

**NTK**

50°6'14.083"N, 14°23'26.365"E

Národní technická knihovna  
National Technical Library

210 mm

# Linked Library Data

Jindřich Mynarz

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

# What is this talk about?

210 mm

- In terms of *Polythematic Structured Subject Heading System*, it will be about:
  - „libraries“ (PSH6536)
  - „type of data“ (PSH6591)

```
<> dc:subject <http://psh.ntkcz.cz/skos/PSH6536> ,  
          <http://psh.ntkcz.cz/skos/PSH6591> .
```

# Knowledge technologies

210 mm

## 1. System

- Meaning of a concept is identical with its place in the system.
- Systematic knowledge is complete, closed and invariant.
- Example: library

# Knowledge technologies

210 mm

## 2. Structure

- Meaning is defined by the *relationships* between concepts.
- Example: dictionary

## 3. Net

- Meaning is created in the process of interpretation of an *actual configuration* of the net.
- Example: the Web

Source: PETŘÍČEK, Miroslav. Síť čili tělo bez orgánů. *Filosofický časopis*, r. 46, č. 1, 1998.

# Linked data

210 mm

„Linked data is the pragmatic implementation of the semantic web.“ (Richard Wallis)

# Linked data

210 mm

- means for achieving the semantic web vision
- web-native data model
- based on Web standards (HTTP) and semantic web technologies (RDF, RDFS, OWL)
- ecosystem of structured data – the Web of Data

# Linked data principles

210 mm

1. Use **URIs** as names for things.
2. Use **HTTP URIs** so that people can look up those names.
3. When someone looks up a URI, **provide useful information**, using the standards (RDF, SPARQL).
4. Include **links to other URIs**. so that they can discover more things.



# Library data

210 mm

**P**

VICKERY, B.C.

On Retrieval System Theory. 2nd.ed. *W*

London, Butterworths 1965.191s.

Přír.č. 16.557/71

Přír.č. 9314/67-D

**SK. PHA**

Source: National Library of the Czech Republic card catalogue

# Library data

210 mm

- Libraries are about **documents** (information resources), not about data.
- And, of course, *libraries 2.0* are all about **people** (users).

# Library data

210 mm

- Bibliographic records are *black-boxes*.
  - They are not easily machine-readable.
  - Low data granularity.

# Library data

210 mm

- Library data is:
  - rather *text* than data (data buried in text)
  - *dumb* data
  - *alive* only within library software, on the Web it is *dead* data

# Library data

210 mm

- Contains duplicate data – **redundancy**
  - Once necessary for unreliable transport mechanisms to ensure consistency *is now not needed.*
  - Some data needs to be entered multiple times, so mistakes can be made in the process of data entry.
  - Makes updates difficult.
  - Uncontrolled redundancy (as opposed to intentional replication) wastes disk space.

# Library data

210 mm

- Library-specific access protocols.
- library://
  - Z39.50
  - OAI-PMH

# Library data

210 mm



Source: <http://www.flickr.com/photos/lwr/144004119/>

# MARC

210 mm

- Machine-readable cataloging
- MARC for mark-up
- Mark-up around text



# MARC

210 mm

- MARC is conceived as a data interchange format.
- But it is way more than just exchange format.
- It is also often used as a default view (interface) for catalogers.

# MARC

210 mm

- Data model – traditional catalog card
  - field & subfield granularity
  - „packages“ of metadata in the form of records (record paradigm)

