Drupal in Libraries .............................................................................................. 30
- ISNI, ORCID and VIAF: Examining the Fundamentals and Application of Contributor Identifiers .... 31

## BOOTCAMPs
- CouchDB: A Database for the Web ............................................................... 33
- Customizing the DSpace Manakin XML User Interface ......................................... 34
- Catmandu: Boost Your Data Processing with Library Oriented ETL ...................... 35
- Evaluating the Value of Named Entity Recognition for Non-structured Metadata with the Help of OpenRefine ......................................................... 36
A Clean Slate

**Name of presenter(s):** Herbert Van de Sompel

**About the presenter(s):** Herbert Van de Sompel graduated in Mathematics and Computer Science at Ghent University (Belgium), and in 2000 obtained a Ph.D. in Communication Science there. For many years, he headed Library Automation at Ghent University. After leaving Ghent in 2000, he was Visiting Professor in Computer Science at Cornell University, and Director of e-Strategy and Programmes at the British Library. Currently, he is the team leader of the Prototyping Team at the Research Library of the Los Alamos National Laboratory. The Team does research regarding various aspects of scholarly communication in the digital age, including information infrastructure, interoperability, digital preservation and indicators for the assessment of the quality of units of scholarly communication. Herbert has played a major role in creating the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), the Open Archives Initiative Object Reuse & Exchange specifications (OAI-ORE), the OpenURL Framework for Context-Sensitive Services, the SFX linking server, the bX scholarly recommender service, and info URI. Currently, he works with his team on the Open Annotation, ResourceSync and Memento (time travel for the Web) projects.

**Keynote:** Discusses Research Objects and the relationship to work my team has been involved in during the past couple of years: OAI-ORE, Open Annotation, Memento.
The Inside-Out Research Library: Managing Research Output Beneficially

Name of presenter(s): Peter van der Togt & Peter van Boheemen

About the presenter(s): Peter van der Togt is working at Wageningen University and Research library in the Netherlands. He has been working as a CRIS manager and Repository manager. Nowadays he is involved in library innovation and is coordinating functional management of the library applications. He is now leading a project to improve the presentation and usage of Wageningen UR output.

Peter van Boheemen is working at the IT department of Wageningen University and Research Centre as a consultant for ‘Library and Search’. He is working on innovating library applications and Enterprise search solutions. He has been involved in library automation for over two decades.

Abstract: Traditionally libraries have given much attention to the careful registration of meta data for shelved publications that were acquired for internal use.

When research libraries got involved in creating repositories for storage and description of locally produced publications, they were satisfied with sloppy meta data according to poor schemes. Many libraries were not even interested in producing good meta data records and left the description of those publications in the hands of the researchers or their administration and concentrated on describing publications that had been acquired elsewhere and have been described by many others before.

Wageningen University and Research Library has invested a lot in building a local repository, reflecting all output described in a coherent way. To be of any use to research staff and management, the library realized that extra meta data had to be added to these descriptions. Meta data that are not used in traditional cataloguing systems. As one of the smallest universities in the Netherlands Wageningen University and Research Center is now the largest contributor to the Dutch national repository NARCIS.

In this presentation we will show what is needed to describe local output in a way to be of any use to your local community. We will explain the problems we face with library cataloguing systems not supporting needs for local output. We will show in which way we provide extra information to research staff about their publications.
Magic Mirror in My Hand, Who is the Fairest in the Land? How to Re-structure Information Services

Name of presenter(s): Beate Rusch

About the presenter(s): Beate Rusch is associate director of the Cooperative Library Network Berlin-Brandenburg (KOBV), Germany

Abstract: What kind of future information infrastructure do we need in order to deliver the most appropriate support to scientists in their digital workplace? Do the existing, scattered local structures in the library field provide the right solutions for urgent issues such as electronic media license management, support for Open Access, long-term availability of digital research data, and last but not least development and operation of smart virtual research environments? How can cataloguing be organized more efficiently? What would be a good ratio between libraries, consortia as service providers for libraries, and commercial software vendors in times of cloud computing?

In order to find the right answers to these questions, in 2011 the German Council of Science and Humanities (Wissenschaftsrat http://www.wissenschaftsrat.de/en/about.html) started a comprehensive procedure for evaluating the library consortia in Germany with respect to their suitability for the future. Although the six library consortia in Germany are very different in organizational structure, funding, size and systems architecture, they are working together in a tight web of cooperative relationships. The German interlibrary loan system for instance is regarded internationally as best practice. For historical reasons however there is only a national database for serials (http://www.zeitschriftendatenbank.de/), but not for monographs, e-books and the like.

Viewing the German library landscape as a whole, the experts acknowledge that there are substantial service and infrastructural redundancies (in the area of data for example) and recommend a structural reorganization in favor of a new innovative division of library services beyond local and federal borders (Wissenschaftsrat, „Empfehlungen zur Zukunft des bibliothekarischen Verbundsystems in Deutschland“, February 2011).

Since then a quite controversial fundamental discussion is going on. Questions are raised that are not only of interest for Germany but also internationally. How much redundancy is necessary? What is the significance of library data? In which field do we start in-house developments? What consequences follow from the Open Data idea? How centralized should data structures actually be? On the other hand, what new models of cooperation are possible? What discipline specific services are necessary? Where do generalized infrastructural services start and where do they end?

