Beyond data management plans, creative data services in libraries — Nov. 13, 2014

Joel Herndon
Head, Data and Visualization Services Duke University

Kimberly Silk
Data Librarian, Martin Prosperity Institute at Rotman School of Management University of Toronto

Mike Mertens
Deputy Director and Data Services Manager, Research Libraries UK
Joel Herndon
Head, Data and Visualization Services
Duke University
@duke_data

Creative Data Services: Mapping and Visualization
Staffing

Staff Expertise

DATA SOURCES
Locate and license data sources

DATA STORAGE AND MANAGEMENT
Learn reliable data management practices

DATA CLEANING AND ANALYSIS
Access data wrangling and statistical software

MAPPING AND GIS
Conduct spatial analyses and create maps

DATA VISUALIZATION
Explore and communicate research results

http://library.duke.edu/data/about

Webinar
Library Connect

DUKE UNIVERSITY
LIBRARIES

#LCwebinar
Joel Herndon
Partnerships for Data Services

Service: Statistical Consulting

Service: GIS Licensing/Programming

Service: Data Visualization
Environment for Data Visualizations

Source: Duke University Visualization Flickr Stream (Various Authors) - http://bit.ly/1qAinql
Environment for Visualization - Tools

From the desktop... ... to the cloud!

ArcGIS
Google Earth
Google Fusion Tables
ArcGIS Online
R
R Studio
Tableau
Shiny by RStudio
STATA
Excel

DUKE UNIVERSITY LIBRARIES

Webinar
Library Connect

#LCwebinar

Joel Herndon
<table>
<thead>
<tr>
<th>Details</th>
<th>Date</th>
<th>Time</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Details</td>
<td>Mon, Sep 8, 2014</td>
<td>10:00 AM</td>
<td>Fusion Tables: simple mapping and charting</td>
</tr>
<tr>
<td>Details</td>
<td>Tue, Sep 9, 2014</td>
<td>1:00 PM</td>
<td>Stata for Research</td>
</tr>
<tr>
<td>Details</td>
<td>Mon, Sep 15, 2014</td>
<td>9:30 AM</td>
<td>Basic Data Cleaning and Analysis for Data Tables</td>
</tr>
<tr>
<td>Details</td>
<td>Tue, Sep 16, 2014</td>
<td>1:00 PM</td>
<td>Analysis with R</td>
</tr>
<tr>
<td>Details</td>
<td>Wed, Sep 17, 2014</td>
<td>3:00 PM</td>
<td>Data Wrangling with Google Refine</td>
</tr>
<tr>
<td>Details</td>
<td>Mon, Sep 22, 2014</td>
<td>10:00 AM</td>
<td>Easy Interactive Charts and Maps with Tableau</td>
</tr>
<tr>
<td>Details</td>
<td>Tue, Sep 23, 2014</td>
<td>1:00 PM</td>
<td>Fusion Tables: simple mapping and charting</td>
</tr>
<tr>
<td>Details</td>
<td>Thu, Sep 25, 2014</td>
<td>1:00 PM</td>
<td>Stata for Research</td>
</tr>
<tr>
<td>Details</td>
<td>Mon, Sep 29, 2014</td>
<td>2:00 PM</td>
<td>Using Gephi for Network Analysis and Visualization</td>
</tr>
<tr>
<td>Details</td>
<td>Tue, Sep 30, 2014</td>
<td>10:00 AM</td>
<td>Basic Data Cleaning and Analysis for Data Tables</td>
</tr>
<tr>
<td>Details</td>
<td>Tue, Sep 30, 2014</td>
<td>1:00 PM</td>
<td>Introduction to ArcGIS</td>
</tr>
<tr>
<td>Details</td>
<td>Wed, Oct 1, 2014</td>
<td>1:00 PM</td>
<td>Analysis with R</td>
</tr>
<tr>
<td>Details</td>
<td>Thu, Oct 2, 2014</td>
<td>1:00 PM</td>
<td>Data Wrangling with Google Refine</td>
</tr>
<tr>
<td>Details</td>
<td>Tue, Oct 7, 2014</td>
<td>1:00 PM</td>
<td>Easy Interactive Charts and Maps with Tableau</td>
</tr>
<tr>
<td>Details</td>
<td>Fri, Oct 10, 2014</td>
<td>1:00 PM</td>
<td>ArcGIS Special Topics: Geocoding &amp; Proximity Analysis</td>
</tr>
<tr>
<td>Details</td>
<td>Wed, Oct 15, 2014</td>
<td>10:00 AM</td>
<td>Analytics and Storage in the Cloud</td>
</tr>
<tr>
<td>Details</td>
<td>Fri, Oct 17, 2014</td>
<td>1:00 PM</td>
<td>Historical GIS</td>
</tr>
<tr>
<td>Details</td>
<td>Mon, Oct 20, 2014</td>
<td>10:00 AM</td>
<td>Splunk for Research Data</td>
</tr>
<tr>
<td>Details</td>
<td>Tue, Oct 21, 2014</td>
<td>1:00 PM</td>
<td>Data Visualization on the Web</td>
</tr>
</tbody>
</table>
Data Visualizations
Mapping – Natural and Social Sciences

http://bit.ly/1Ev1KEJ
Data Visualization - Spatial Humanities

When activists protested the Biennale Festival’s capitalistic structure of art sales a number of photos were taken. This project overlays them with contemporary shots of the urban space. A growing collection of 15 photos make an interactive Google map of the event. Visit http://digitalcities.univiu.org/~ssennett/biennaleprotest.html for more information.