- Binary MARC, ISO 2709

```

00772aam a2200241 a
450000100090000000300080000900500170001700800410003402000220007504000150009704
100130011204200070012508400200013210000430015224000220019524500660021725000140
0283260006200297300002500359500003300384700003300417856008000450-11288890-SE-
LIBR-20090415151036.0-090126s2009 sw |||| |00| f swe - a978-91-1-302192-8- aLGENdL-
dS-1 aswehfre- 9NB- aHj.01=c2kssb/8-1 aLe Clézio, Jean-Marie Gustave,d1940--10aDésert.-
ISvenska-10aÖken /cJ.M.G. Le Clézio ; översättning av Ulla Bruncrona- a[Ny utg.]- aStockholm :-
bNorstedt,c2009e(Smedjebacken :fScandbook)- a400, [1] s. ;c18 cm- aFörsta svenska upplaga
1984-1 aBruncrona, Ulla,d1943-4trl- zOmslagsbild-
uhttp://bilder.panorstedt.se/bilder/omslag/224/13021928_O_1.jpg-000804cz a2200241n
450000100060000000300080000600500170001400800410003103500230007204000110009504
300070010610000430011340000550015640000430021140000330025440000530028767000610
03406700007004016700000800408670001200416678006200428700007200490-70378-SE-
LIBR-20090609102922.0-011206n| ac|nnaabn |n aaa d- a(LIBRIS)A000039332- aSbswe-
afr-1 aLe Clézio, Jean-Marie Gustave,d1940--1 aLe Clézio, J.-M. G.q(Jean-Marie Gustave),d1940-
-1 aClézio, Jean Marie Gustave Le,d1940--1 aLe Clézio, J. M. G. ,d1940--1 aLe Clézio, J.M.G.-
q(Jean Marie Gustave),d1940-- aAfrikanen : [porträtt av en far] / J. M. G. Le Clézio- aNE- aBnF-
aLC Auth-D aSkönlitterär författare. Nobelpriset i Litteratur 2008-10aLe Clézio, J.-M. G.q(Jean-
Marie Gustave),d1940-0(DLC)n 50039803-000196nxa a22000971
4500001000900000003000800009004000900017005001700026008003300043852002200076-23
735790-SE-LIBR-11288890-20090414083906.0-090414|| |||||001|||||000000- bLhb09/j2767 xta-0
  
```

(Source image)

- MARC-XML (or OAI-MARC) is just MARC with angle brackets.

## MARC21

```
245 10 |a Programming the semantic web / |c Toby Segaran, Colin Evans,  
and Jamie Taylor
```

## OAI-MARC

```
<varfield id="245" i1="1" i2="0">  
  <subfield label="a">Programming the semantic web /</subfield>  
  <subfield label="c">Toby Segaran, Colin Evans, and Jamie Taylor</subfield>  
</varfield>
```

# MARC

210 mm

- Over **40** years old.
- Should be a dead language, just like Latin.
- IE6 of libraries.

# Links in library data

210 mm

- There already are links in library data.
- Rather *implicit* than explicit.

# Links in library data

210 mm

```
LDR 00000nam-a22-----a-4500
001 000660086
003 CZ-PrSTK
005 20100114152838.0
008 090430s2009----xxua---f-----001-0-eng-d
020 |a 978-0-596-15381-6
040 |a DDB |d SzZuIDS NEBIS FHNW-TW |d ABA013 |b cze
050 4 |a TK5105.88815 |b .S34 2009
072 7 |a 004.7 |x Počítačové sítě |2 Konspekt |9 23
080 |a 004.738.52
080 |a 004.65
1001 |a Segaran, Toby |7 ntk2009537638 |4 aut
24510 |a Programming the semantic web / |c Toby Segaran, Colin Evans, and Jamie Taylor
250 |a 1st ed.
260 |a Sebastopol : |b O'Reilly, |c 2009
300 |a 280 s. : |b il.
500 |a Obálkový podnázev: build flexible applications with graph data
500 |a Obsahuje rejstřík
65007 |a www |x vt |2 psh
65007 |a internet |x vt |2 psh
65007 |a vyhledávání informací |x in |2 psh
655 4 |a monografie
65007 |a sémantický web |x vt |2 psh
7001 |a Evans, Colin |q (Colin Hearne) |7 stk2010477033 |4 aut
7001 |a Taylor, Jamie |7 stk2010477034 |4 aut
910 |a ABA013 |b A 45734
SYS 000660086
```

# Links in library data

210 mm

LDR 00000nam a22

Code for Deutsche  
Digitale Bibliothek

United States  
country code

Language code  
for English

008 090 430s2009 --- **xxu**a --- f --- 001-0-**eng**-d

020 |a 978-0-596-15381-6

040 |a **DDB** |d **SzZuIDS NEBIS FHNW-TW** |d **ABA013** |b cze

Code for  
Hochschulbibliothek  
Technik und Wirtschaft

Code for  
National Technical Library



# Links in library data

210 mm

```
050 4 |a TK5105.88815 |b .S34 2009  
072 7 |a 004.7 |x Počítačové sítě |2 Konspekt |9 23  
080 |a 004.738.52  
080 |a 004.6
```