From the perspective of the smallest and youngest of the library networks, the Cooperative Library Network Berlin-Brandenburg (KOBV), the German discussion will be presented for an international audience with a wink, under the title “Magic mirror in my hand, who is the fairest in the land? How to re-structure information”.

presentation
video
An Open Source Infrastructure for Preserving Large Collections of Digital Objects

Name of presenter(s): Sven Schlarb

About the presenter(s): Sven Schlarb holds a PhD in Humanities Computer Science from the University of Cologne. Before joining the Research and Development Department of the Austrian National Library in 2008, he worked as a software developer in Cologne and Madrid and as support consultant at SAP in Madrid. His work currently focuses on digitisation and long term preservation.

Abstract: Today's libraries are curating large digital collections, indexing millions of full-text documents, and preserving Terabytes of data for future generations. This means that libraries must adopt new methods for the processing of large amounts of data. And this is exactly where the SCAPE project (www.scape-project.eu) comes into play. The SCAPE project offers an open source infrastructure, as well as a variety of tools and services for the distributed processing of large data sets with a focus on long-term preservation.

In this project context, we are here presenting an open source infrastructure for preserving large collections of digital objects created at the Austrian National Library for quality assurance tasks as part of the management of a large digital book collection. We describe the experimental cluster hardware and the software components used for creating the infrastructure. More concretely, we will show a set of best practices for the data analysis of large document image collections on the basis of Apache Hadoop. Different types of hadoop jobs (Hadoop-Streaming-API, Hadoop MapReduce, and Hive) are used as basic components, and the Taverna workflow description language and execution engine (www.taverna.org.uk) is used for orchestrating complex data processing tasks.
Linked Data Enhanced Publishing for Special Collections

Name of presenter(s): Joachim Neubert

About the presenter(s): Joachim Neubert works as a scientific software developer at the ZBW German National Library of Economics – Leibniz Information Centre for Economics. He published the “STW Thesaurus for Economics” (http://zbw.eu/stw) and the 20th Century Press Archives (http://zbw.eu/beta/p20) as Linked Open Data and developed linked data based web services for economics (http://zbw.eu/beta/econ-ws). As an “invited expert”, he took an active part in the Library Linked Data Incubator Group of the World Wide Web Consortium (W3C). Currently, while migrating and enhancing the “ZBW Labs” web site (http://zbw.eu/labs) on Drupal, he is exploring the potential of CMS based LOD publishing.

Abstract: In the ambition to publish hidden gems and make unique collections most useful to the world, Linked Open Data is a much valued approach. For institutions with limited resources, it is, however, not easy to get this done. Out-of-the-box solutions are not available, and building an application and implementing the relevant standards from scratch often is prohibitively expensive. Content management systems (CMS) are a viable solution to this dilemma, and in the case of Drupal, one of the most popular CMS nowadays, Semantic Web enrichment is provided as part of the CMS core. In a simple declarative approach, classes and properties from arbitrary vocabularies can be added to Drupal content types and fields, and are turned to Linked Data on the web pages automatically. The embedded RDFa marked-up data can be easily extracted by other applications. This makes the pages part of the emerging Web of Data, and in the same course helps discoverability with the major search engines.

The presentation will demonstrate how custom content types and fields can be added to match the requirements of a particular collection. It will show how to enhance them semantically with mixed vocabularies, including the search-engine-friendly schema.org. Exposing the data in a RESTful application programming interface or as a SPARQL endpoint are additional options provided by Drupal modules.

The presentation will also introduce modules such as Web Taxonomy, which allow linking to thesauri or authority files on the web via simple JSON-based autocomplete lookup, or SPARQL Views as a mighty tool for reaching out to other Linked Data sources. Thus the Web of Data can be made instrumental for local collections, providing context and pulling in content from general resources like Wikipedia or complementary special collections.
Facing the Music: Are Information Professionals and Researchers Dancing to Different Tunes?

Name of presenter(s): Jane Stevenson & Lukas Koster

About the presenter(s):

Jane Stevenson is a qualified archivist with over 20 years experience. She works for the Mimas National Data Centre at the University of Manchester, a centre of expertise to support the advancement of knowledge, research and teaching. Jane is responsible for the management and development of the Jisc-funded Archives Hub service, a gateway to descriptions of UK archives, representing over 200 repositories across the UK. She has taught and presented widely on archives discovery, cataloguing standards, EAD and issues around interoperability. She is an active participant in the UK Archives Discovery Network and a Trustee of the UK National Jazz Archive. She is a bit of a jazzer, and tickles those ivories in her spare time, as well as enjoying bread-making and running (not necessarily at the same time).

Lukas Koster is Library Systems Coordinator at the Library of the University of Amsterdam, responsible for digital library information systems, mostly Primo at the moment.

After a sociology degree and additional training in ICT, he worked as systems designer and developer in institutions of higher education and scientific information. Since 2003 he has been working with library search and discovery systems at the National Library of the Netherlands and the University of Amsterdam.

Lukas is currently working on a pilot project linking publications in Primo to research projects data in external research information systems and corresponding datasets. Lukas has been seen playing guitar and singing at library conferences. He likes bread eating and running, possibly at the same time.

