Rise and fall of the cataloguer’s empire - an update -

Daniel van Spanje
Senior product operations manager
Metadata services
Global product management
OCLC Leiden – The Netherlands
The paradigm shift
The paradigm shift

- The result of singing is a song
- The result of talking is a talk
- The result of playing is a play

- But the catalogue is dissolving in discovery so what are we doing when we are cataloguing!
The world's libraries. Connected.
The online information industry

- remote user devices were Acoustic Couplers that accepted a standard GPO handset placed into foam cups ....
  This link worked at an amazing 110 baud and these devices were the wonder of the age at the time.
  
Web scale discovery services

• “tools that search seamlessly acrosss a wide range of local and remote content and provide relevance-ranked results”

• “have the ambitious goal of providing a single point of entry into a library’s collection”

De wereldwijde bibliotheken. Verenigd.
The world's libraries. Connected.
Meaning in life
MF Steger, MJ Bundick, D Yeager - Encyclopedia of adolescence, 2011 - Springer
Adolescents use marijuana more than any illegal substance. This essay reviews rates of marijuana usage and perceptions, and identifies risk and protective factors that may be effectively targeted in research efforts and prevention aimed toward adolescents. The
Cited by 357 Related articles All 4 versions Cite Save

[BOOK] Meaning in life
T Metz - 2013 - books.google.com
What makes a person's life meaningful? Thaddeus Metz offers a new answer to an ancient question which has recently returned to the philosophical agenda. He proceeds by examining what, if anything, all the conditions that make a life meaningful have in common.
Cited by 126 Related articles All 8 versions Cite Save

The meaning in life questionnaire: Assessing the presence of and search for meaning in life.
Abstract 1. Counseling psychologists often work with clients to increase their well-being as well as to decrease their distress. One important aspect of well-being, highlighted particularly in humanistic theories of the counseling process, is perceived meaning in life.
Cited by 1533 Related articles All 19 versions Cite Save

[BOOK] The passions: Emotions and the meaning of life
RC Solomon - 1976 - books.google.com
20-20-3456789-10 Copyright © 1976 by Hackett Publishing Company, Inc. All rights reserved.
Some of the most popular social media websites are Baidu Tieba, Facebook (and its associated Facebook Messenger), Gab, Google+, Instagram, LinkedIn, Pinterest, Reddit, Snapchat, Tumblr, Twitter, Viber, WeChat, Weibo, WhatsApp, and YouTube. These social media websites have more than 100,000,000 registered users.

Social media - Wikipedia

People also ask
Is Wikipedia a social networking?
What social networks are there?
What is a space locator in social media?
What is a social media platform?

List of social networking websites - Wikipedia
https://en.wikipedia.org/wiki/List_of_social_networking_websites
This is a list of major active social networking websites and excludes dating websites. For defunct social networking websites, see List of defunct social ...
The world's libraries. Connected.
Ludwig Wittgenstein

Philosopher

Ludwig Josef Johann Wittgenstein was an Austrian-British philosopher who worked primarily in logic, the philosophy of mathematics, the philosophy of mind, and the philosophy of language. Wikipedia

Born: April 26, 1889, Vienna, Austria
Died: April 29, 1951, Cambridge, United Kingdom

Movies: Wittgenstein, Tractatus: M.A. Numminen Sings Wittgenstein, As from Afar

Influenced by: Bertrand Russell, Gottlob Frege, Sigmund Freud, more

Quotes

The limits of my language mean the limits of my world.
The world's libraries. Connected.

Ludwig Wittgenstein

Philosopher

Ludwig Josef Johann Wittgenstein was an Austrian-British philosopher who worked primarily in logic, the philosophy of mathematics, the philosophy of mind, and the philosophy of language. Wikipedia

Born: April 26, 1889, Vienna, Austria
Died: April 29, 1951, Cambridge, United Kingdom

Movies: Wittgenstein Tractatus, M.A. Numminen Sings Wittgenstein, As from Afar

Influenced by: Bertrand Russell, Gottlob Frege, Sigmund Freud, more

Quotes

The limits of my language mean the limits of my world.
The human body is the best picture of the human soul.

Whereof one cannot speak, thereof one must be silent.