http://bit.ly/1sw7E0R
"Mapping" Patterns

Diseases Across the Top Five Languages of the MEDLINE Database: 1961-2012

Angela Zoss, Trevor Edelblute, Inna Kouper

http://bit.ly/1xF8dJg
Looking Forward: Future (Data) Services

• Increased use of data visualization for both data management and analysis

• Expanded instruction of data literacy across all disciplines (not just quantitative disciplines)

• Increase in “data science” projects
Joel Herndon
Head, Data and Visualization Services
Duke University
@duke_data

Creative Data Services: Mapping and Visualization
Kimberly Silk
Data Librarian, Martin Prosperity Institute at Rotman School of Management, University of Toronto
@kimberlysilk | kimberly.silk@rotman.utoronto.ca

The Embedded Data Librarian
About Kim and her job:

• Data Librarian at a research institute within a leading business school at Canada’s largest university.

• The Martin Prosperity Institute (MPI) explores the possibilities for and the requirements needed to achieve a prosperous future, one in which democracy and capitalism work in support of each other to the benefit of all.

• In terms of data, we predominantly use social science data from Canada and international sources.
The Popular Kind of Data Librarian

• Many data librarians in academic institutions serve the needs of the entire university
  – Undergraduate students needing assistance learning how to use data for assignments and theses
  – Graduate students using data for their dissertations
  – Researchers and faculty using data for funded work

• Services include
  – Data acquisition (from national statistical agencies and other licensed sources)
  – Data discovery, access & distribution (online data catalogue)
  – Assisting with data analysis, conversion, instruction
  – Beginning to look at Research Data Management (RDM) services
The Data Lifecycle – University-wide
A Different Kind of Data Librarian

- More “special” than “academic”
- Embedded in the organization – not in the library – and an integral member of the research teams for all research verticals
- Research projects begin and end with the data librarian:
  - Literature search and data discovery – open and licensed – to determine viability of the project (are data available?)
  - Data acquisition; some help w/ data collection
  - Keeping track of data sources, and methodologies applied
  - Data publishing in tandem with article or report (Repositories)
  - Data cataloging & archiving for discovery & reuse (Dataverse)
The Data Lifecycle - MPI
What I’m wrestling with:

• Data metadata – describing the data sets we’re using, and what models are applied, and analyses used
• Ensuring consistency with other social science data repositories to facilitate sharing
• Planning for expectations of funding agencies – publishing the paper and the supporting data
• Looking for ways to encourage scholars to share their data – using data citation counts as incentive
• Archiving data is easy; challenge is in developing a database of metrics where researchers can pick and choose variables to apply models quickly, i.e. indices of city performance
Consistency within an Institution

• UToronto is currently developing a service delivery plan for Research Data Management (RDM) across disciplines, university-wide.
• My approach at the MPI must be consistent with my institution’s approach.
• Canadian institutions are looking to develop consistency across Canada, starting with the major research universities.
• Still lots to do.
Thank you for listening.

- Questions?
Mike Mertens
Deputy Director and Data Services Manager
Research Libraries UK
@RLUK_Mike

Linked Open Data. Identifying Opportunities.
Overview

• A little about RLUK
• Linked data – RLUK Strategic reasons for and method of publishing (plus challenges!)
• Semantic services: seeing linked open data consumed and re-used in the Library sphere.
• In conclusion: A changing skills and research support base for librarians?
RLUK Background #1

- Founded in 1983
- Originally seven members, now 34 of the UK & Ireland’s research intensive libraries
  - A Collective Approach: Re-shaping the modern research library collection
  - Open Scholarship: Creating a new environment for the communication of research outputs
  - Nothing Hidden, Nothing Lost: Exposing and exploiting our collections
  - Mapping a Changing Research Landscape: The role of libraries in research and researcher training
  - A Creative Community: Nurturing leadership, innovation and skills throughout our libraries
RLUK-Background #2

19th Century British Pamphlets
Linked Open Data - The Why

• RLUK’s commitment to the principles of open extends also to metadata
• Getting ahead of the curve: making best use of legislative change - European PSI (Public Sector Information) Directive, which includes museums, libraries and archives for the first time
• Bring the stuff together – it's what users want!
• Provide more impetus to libraries in the UK consuming and using Linked Open Data as well as producing it
• Push in new directions regarding overall (liaison) library skills base
LOD - The What

• Subset of the RLUK database comprising nearly 20 million bibliographic records from 34 research intensive & national libraries.
• Fourth largest LOD Dataset in the library world (OCLC Research International Linked Open Data Survey, 2014)
• LOD provides a set of procedures and technical standards to allow the reuse of data consistently and reliably across communities and content silos
• At its simplest, Linked Open Data is based on the Semantic Web notion of discovery we are used to, following information from URL to URL. Data and Web resources are linked via stable Uniform Resource Identifiers (URIs)
• See http://linkeddata.org/ for more information!
If you want the mechanics...