Abstract:

A researcher walks into a library/archive*. Do they:

a) Find exactly what they want quickly and easily

b) Become increasingly frustrated by the mysteries of cataloguing and collection management and just settle for whatever they can find

c) Give up, go home, and use Google instead

* Delete as applicable

Are we, as information professionals, really designing our discovery systems to meet the expectations and search patterns of our users?

One of the key responsibilities of university libraries and archives is supporting researchers. Libraries and archives have put a great deal of time and effort into making their collections available and accessible, typically by providing search-and-find tools for different types of collection, focusing on their own content types and taking their own expertise and methods as a starting point.

Does this approach really match the needs of researchers? Researchers tend to want subject or discipline specific publications, research project information; data of all kinds, from any relevant source. They are not usually concerned with collection types. Besides that, they also need support for workflows, collaboration, publishing, data collection and management.

We will highlight how perspectives can differ and conflicts of interest may occur in the dialogue between information professionals and researchers, setting the stage for the exploration of some intermediate and long term solutions, which will be set against a backdrop of projects and user evaluations, carried out at our own institutions, that put user needs and requirements into the spotlight.
Partners in Research. 
Outside the Library, Inside the Infrastructure 

Name of presenter(s): Sally Chambers & Saskia Scheltjens 

About the presenter(s): 
Sally Chambers works for DARIAH, the Digital Research Infrastructure for the Arts and Humanities based in the Göttingen Centre for Digital Humanities, Germany. Before joining DARIAH, Sally worked for The European Library focusing on interoperability, metadata and technical project coordination. Sally has been working in libraries since the mid-1990’s primarily in digital service provision. Sally is convinced that libraries have a key role to play in digital scholarship and is dedicated to understanding this role and encouraging libraries to rise to the challenge.

Saskia Scheltjens is a faculty librarian at Ghent University, Belgium. She’s busy setting up a brand new faculty library for Arts & Philosophy where she tries to develop a place that is in tune with 21st century information needs of students and researchers. She is co-founder of the Ghent Centre for Digital Humanities and co-organises the first THATCamp in Belgium. Before Ghent, Saskia worked and lived in Amsterdam where she worked as library coordinator for the Rijksmuseum. She is interested in ‘all things in between’, like digital humanities, and doesn’t believe anymore in the dichotomy between analog and digital, even when it involves libraries.

Abstract: More than ever before, libraries are having to prove their value to guarantee their existence. However, even from within the library community there is an increasing awareness that librarians should not sustain relics of service of a bygone age, but rethink their (digital) service provision (Koster, 2013). For several decades, libraries have been rising to the challenge of ‘becoming digital’. However, increasingly scholarship is also ‘going digital’ (Borgman, 2009). This rise of ‘digital scholarship’ presents a unique opportunity for libraries to show their value and improve their relevance (Calhoun, 2012; Chambers, 2012; Scheltjens, 2012). With this context in mind, our presentation will explore the role that libraries could play in digital scholarship. We will particularly focus on the (digital) service provision for humanities research, where the library has always, and in our belief should continue, to play as essential role.

We will build on the ideas such as those contained in the recently published (February 2013) issue Journal of Library Administration - Digital Humanities in the Library: New Models for Engagement. Key themes will include open licences, open access and open knowledge; the importance of librarians becoming (equal) partners in digital (humanities) research projects (e.g. the development of digital scholarly editions, virtual research environments, linking research data and publications etc.), forging new cooperations with faculty via international research infrastructures, undertaking research data management and active participation in developing new models of scholarly publication.

To illustrate these ideas, we will use practical examples of the roles that libraries have already started to play in the digital research arena. At European level, we will look at the recently EU-funded project, Europeana Cloud: Unlocking Europe’s Research via The Cloud which brings together researchers from the humanities (via DARIAH, the Digital Research Infrastructure for the Arts and Humanities) and the social sciences (via CESSDA, the Council of European Social Sciences Data Archives) to explore how the content available in Europeana and The European Library can be further developed to better meet the needs of Europeana humanities and social sciences researchers.

At present the involvement of research libraries in research infrastructures has been low (Lossau, 2012). However, increasingly libraries are starting to play a more active role in digital research infrastructures such as DARIAH. We will use specific case studies of libraries already actively involved in DARIAH (such as the State and University Library in Göttingen) and institutions that are about to embark on this kind of cooperation (such as Ghent University, KULeuven and Antwerp University together with their respective libraries). In this way, we can move towards realising DARIAH’s vision for the ultimate ‘library-scholar collaboration’ (Romary, 2013) in order to develop a ‘general charter of good practices for libraries and (digital humanities) research projects’.
The Library Happens Elsewhere: Using and Adapting Library Workflows as Infrastructure for Stockholm University Press

Name of presenter(s): Anders Söderbäck

About the presenter(s): Anders Söderbäck works as Head of the Department of Publishing at Stockholm University Library. Apart from publishing, the departments responsibilities includes repository, metadata, digitization and discovery services. Before joining the Stockholm University, Anders worked at the National Library of Sweden as coordinator for the Libris union catalog.

Abstract: The web, fundamentally a distributed hypermedia application, immediately accessible from billions of computers worldwide, provides many challenges for even forward thinking libraries. Open Access and Patron Driven Acquisitions marks the beginning of the end of libraries as we know them. On the horizon, new models of research, education and publishing promises an end to the things we once held dear: Control, Collections, Journals, Elsevier and the Dewey Decimal System.