Books

Tractatus Logico-Psychologicus

Philosophical Investigations

Blue and Brown Books

On Certainty

Culture and Value

People also search for
The world's libraries. Connected.
The world's libraries. Connected.
Tractatus logico-philosophicus

Auteur: Ludwig Wittgenstein; C. K. Ogden


Editie/Formaat: Print book: Engels

Database: WorldCat

Samenvatting: The author is widely regarded as the most important philosopher of the twentieth century. In this title, he offers a radical new theory of logic, along the way addressing such problems as the

Beoordeling: ★★★★★ (nog niet beoordeeld) 0 met beoordelingen - U bent de eerste

Onderwerpen: Logic, Symbolic and mathematical, Language and languages -- Philosophy.

Meer in deze trant: Vergelijkbare items
The world's libraries. Connected.
About The Project

At the heart of our work are oral histories of jazz musicians. We are currently expanding our data sources to include other document types and music-related datasets.

In the first phase of the project — funded through an OCLC/ALISE Library and Information Science Research grant — we experimented with the actual content of interview transcripts, generating new triples based on data rather than converting existing metadata. The 50+ interview transcripts we’ve used to create LOD come from the Rutgers Institute for Jazz Studies Archives, Smithsonian Jazz Oral Histories, the Hamilton College Jazz Archive, UCLA's Central Avenue Sounds series, and the University of Michigan's Nathaniel C. Standifer Video Archive of Oral History. For a full list of transcripts, see our Data Sources page.

To create LOD, we developed a suite of tools: a transcript analyzer, a name mapping and curator tool, and a crowdsourcing tool. These tools operate together to find names mentioned during the interview in order to assign a positive identification to each, disambiguating names using online resources like DBpedia and VIAF. The transcript analyzer also recognizes the question and answer structure of the oral history. As people are mentioned by an interviewee, simple RDF triples between interviewee and persons mentioned are created in the form of linkedOf. These triples are then be mapped to the semantic web. For more on the transcript...
The catalogue: characteristics

- Discovery happens at the library
  - The physical library
  - The online catalog

- Collection centered

- Item oriented
The catalogue: characteristics

• Discovery happens at the library
  – The physical library
  – The online catalog

• Collection centered

• Work oriented
Web-scale discovery services

• Discovery happens at the online library

• From collection to connection

• Subject and discipline oriented
The web

• Discovery happens outside the library

• From connection to navigation and linking

• Context oriented
The web

• Discovery happens outside the library

• From connection to navigation and linking

• Engagement oriented
The world's libraries. Connected.

- Item / work
- Subject
- Context
The world's libraries. Connected.

Subject

Item / work

Physical Catalogue

Context

Digital Discovery
The world's libraries. Connected.

Context

Subject

Item / work

Physical Catalogue

Digital Discovery

Centralized searching

Decentralized searching
The world's libraries. Connected.

**Item / work**

**Subject**

**Context**

**Physical**

**Catalogue**

**Centralized searching**

**Local**

**Digital**

**Discovery**

**Decentralized searching**

**Global**

Centralized searching vs. Decentralized searching.
Metadata management process

Item → Cataloguer → Catalogue
Metadata management process

- Item
- "Cataloguer"
- Bibliographic records
- Authority control
- Item records
- OPAC
- IT staff
- Discovery
- Web
Metadata management process

Collection manager

IT staff

IT staff

Discovery

Web

The world's libraries. Connected.
<table>
<thead>
<tr>
<th>Collection</th>
<th>Shared by</th>
<th>In My Holdings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oxford Scholarship Online</td>
<td>Not Held by my library</td>
<td>[Select Collection]</td>
</tr>
<tr>
<td>Provider: Ingenta</td>
<td>(0 / 4699) Titles</td>
<td></td>
</tr>
<tr>
<td>Customizable: Discontinued</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oxford Scholarship Online Biology E-Books</td>
<td>Not Held by my library</td>
<td>[Select Collection]</td>
</tr>
<tr>
<td>Provider: Oxford University Press</td>
<td>(0 / 131) Titles</td>
<td></td>
</tr>
<tr>
<td>Customizable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oxford Scholarship Online Business and Management E-Books</td>
<td>Not Held by my library</td>
<td>[Select Collection]</td>
</tr>
<tr>
<td>Provider: Oxford University Press</td>
<td>(0 / 273) Titles</td>
<td></td>
</tr>
<tr>
<td>Customizable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oxford Scholarship Online Classical Studies E-Books</td>
<td>Held by my library</td>
<td>[Refresh]</td>
</tr>
<tr>
<td>Provider: Oxford University Press</td>
<td>(308 / 308) Titles</td>
<td></td>
</tr>
<tr>
<td>Customizable</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Metadata management process