**Library of Congress  
Classification**

**Conspectus  
subject category**

**Universal Decimal  
Classification Numbers**

# Links in library data

210 mm

## Main author

```
1001 |a Segaran, Toby |7 ntk2009537638 |4 aut
24510 |a Programming the semantic web / |c Toby Segaran, Colin Evans, and Jamie Taylor
250 |a 1st ed.
260 |a Sebastopol : |b O'Reilly, |c 2009
300 |a 280 s. : |b il.
500 |a Obálkový podnázev: build flexible applications with graph data
500 |a Obsahuje rejstřík
65007 |a www |x vt |2 psh
65007 |a internet |x vt |2 psh
65007 |a vyhledávání |x vt |2 psh
655 4 |a monografie |x vt |2 psh
65007 |a sémantický web |x vt |2 psh
7001 |a Evans, Colin |g (Colin Hearne) |7 stk2010477033 |4 aut
7001 |a Taylor, Jamie |7 stk2010477034 |4 aut
```

## Additional authors

# Links in library data

210 mm

**Geographical  
place of publication**

260 |a **Sebastopol** : |b **O'Reilly,** |c 2009

**Name of publisher**

# Links in library data

210 mm

## Polythematic Structured Subject Heading System headings

```
65007 | a www |x vt |2 psh  
65007 | a internet |x vt |2 psh  
65007 | a vyhledávání informací |x in |2 psh  
655 4 | a monografie  
65007 | a sémantický web |x vt |2 psh
```

Genre/form data  
or focus term

# Links in library data

210 mm

- Links use **words** as identifiers.
  - bound to a natural language
    - ambiguous,
    - informationally incomplete (many concepts require more words; e.g., balloon flying),
    - differences and errors in spelling
  - partial
  - non-unique, non-universal

# Links in library data

210 mm

- Can URIs be used as identifiers?
- Hacking MARC to accept URIs.
  - Is RDFa hacking HTML?
  - using newly added subfields
- New standards - RDA: Resource Description and Access

# Linked library data

210 mm

- What does it mean for library data to exist within the capabilities and constraints of the Web architecture?
- Linked data as a **single, unified interface** for library data.
- **Web-compatible** data carrier.

# Legacy data to linked data

210 mm

## 1. Choosing datasets

- interesting for other parties
- machine-readable
- distributed with an open licence



# Legacy data to linked data

210 mm

## 2. Choosing the way of description

- established and popular ontologies
- controlled vocabularies
- code lists

# Legacy data to linked data

210 mm

## 3. Creating mapping

- internal and external
- identity resolution based on partial identifiers

# Legacy data to linked data

210 mm

## 4. Storing data

- RDF triple store

# Legacy data to linked data

210 mm

## 5. Lightweight access API

- URI patterns design
- Content negotiation
- SPARQL endpoint

# Linked library data

210 mm



Source: <http://www.flickr.com/photos/thebirdwells/2721654446/>

# Linked library data at National Technical Library

210 mm

- Polythematic Structured Subject Heading System
  - <http://psh.techlib.cz/skos>
- [data.techlib.cz](http://data.techlib.cz) ?
  - ISSN database
  - Bibliographic records
  - Name authority records



# Benefits of linked data for libraries

210 mm

- *„Future for bibliographic control will be collaborative, decentralized, international at scope, and Web-based.“ (On the Record, 2008)*

# Benefits of linked data for libraries

210 mm

- Multiple distribution channels
- Interoperability
- Data integration
- Decentralization of the cataloging process



# Benefits of linked data for libraries

210 mm

- **Linked library data:**
  - resource discovery
  - navigation interface for library users (data as a navigation tool)
  - explicit paths from vague to precise queries
  - „outsourcing” library authorities to external datasets (e.g., DBPedia)
  - labeled links are amenable to manipulation by computers
  - explicit bibliographic relations

# Benefits of linked data for libraries

210 mm

- **Linkable** library data:
  - incoming links
  - search engine visibility

# Benefits of libraries for linked data

210 mm

- „We appreciate libraries, we can not live without them, but we take them for granted.“  
(Alex Iskold)

# Benefits of libraries for linked data

210 mm

- Brand of quality and trustworthiness
- Vast experience in managing information resources and services
  - But, this knowledge may not be useful any longer under current circumstances.
- Bibliographic vocabularies as domain ontologies

# Barriers to innovation

210 mm

- Libraries
  - long-established library processes
  - out-of-date library standards
- Library software vendors
  - libraries are often bound by long-lasting contracts
  - use of MARC eliminates the possibility to use standard business software

# Barriers to innovation

210 mm

- Technological determinism: can the linked data technology drive the change and innovation?
- Chicken-egg problem
  - without the necessary technology backing it is not possible to re-structure library processes to produce more linked data,
  - and without the library's demand, there will be no software solution that enable that.