While the library is now found in the researchers computers and reference management systems, the former library organizations take on new responsibilities in supporting research and researchers. From 2013, the Stockholm University Library is responsible for Stockholm University Press – a new publishing house for open access digital and print on demand books. SUP builds on the library’s previous efforts in publishing, digitization and repository management, but adds more power, focus, and last but not least a clear mandate from the university.

This presentation describes the road to SUP, the prides and pitfalls of building a new university press on library workflows, and in what changes this has brought on the library. It will also present the current infrastructure (technical and organizational) used by SUP for publishing and dissemination, as well as a discussion on where SUP and publishing (and university libraries) are heading.
Making Future-Proof Library Content for the Web: Metadata-Driven Workflows and Doing Things the “Right” Way

Name of presenter(s): Rurik Thomas Greenall

About the presenter(s): Rurik Thomas Greenall works at NTNU university library with semantic technologies.

Abstract: Here, we present an up-to-date model for providing library content that ensures longevity and accessibility through application of standards and best-practices without breaking your back or bank.

The role of libraries as a content provider has changed over time; while the rise of the Web as the major delivery platform has changed the way we acquire and provide content, our working has typically been limited by workflows and thinking from the traditional paper-based world. We argue that traditional systems — designed for content provision via physical media — aren’t players on the Web, backed up by the idea that current criticisms of library “silos” are simple criticism of non-Web-friendly approaches.

We believe that a transition to a web-based world means that several important aspects of the library and the library workflows must be addressed in order to provide means of finding and acquiring content that will work now and in the future. The core of this is learning from current trends in Web programming, including, but not limited to Semantic Web and HTML5, and also techniques derived from print and graphical industry.

Here, we also present a simple, workable implementation from our institution, which shows how these techniques can be integrated into all layers of the process. Including, but not limited to: data-driven acquisition, metadata-driven processing, content delivery via content negotiation with distributed RDF database architectures to provide data and content just in time in a way that everyone can use (RDFa, Schema.org, RDF, HTML5).
Collecting Twitter Data with Social Feed Manager

**Name of presenter(s):** Daniel Chudnov

**About the presenter(s):** Daniel Chudnov is Director, Scholarly Technology, at the George Washington University Libraries in Washington, DC, and in this role has management responsibilities in all aspects of technology, digitization, software, and systems development and operations at GW Libraries. Prior to working at GW, he was a software developer and project manager at the US Library of Congress, including as technical lead and project manager in the first years of the Twitter Archive project there. Earlier in his career he worked in multiple jobs at the Yale University School of Medicine and also at MIT Libraries.

**Abstract:** Social media data, like that produced in Twitter, is a rich source for studying how we communicate and exchange ideas on a global scale. Because most social science researchers and students are not trained in software development and working with APIs, however, collecting data can be a painstaking manual process for them. Social Media Manager, a prototype application developed at GW Libraries, automates the collection of data from Twitter’s API and makes it possible for scholars, students, and cultural heritage institution staff to identify, select, collect, and preserve Twitter data for research purposes at little cost. We will review the features of the application, how it can complement third-party services reselling access to historical data, and how we are using it at GW Libraries.
Reinventing Discovery: An Analysis of Big Data

Name of presenter(s): Andrew Nagy

About the presenter(s): Andrew Nagy is a Senior Product Manager, Discovery Services at Serials Solutions. Since joining the company in 2008, he has played a key role with the development team that build and launched the Summon service. In his current role, he oversees the User Experience and Search component teams for Summon as well as its API. With more than a decade of experience in web-based software engineering and product management, Andrew has managed national grant-funded software projects, developed globally adopted open source software, prototyped applications for dot-com start-ups and played a key role in the launch of the Summon service, which was awarded best-in-class from the Software and Information Industry Association. Prior to joining Serials Solutions, Andrew was a Technology Development Specialist at Villanova University, where he developed VuFind, an open source search engine for the library, a Digital Asset Management system and a multi-lingual bibliographic database. Andrew’s education includes a B.S., Information Management and Technology from Syracuse University, M.S., Computer Science and Master of Technology Management from Villanova University.

Abstract: As the late Steve Jobs would say, don’t listen to your customers, they don’t know what they want. Usability studies and live user analysis provide valuable feedback about your product or web site in terms of how the tool is used, but listening to the users about what they want out of the tool can result in a “whack-a-mole” scenario where you solve a problem for one user, but create new problems for other users. Analyzing usage data can provide a very different perspective on how live users actually use the tool and allow you to identify different personas and use cases. This talk will share how Serials Solutions collects and analyzes a dataset of queries and clicks generated by millions of users at hundreds of libraries around the world to find behaviors, patterns, successes and failures in the interface design and search algorithms and then how we leverage that to improve and redesign. We will share the details of our custom developed data warehouse system and how we leverage these tools to perform our analysis. We will also share with you before-and-afters that were developed based on the results of the ongoing analysis.
“The Good, The Bad and The Ugly!”
Using APIs to Develop Reading List Software at the University of Huddersfield

Name of presenter(s): Dave Pattern

About the presenter(s): Dave has been the Library Systems Manager at the University of Huddersfield (UK) since 2004 and has been a web developer since 1995. He was one of the Library Journal’s “Movers and Shakers” in 2009 and was IWR’s “Information Professional of the Year” in 2010. He is also proud to be a shambrarian!