- Library materials
- Metadata / Collection manager
- Bibliographic records
  - Identifier control
  - URL control
- Knowledge base
- IT staff

Discovery

IT staff

OPAC

Web

Google™
The fundamental question

How to connect users to library collections on the web?
How to expose library collections on the web?
is a global library cooperative representing 84,000 institutions than 120 countries. On behalf of the cooperative, OCLC manages WorldCat, the largest, most authoritative and most consulted bibliographic database in the library industry.

g value for you and your customers

with hundreds of the leading publishers and application service providers in the information industry through the following

Record creation and metadata services

OCLC is the leading authority on library data and metadata creation. Learn how to license WorldCat data and outsource your cataloging needs, including record delivery to your customers.

Discovery and syndication services

OCLC’s syndication program is designed to drive traffic back to the content that OCLC libraries purchase or license, which supports your bottom line.

APIs and web services

Whether your business serving libraries or else integrate a wealth of data into your application quickly and easily through OCLC’s APIs and web services.
Publishers, library service vendors and consumer web services all work with OCLC to increase sales, improve information services, save on infrastructure costs and get access to high-quality library data. If your organization would benefit from a similar partnership with OCLC, we’d love to hear from you.

Submit your inquiry →

Become an OCLC partner like these industry leaders
What the web wants

Some things the web wants:

1. Familiar structures
2. A network of links
3. Entity identifiers
First, define ALL THE THINGS
First, define ALL THE THINGS.

Linked Data "Entities"
So the goal of linked data is to produce machine-understandable knowledge about things we are interested in.

As librarians, we can jumpstart this process by upgrading the descriptions of things that librarians have always collected information about...authors, works, subjects.
How to get data on the web

• Human readable data

• Machine readable data

• Machine actionable data
OCLC’S 2015 INTERNATIONAL LINKED DATA SURVEY

SOURCE: KAREN SMITH-YOSHIMURA
2015 responding institutions by type

- Academic library: 31%
- National library: 20%
- Network: 14%
- Government: 10%
- Scholarly: 8%
- Public Library: 7%
- Museum: 4%
- Other: 6%

71 institutions total
What is published as linked data

- Other
- Ontologies/vocabularies
- Geographic data
- Encoded archival descriptions
- Digital collections
- Descriptive metadata
- Datasets
- Data about museum objects
- Bibliographic data
- Authority files
<table>
<thead>
<tr>
<th>2015 linked data sources most consumed</th>
<th>2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>VIAF (Virtual International Authority File)</td>
<td>41</td>
</tr>
<tr>
<td>DBpedia</td>
<td>36</td>
</tr>
<tr>
<td>GeoNames</td>
<td>35</td>
</tr>
<tr>
<td>id.loc.gov</td>
<td>35</td>
</tr>
<tr>
<td><strong>Resources we convert to linked data ourselves</strong></td>
<td><strong>17</strong></td>
</tr>
<tr>
<td>Getty's AAT</td>
<td>16</td>
</tr>
<tr>
<td>FAST (Faceted Application of Subject Terminology)</td>
<td>15</td>
</tr>
<tr>
<td>WorldCat.org</td>
<td>15</td>
</tr>
<tr>
<td>data.bnf.fr</td>
<td>12</td>
</tr>
<tr>
<td>Deutsche National Bib Linked Data Service</td>
<td>12</td>
</tr>
</tbody>
</table>
VIAF aggregates identifiers

Chomsky, Noam, 1928–
Chomsky, Noam.
Chomsky, Noam (Noam Avram), 1928–
Chomsky, Noam (American linguist, philosopher, born 1928)
VIAF ID: 89803084 (Personal)
Permalink: http://viaf.org/viaf/89803084
ISNI: 0000_0003_8149_3996
Wikidata disseminates identifiers

Jerry Brown (Q152451)

VIAF identifier: 1320107

Authority control:
- WorldCat: VIAF: 1320107
- LCCN: n80131844
- ISNI: 0000 0001 1487 0443
- GND: 119129352578
- ULAN: 500274877