Linked Data for Libraries, Archives and Museums
How to clean, link and publish your metadata

“Metadata experts Seth van Hooland and Ruben Verborgh introduce the key concepts of metadata standards and Linked Data and how they can be practically applied to existing metadata.”

http://book.freeyourmetadata.org/

Webinar
Library Connect
The How

Linking discrete sets of data through URIs (person, date, place, event) makes any searched for information more contextualised and informative.

Subject: Québec – History -- French and Indian War, 1755-1763
The Who (so far in the UK)

• Observations on UK LOD Context:
  – “Infrastructure”=Datasets, Projects, Systems, Capacity and Skills based on Linked Open Data
    • Headlining datasets from The British Library, University of Cambridge, Archives Hub, British Museum
    • Projects by Archives Hub, King’s College London and others
    • Internal Systems at British Library and University of Cambridge
    • Developer Capacity at BL_Labs and a relative handful of libraries
  • However - LOD is not being used in a systematic way to deliver services! We wanted and want this to change.
The What Next?

• Library world is typically focused on adding to the LOD or semantic one by publishing LOD.
• Of 172 LOD projects or services surveyed by OCLC Research, only just over 27% also consumed LOD.
• Yet some are well advanced in terms of systems, not projects:
  • AMSL-A linked data Electronic Resource Management at Leipzig University Library.
  • D:SWARM LOD-based data integration and search disambiguation SLUB Dresden.
  • Oslo Public Library OPAC.
  • Linked Jazz, a research project based at the Pratt Institute School of Information and Library Science.
AMSL-Univeristy of Leipzig

- AMSL - A linked data basis for Electronic Resource Management based at SLUB Dresden and Leipzig University Library.
- Funded by the European Union (EFRE) and the Free State of Saxony.
- Uses tools (OntoWiki) developed at the Institute of Applied Informatics (InfAI), University of Leipzig under the LOD2 project (7th EU Framework Programme).
- Business models regarding e-media acquisition are changing rapidly. “Our acquisition department longed for a very flexible way to manage e-resources in future”.

http://amsl.technology
D:SWARM #1

d:swarm is an Open Source Middleware for Data Integration

http://dmp.slub-dresden.de/en/

ERM Frontend  Discovery System  Frontend X

ERM Backend  Integrated Library System  System A

Web, LOD  d:swarm – Middleware for Data Integration  LOD

Research Libraries UK

Webinar
Library Connect

#LCwebinar
Mike Mertens
D:SWARM #2

- Increases the quality of library data by way of deduplication, merging, FRBRization, semantic enrichment and its publication as LOD.

- d:swarm is a data modeling tool with a graphical and easy-to-use interface that poses no obstacles to non-programming librarians.

- Can bring together heterogeneous data from various sectors: museums, archives, libraries, galleries: the discovery holy grail!
Anyone can do it...!
Oslo Public Library #1

- Announced June that its catalogue data will be stored as linked data (within open source Koha ILS)
- Finds linked data a rich metadata model more suitable for future needs than the library specific MARC-format.
- LOD enables it to *use the same metadata format for its physical collection and its digital content*
- Provides a good foundation for search, presentation and integration with other content, internally and externally
“Research into the possibilities for linked open data applications within digital scholarship remains wide open. While a number of research projects are currently exploring methods and tools to clean and open up data for use in linked open data environments, the field of digital scholarship lacks a critical mass of these efforts.”

(Linked Jazz: Building with Linked Open Data, Leanora Lange, the Center for Jewish History & Cristina Pattuelli, School of Information and Library Science at the Pratt Institute). 30 June, 2014.
Linked Jazz #2
LOD DIY - Just a nice theory?

- Improve on or augment existing offerings
- Broaden library skills base (use & make sense of data, coding, UX)
- Open data - reclaim agility for research support development
- Greater ability to match speed of electronic content business models
- Overcome “format hell” and information siloing (link and relate information from galleries, libraries, archives and museums)
- Linking information from different campus systems
- Greater support for digital humanities
LOD-What do we need still for re-use?

- Need for vision: library and vendor engagement ("our system only uses linked data-like concepts")
- Need for developer base
- Awareness of existing tools
- Partnerships between developers and librarians
- Librarians as developers
- Encourage greater documentation of existing APIs
Contact & Acknowledgements

Twitter: @RLUK_Mike
Email: Mike.Mertens@rluk.ac.uk

*With thanks to:* Nuno Freire, The European Library; Owen Stephens, Owen Stephens Consultancy, Leanora Lange, the Center for Jewish History & Cristina Pattuelli, School of Information and Library Science at the Pratt Institute.
Questions and Thank You!

Joel Herndon
joel.herndon@duke.edu
@duke_data

Kimberly Silk
kimberly.silk@Rotman.Utoronto.Ca
@kimberlysilk

Mike Mertens
Mike.Mertens@rluk.ac.uk
@RLUK_Mike

Library Connect
Webinar