Abstract: This presentation will provide an overview of the development of “MyReading”, an award-winning bespoke reading list solution for the University of Huddersfield in the UK. By using a varied selection of APIs -- including Summon and 360 Link from Serials Solutions, SirsiDynix HIP, OCLC’s xISBN, COPAC and Amazon -- the total development time was greatly reduced and the library has successfully delivered a service that has been widely embraced by academic staff. At present, over 99% of all the active course modules have a reading list available to students. Some of the APIs were good and some were ugly, but is there such a thing as a bad API?
Triple Bypass - Open MARC Surgery

Name of presenter(s): Martin Malmsten

About the presenter(s): Head of Software Development, Libris

Abstract: As more and more datasets become part of the semantic web, originating both within and outside the library community, actually consuming and using Linked Data becomes increasingly relevant. As a part of building a new version of Libris, the Swedish Union Catalogue, Linked Data technologies are now an integral part of the core of the system. This has brought about a radical change in how we do systems design making the distinction between internal and external datasets blurry at best. It also raises a couple of questions: how do we deal with changes in datasets outside our control, how do we tell people that data they relate to has changed? And ultimately: what about MARC? As a part of a larger library infrastructure dealing with MARC-records is not optional, but it does not have to be at the centre of the system. And it certainly does not need to have a large impact on the design.

Ultimately, building a system that is “linked by default” provides something that the older systems could not: context. Context in turn is what needed if we want to bring the library to the user. Being able to draw from data from other cultural institutions, Wikipedia, etc. makes it easier, or rather possible, to create compelling interfaces for our users. This, is the driving force for Linked Data when we move beyond simply publishing it.

This presentation details the systems design decisions made while building a union catalogue and cataloguing interface on Linked Data rather than traditional MARC-records as well as the benefits and examples of the same.

(A beta version will be released at the beginning of April, which would be part of the presentation)

Note: The main point is that when we move from publishing Linked Data to actually using it in our systems, things becomes really interesting. And also we killed of MARC in the process. The title is perhaps a bit too heavy on the semweb/triple puns, alternative: “Building a Linked Library Catalogue from the Ground Up”.

presentation
video
Who is the Authority?
Authority Control in Linked Data Environment

Name of presenter(s): Jing Wang

About the presenter(s): System Integration Engineer at the Systems Department, Sheridan Libraries, Johns Hopkins University. Responsible for Library Management Systems integration and serve as technical lead for VIVO implementation at Johns Hopkins University

Abstract: This presentation discusses the challenges of managing institutional scholarship information and opportunities for metadata specialist based on the experience of VIVO implementation at Johns Hopkins University. VIVO is an open source, semantic-web tool for research and scholarship discovery. Emerging tools enable participants to visualize research activity, discover linkages between people, programs, funding, scholarly works, events and more.

The presentation will start with a short introduction to VIVO, and present several challenges of identifying authoritative data source, establishing institutional data model, breaking the silos of various internal data systems, and identity management. The goal of this presentation is to start the dialog on the role of metadata specialist and data librarian in facilitating data interoperability, exchange, discovery, and usability.
MOOCs and Libraries: An Overview of the Landscape, and how Libraries can serve the “Inside Out” Classroom

Name of presenter(s): Merrilee Proffitt

About the presenter(s): Merrilee Proffitt is a Senior Program Officer in OCLC Research. She provides project management skills and expert support to institutions represented within the OCLC Research Library Partnership. Her current projects and interests include: archival description, increasing access to special collections, the impact of copyright on primary source material, digital library initiatives, looking at developing better relationships between Wikipedia and cultural heritage institutions, and how Massively Open Online Courseware (MOOCs) may impact libraries.

Abstract: MOOCs, or Massive Open Online Courses, have become all the rage, with numerous institutions joining forces with both commercial and non profit partners: Udacity, Coursera, edX, FutureLearn, and others. The US-based Babson Survey Research Group recently found that although 55 percent of institutions said they were undecided about their plans for offering MOOCs, 9.4 percent said they were in the planning stages of offering one, and 2.6 percent have already taken the plunge; the same survey showed the number of students taking at least one course online has reached an all-time high of 32 percent. With several European institutions signed on as Coursera partners and the launch of the Open University-backed FutureLearn in 2012, and the University of Amsterdam announcing plans to offer their own MOOC in early 2013, MOOCs will soon spread across and throughout Europe.

This presentation will address how libraries are engaged in MOOC efforts on campus, and how libraries are rethinking services with the prospect of an “inside out” classroom. We’ll report on how libraries are supporting early MOOC implementations by engaging in discussions around copyright, licensing, and open access; how libraries are supporting course production; how librarians are becoming “embedded” in MOOC environments in order to provide evolved research services; and finally, what we know about the “massive” audiences for these online courses.
Annotated Books Online: The Public as a Contributor

Name of presenter(s): Dafne Jansen, Bert Massop, Tom Tervoort & prof. dr. Arnoud Visser

About the presenter(s):
Dafne Jansen is a project manager in the Innovation and Development division of Utrecht University Library. She has worked on several projects that focus on scholarly communication and online collaboration.