전기 동제

Wikipedia

The Free Encyclopedia
LINKED DATA INITIATIVES

- BIBFrame 2.0 [https://www.loc.gov/bibframe/](https://www.loc.gov/bibframe/)
- LD4P (piloting the transition of technical services workflows to a linked data environment) [https://ld4l.org/](https://ld4l.org/)
- BIBFLOW - Reinventing Cataloging: Models for the Future of Library Operations [https://bibflow.library.ucdavis.edu/about/](https://bibflow.library.ucdavis.edu/about/)
- Public Library of OSLO [https://sok.deichman.no/work/w903204ae59a040426baf680b42c1c2b2](https://sok.deichman.no/work/w903204ae59a040426baf680b42c1c2b2)
  - VIAF / Schema.org / WorldCat Works / WorldCat Persons / WorldCat Translations (coming soon!)
But how about MARC? ....
The world's libraries. Connected.

- Use uniform titles
- Use added entries with role codes (7xx and $4)
- Use 041 for translations, including intermediate translations
- Use indicators to refine the meaning
- Use the most specific fields appropriate for a descriptive task
- Minimize the use of 500 fields
- Obey field semantics
- Avoid redundancy

If you must use free text:
- Use established conventions
- Use standardized terms

**Making MARC “Linked Data Ready”**

- Least machine-processable
  - Most machine-processable
  - Algorithmically recoverable

- Use uniform titles
- Use added entries with role codes (7xx and $4)
- Use 041 for translations, including intermediate translations
- Use indicators to refine the meaning

- Use the most specific fields appropriate for a descriptive task
- Minimize the use of 500 fields
- Obey field semantics
- Avoid redundancy

- Use established conventions
- Use standardized terms

Least machine-processable

Most machine-processable

Algorithmically recoverable

OCLC®
Working With the PCC To Make MARC LD Ready

‘Work’ Task Force
Analyze the ‘Work’ definitions referenced in library linked data.
• How are they similar or different?
• How do they relate to the classic FRBR definition?
• What are the use cases for ‘Work?’

How should Work URIs be represented in MARC records?
• What are the best practices for adding URIs to MARC records to ease the conversion to linked data?
• How will cataloging or resource description workflows be affected?

‘URI’ Task Force
Library data stored as records

- **person**
  - author
  - edition
  - holding
  - classification
  - publisher
  - date of publication

- **place**
  - location
  - title
  - source
  - ISBN

- **object**
  - organization

- **concept**
  - work

The world’s libraries. Connected.
Building a library knowledge graph

Field in a record vs. entity in knowledge graph
Field in a record vs. entity in a knowledge graph

- Martin Heidegger
- Germany
- this copy of "What is a thing"
- "Die Frage nach dem Ding"
- "What is a thing"
- Metaphysics
- library
- organization
- work
- concept
- object

Field in a record vs. entity in a knowledge graph
The world's libraries. Connected.
Our Challenges

• End of cataloging? Cataloging becomes license management and checking subscriptions. At the same time the focus of many libraries shifts from acquiring externally created content toward disclosing internally curated assets! Think of special collections, digitized collections, data sets, research data.

• Cataloguing needs to change from record management to entity management! And we need persistent identifiers for these entities. How can we develop these new workflows? Can we restrict ourselves to publishing data in RDF or do we need to create RDF triples?

• Do we know where the user is. Should we not move away from the focus on descriptive metadata and focus more on transaction data to engage with the user?
Further reading

• Lorcan Dempsey’s blog: http://orweblog.oclc.org/towards-the-facilitated-collection/

• Jean Godby, Linked data, the bibframe initiative and OCLC
  https://www.oclc.org/content/dam/oclc/membership/canadacouncil/presentations/BIBFRAME-Initiative.pdf

• Roy Tennant, Data designed for discovery
  https://www.oclc.org/content/dam/oclc/events/member-forums/datafordiscovery_tennent.pdf

• Karen Coyle, FRBR, the Domain model, chapter 3 in: RDA Vocabularies for a Twenty-First Century Data Environment, 2010
  https://journals.ala.org/index.php/ltr/article/view/4676/5553

• Karen Coyle, RDA in RDF, ch. 4 in: RDA Vocabularies for a Twenty-First Century Data Environment, 2010
  https://journals.ala.org/index.php/ltr/article/view/4672/5545
Rise and fall of the cataloguer’s empire - an update -

Daniel van Spanje
Senior product operations manager
Metadata services
Global product management
OCLC Leiden – The Netherlands

Thank you!