# Consequences of maintaining the status quo

210 mm

- Navigation restrictions.
- Cooperation restrictions.
  - Standardization is a requirement to cooperation.
  - Elimination of „local practices“ in favour of standards.
- Cost of maintenance.
- Quality of data influences the quality of services, that can be built upon them.
- Library as „a book museum“.
- Need for internationalization puts new requirements on bibliographic data.

# What is happening?

210 mm

- RDA: Resource Description and Access
  - Metadata Registry Project
  - <http://metadataregistry.org>



# What is happening?

210 mm

- OpenLibrary
  - <http://www.openlibrary.org>
  - „One web page for every book.“

# What is happening?

210 mm

[Log in](#) / [Sign Up](#)

**SUBJECTS**

**AUTHORS**    **ADD A BOOK**

**LISTS**    **RECENTLY**    **ABOUT US**

**OPEN LIBRARY**

One web page for every book.

Only show ebooks    [More search options](#)

**Ever wanted to play librarian?  
It's OK. We all do.**

If you love books, why not *help build a library*?

**Browse the Catalog**

You can always search, but try jumping in to **subjects**, like [Astronauts](#), [Texas](#), [Diaries](#), [Riots](#), [Assam](#), [Pleistocene](#), [Colette](#) or [Carpentry](#)...

**Add a Book**

There are millions of records in Open Library, but nowhere near every book ever published! Would you like to [try adding a book](#) to the library?

**Open an Account**

You can keep track of your edits if you [open an account](#).

It's a wiki too, which means anyone can edit or correct it anytime.

[blog.openlibrary.org](http://blog.openlibrary.org)

- [We're Hiring!](#) - April 6, 2010
- [jQuery Magic Goes Upstream](#) - March 25, 2010

[Top](#) | [Home](#) · [Add a Book](#) · [Subjects](#) | [Authors](#) | [Lists](#) · [About Us](#) | [Help](#) · [Developers](#)

# What is happening?

210 mm

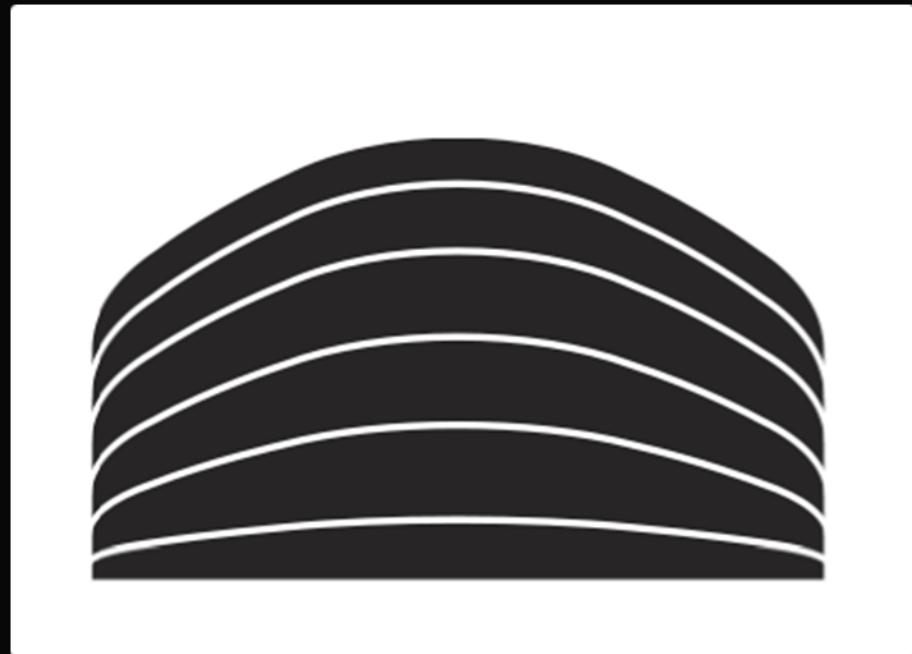
- German National Library has released its linked data prototype for authority data.
  - <http://d-nb.info/gnd>



# What is happening?

210 mm

- National Technical Library
  - <http://data.techlib.cz> (?)



# Summary

210 mm

- „Libraries must transition to the Web future without delay in regard to retain their significance as information providers.“ (On the Record, 2008)

# Summary

210 mm

- Interact with data from other information ecologies.
- Machine applications are also users of the data libraries produce.
- Becoming a part of the data supply chain.

210 mm

# Thank you!

Jindřich Mynarz  
E-mail: [jindrich.mynarz@techlib.cz](mailto:jindrich.mynarz@techlib.cz)  
Twitter: [@jindrichmynarz](https://twitter.com/jindrichmynarz)