Bert Massop and Tom Tervoort are students Computing Science at the Utrecht University Graduate School of Natural Sciences. They are the active developers of the Annotated Books Online application and derivative projects in cooperation with the Utrecht University Library.

Prof. dr. Arnoud Visser is professor extraordinarius of Textual Culture in the Renaissance Low Countries at Utrecht University. He is the Utrecht representative and coordinator of the Annotated Books Online project.

Abstract: Over the past decade, libraries have been working hard to make their special collections digitally available to both researchers and the general public. Libraries strive to provide high-quality information from reputable sources, for example by making objects such as manuscripts, early and rare printed books (or even just the texts thereof) available online. The public is treated solely as a consumer, not a contributor. This one-directional approach leaves no room for one of the most fascinating aspects of the digital age: on-line participation and collaboration.

We present Annotated Books Online, an on-line interactive archive of early modern annotated books. In the early modern period, readers were used to writing their thoughts and remarks on the margins of their books; an activity that is now frowned upon by most librarians. Such annotations can however tell us a lot about the mindset of the time. Annotated Books Online enables researchers to collaborate on transcribing and interpreting such annotations. Instead of a traditional textual per-book or per-page approach, Annotated Books Online provides graphical tools to attach relevant information to any specified area on a scanned page. Researchers can upload their own digitized books and collaborate on material provided by others, thereby eliminating the traditional one-directional approach. Results are immediately made available to the general public, bringing the spirit of Wikipedia to the world of early modern books.

While Annotated Books Online is intentionally limited to early modern annotated books, its open-source nature allows it to be adapted for any other domain of knowledge where position-dependent information on digital objects is involved. We discuss several questions that projects like Annotated Books Online raise for us librarians, such as:

Instead of participating solely as content providers, do we envision a bigger role for ourselves as information brokers?

What do we want to do with the valuable information that the public (in this case even high-quality information by specialized researchers) is generating when transcribing and interpreting these digital objects?

Annotated Books Online enables its users to upload objects from libraries from around the globe, creating their own collection based on content instead of physical location. Should we reconsider the way we treat our digitized collection as an extension of our physical collection?
Improving the Discoverability of Learning Materials with Linked Data

**Name of presenter(s):** Sarah Brown

**About the presenter(s):** STELLAR Project Manager, Library Services, The Open University

**Abstract:**

The STELLAR (Semantic Technologies Enhancing the Lifecycle of Learning Resources) project, led by the Open University (OU) and funded by JISC*, seeks to explore the extent to which the use of semantic technologies in a digital library environment can add value, as perceived and defined by academic users, to the sustainability of legacy learning resources.

As one of the earliest digital learning providers the OU has a rich heritage of archived learning materials. This unique collection also provides a comprehensive history of distance learning. Whilst some of these materials have been converted into Open Educational Resources (OERs) the wider archive of learning materials has been largely unexploited to date.

This presentation will discuss how this experimental project approached the transformation of old course materials through the application of linked data and semantic indexing, in order to make them more discoverable, flexible and reusable. The content selected included print, audio and video materials from non-current courses which were digitised and stored in the Fedora-based OU Digital Library (OUDL).

The project uses linked data to recommend legacy content to academic users who are writing new course materials. The presentation demonstrates how semantic analysis allows the retrieval of material which matches the meaning of their work in progress. Rather than spending time searching and often not finding content to reuse or reference, academics will be delivered relevant assets and data; perhaps from sources they would not previously have considered. In addition, new views of course materials can be provided; for example providing access to the “history” of a collection of learning materials or enabling comparison of how learning design evolves over time.

The presentation will describe how the project went about understanding the perceived value of legacy learning materials and determining if this value can be enhanced through the use of semantic technologies. It will describe the approach taken and whether improvements were successfully delivered by transforming the materials. Finally it will conclude by offering some thoughts on the sustainability implications of using semantic technologies in this context and how the project findings might influence digital preservation policies within the university.

*A registered charity which works on behalf of UK higher education, further education and skills to champion the use of digital technologies ([http://www.jisc.ac.uk/](http://www.jisc.ac.uk/))
The GLAM-wiki Toolset Project

Name of presenter(s): Valentine Charles

About the presenter(s): Valentine Charles is an Interoperability Specialist for the Europeana Foundation with special responsibility for the Europeana Data Model and activities related to data interoperability.

Abstract: Libraries are currently investing great effort in the digitisation of their resources. However the access to digitised content is still largely restricted to the main catalogues’ interfaces or some thematic portals. It is even more difficult to access it freely for re-use and sharing. On the other hand some initiatives have been built around the concept of free access and openness of digital content. Wikimedia Commons[1], in particular is undertaking this effort since a few years already and has been building a vast repository of media that is freely accessible. So far the contribution from the GLAM community (Galleries, Libraries, Archives and Museums) and especially the libraries has been low[2] . The lack of user friendly tools for (mass) uploads and the lack of support in the mapping from libraries metadata standards to the Wikimedia Commons templates has limited the share of library collections in Wikimedia Commons.

The Wikipedia movement and Europeana have both a strategic aim to open access to knowledge and in this particular context, to Cultural Heritage. Europeana (www.europeana.eu) provides access to digital content from libraries, museums, archives and audio-visual collections across Europe. Europeana also has experience in setting standards, has access to a network of GLAM partners and is therefore well positioned to facilitate such an infrastructure in close cooperation with Wikimedia[3].

The GLAM-wiki toolset project is developing a scalable and maintainable system for mass uploading (open) content from GLAMs to Wikimedia Commons. One of the aspects of the project is also focusing on the support of the main metadata standards used by libraries and other GLAMs and their mappings to the Wikimedia Commons templates. With the toolset libraries and other GLAMs will be able to themselves map their metadata and perform batch uploads to Wikimedia Commons. The project is also creating GLAM-specific requirements for usage statistics with the aim to ensure that the in-development Wikipedia statistics solutions will fit the need also of the Wikimedia movements GLAM-partners.

This presentation will address the different challenges libraries have to face when contributing to Wikimedia Commons. It will also detail the main requirements and principles on which the GLAM-wiki toolset is based. At the time of writing the development of the GLAM-wiki toolset and the first survey on usage statistics have started. This presentation will provide a first demonstration of the GLAM-wiki toolset and will present the results of the survey on usage statistic.
WORKSHOPS
GHENT, MAY 28 – MAY 31
Creating Better User Interfaces for Library Catalogs: How to Present and Interact With (Frbr-Based) Bibliographic Data?

Name(s): Tanja Merčun & Maja Žumer

Description:
The majority of library catalogues today still fails to truly fulfil the second objective of the library catalogue: to collocate all works of an author, all editions of a work as well as all works related to the author or a given work. Creating displays that would enable this requires changes in our bibliographic data, which has been the main incentive behind the FRBR conceptual model.

With the first implementations of RDA (Resource Description and Access) and other FRBR-based cataloguing rules, we are beginning to see a wider adoption of FRBR principles in library catalogues. While the process is just beginning and there are still a number of unresolved issues connected to the frbrization of legacy data, a crucial question that the library community should start addressing at this point is also the presentation of FRBR-based data in user interfaces of bibliographic information systems. Current list-based displays are not really best suited for displaying the complex networks of relationships in the bibliographic universe that FRBR is trying to expose, which is why we need to rethink how to make the best use of bibliographic data. Knowing which entities, attributes and relationships we wish to present to our users and what kind of interaction and exploration we want to encourage is not only necessary for exploiting the full potential of the data in our user interfaces, but also for informing our future cataloguing practice, rules and formats.

Workshop outcomes:
The workshop will first shortly explain the FRBR model, its potentials for improving user interfaces of library catalogues and the challenges ahead. It will also present some FRBR-inspired user interfaces to date as well as the author’s own experiment using information visualization approach for presenting and exploring bibliographic families. However, the main part of the workshop will be devoted to the discussion between participants, their sharing of experience and views on the topic. Although deriving from the FRBR, the workshop could be valuable to a broader audience interested in creating better bibliographic information systems as well as those looking for ideas on designing more exploratory user interfaces.

Workshop outcomes:
• creating awareness of an important aspect that has not really been addressed yet, but needs to be better explored and defined in the near future;
• exploring alternative ways for presenting and interacting with bibliographic data-
• identifying entities, relationships, attributes and features needed to support the creation of a more powerful display and exploration of bibliographic data
Very Gentle Linked Data

Name(s): Adrian Stevenson

Description: The aim of this workshop is to give non-to-moderately-technical people a chance to create some linked data themselves.

We will first take a look at some simple data modelling. Participants will create some RDF (Resource Description Framework) on paper.

We will then use a text editor to convert the RDF into electronic form.

Time and technology permitting, those feeling brave enough will load their RDF into a Fuseki triple store running on their laptop. (http://jena.apache.org/documentation/serving_data/index.html). I will show how to get the triple store running. This step is optional.

We will then learn a little about SPARQL, the query language for linked data. We will perform some simple queries using our own data where possible, or on a publicly available triple store (internet connection permitting).
An Open Access Workflow to Extract, Transform, Map, and Publish Dynamic Metadata

Name(s): Miel Vander Sande, Pieter Colpaert & Erik Mannens (Inspired and assisted by: Patrick Hochstenbach & Dries Moreels)

Description: LibreCat is an open collaboration to provide freely available tools for library and research services. It allows a librarian to define a “menu” which can be repeated for dataset extraction, transformation and loading. The DataTank is an Open Source data adapter for publishing Open Data sets. The DataTank is a RESTful data publishing tool. By daisychaining LibreCat and The DataTank’s Input project, we can now also map these data to an ontology and publish the data in a RESTful interface.

The workshop will go deeper into the latter: an ontology will be chosen, a mapping file will be created and a recipe will be scheduled. The data ingested in the triple store (a data base for semantically enriched data), will then be published through a RESTful interface.
Scoremodel.org: Audit Your Digital Preservation

Name(s): Robert Gillesse & Henk Vanstappen

Description: Early 2013, DEN and PACKED launched a Scoremodel for the preservation of digital collections (scoremodel.org). The model is intended as a self-evaluation tool: organizations that host and/or manage digital collections can use the Scoremodel to identify the potential threats to the long-term viability. In contrast to the OAIS-oriented tools such as ISO 16363, the Scoremodel is aimed towards smaller institutions, who have limited technical and organizational competencies. The criteria in this checklist are grouped around 5 domains, in each of which a limited number of criteria are given, using as little technical jargon as possible.

The resulting report can be used as a planning tool to systematically tackle the threats to digital preservation. By introducing risk levels, organizations are triggered to prioritize actions and to focus on measurements that are most important to take.

In this session, the why and how of the scoremodel.org will be explained, after which the first findings in the use of the tool will be discussed. Next, attendants will be invited to discuss major issues in digital preservation by confronting the criteria from the Scoremodel with their own experience.
Drupal in Libraries

Name(s): Péter Király & Paul Poulain

Description: The first session is about introductions: we would like to see 3-4 15 minutes presentations (focusing on available library-centric Drupal modules), and some 5 minutes “lighting talks” (focusing implementations – show cases of library related Drupal sites). The second meeting is about discussing the possibilities of future cooperation between developers and implementors. What are those task our modules can work together? What are the “terra incognitas” which are not covered by modules?
ISNI, ORCID and VIAF: Examining the Fundamentals and Application of Contributor Identifiers

Name(s): Anila Angjeli & Thomas Hickey

Description: We will discuss the fundamental principles of identifiers, such as trust, interoperability and persistence in terms of:

- Referent and Range (what the identifier attempts to identify and the types of things it covers)
- Community (who is the audience, what are their processes and priorities, local vs. regional vs. global)
- Control (who creates and modifies the identifiers and associated information)

We will then examine each of these identifier systems in some detail, going into the motivation for their creation, their characteristics and the rather complicated relationships that are developing between them. One aspect of their creation is the possibility to turn bibliographic data ‘inside-out’, creating identifiers by using it to match, create and enhance entities found in it.
CouchDB: A Database for the Web

Name(s): Sven-S. Porst

Description: This bootcamp is an introduction to CouchDB and the technologies (HTTP, JSON, Map/Reduce) related to it. Participants will then build their own database and use its data into a web page. CouchDB is a schemaless »NoSQL« database, which can store arbitrary fielded documents – formatted as JSON objects familiar to JavaScript programmers. It does not require a fixed schema definition, thus allowing for a lot of flexibility in settings with a huge number of varying fields like bibliographic records.

Designed as a database for and on the web, CouchDB is controlled and queried using the familiar HTTP protocol. This makes it natural and easy to integrate CouchDB in web pages. Indexes are defined by writing simple JavaScript functions which are then used in a Map/Reduce algorithm.

Depending on the interests of the participants and the time available, we can also look into other interesting features of CouchDB such as its changes feed for tracking modifications, its ability to replicate between instances or using Lucene to create a full text index of the documents in CouchDB.
Customizing the DSpace Manakin XML User Interface

Name(s): Bram Luyten & @mire colleagues

Description: The DSpace XML user interface is based on Apache's Cocoon framework, a framework built around the concepts of separation of concerns and component based development. During the bootcamp, @mire developers will provide a primer on customizing the DSpace interface, with concrete examples to allow hands-on coding. The envisioned outcome of this bootcamp is that participants get a good sense on “what goes where”: while customizations can be implemented in different ways, @mire's developers will go deeper into the pro's and cons of implementing on different layers.
Catmandu: Boost Your Data Processing with Library Oriented ETL

Name(s): Nicolas Steenlant

Description: To create any data oriented application, one of your recurring tasks will be to import data from various sources, map the fields to a common data model and put it all into a database or search engine. Stores such as MongoDB or ElasticSearch provide a developer friendly API, but you keep writing a lot of boilerplate or throwaway code. We tried to abstract this problem into a set of Perl tools called Catmandu which can work with library data such as MARC, Dublin Core, EndNote, protocols such as OAI-PMH, SRU and repositories such as DSpace and Fedora.

In data warehouses these processes are called ETL, Extract, Transform, Load. Many (often heavyweight) tools exist for ETL processing but none address typical library data models and services.

In this bootcamp we will provide an introduction into these tools. We will show how easy it is to import data and transform it with the help of a small DSL language. Storing and indexing become one-liners.
Evaluating the Value of Named Entity Recognition for Non-structured Metadata with the Help of OpenRefine

Name(s): Seth van Hooland, Ruben Verborgh & Simon Henghen

Description: Certain metadata (notoriously description) contain unstructured text, yet they usually convey a high amount of interesting information. To capture this in a machine-processable format, Named Entity Recognition (NER) can be used. Thanks to the brand new OpenRefine NER extension the Free Your Metadata (freeyourmetadata.org) team developed, you can enrich your description fields right from your workspace. During the boot camp, participants will be asked to perform the NER upon a sample of their own metadata. During the second half of the boot camp, we will discuss the quality of the outcomes of the different services (DBPedia Spotlight, Zemanta and Alchemy) and how the obtained URIs can be used for disambiguation and vocabulary reconciliation.
We wish to thank the Ghent University Library, the Ghent University Faculty Library Arts & Philosophy and our partners.

Elag 2013 would not have been possible without the help of the local organizing committee. Thanks to Veerle, Saskia, An, Pieter, Gwen, Pat and to the conference team Omar, Geert, Femke and Lander. Conference pictures by Geert can be found on www.flickr.com/search/?q=elag%202013.