The New Era of Bibliographic Data in Libraries

2012 MOBIUS Annual Conference
Wednesday, June 6, 2012

Felicity Dykas

Head, Catalog Department
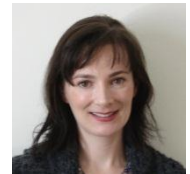
MU Libraries

University of Missouri

Heather Moulaison

Assistant Professor

School of Information Science & Learning Technologies

University of Missouri

@libacat

Bibliographic records today

- **ILS / Catalogs**
  - AACR2
    - With many rules still based on requirements of the card catalog
  - MARC encoding
    - Cannot accommodate features now used in web documents

- **Separate metadata banks**
  - Databases and indexes

- So …

You say you want a revolution in librarianship?

- Start with the catalog: Library metadata today
  - Imagine the possibilities
  - Entity-relationship models are helping us plan
  - RDA will help us implement
- The revolution has entered cyberspace
  - Web documents
  - Cloud computing
  - Semantic web technologies/linked data

You know them well…

- Our current library systems:
  - Contain some encoded data
    - Content the system can understand
    - E.g. fixed fields for language when patron wants to **limit** to Spanish

You know them well…

- Our current library systems:
  - Contain a lot of text-strings
    - Characters that are meaningless to the computer
    - Must be typed as such to be retrieved
    - Can include typos

  Place of publication may be given as
    - Kansas City, Mo. OR Kansas City, MO
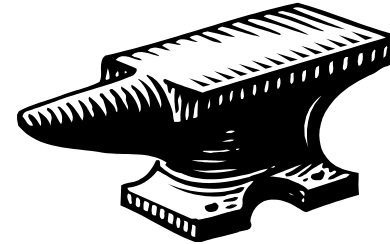   OR KC OR Kansas Ciy, Mo.

Catalogs have not been stagnant

- We are seeing some big changes in access via more user-friendly front-ends
  - Social interfaces allowing for tagging (e.g. , LibraryThing for Libraries)
  - Discovery interfaces (e.g. Summon)
  - New options (e.g., facets)

…yet the underlying system remains the same!

You may not realize…

- Our current library data
  - Is housed in …

flat files

100 1	**McGill, Catherine, |d 1945-**	
245 10	**Cataloging for beginners / |c by Catherine McGill.**	
260	**Manhattan, KS : |b Space Pr., |c 1995.**	
300	**136 p. : |b ill. ; |c 28 cm.**	
650 0	**Cataloging.**	

Current structure

Bibliographic Record

100 1   **McGill, Catherine, |d 1945-**

245 10   Cataloging for beginners / c by Catherine McGill.

260   Manhattan, KS : |b Space Pr., |c 1995.

300   136 p. : |b ill. ; |c 28 cm.
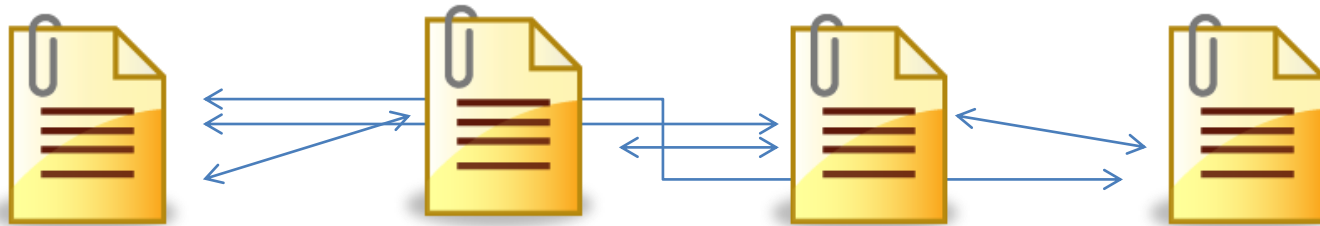
650 0   **Cataloging.**

Authority Record

100 1   **McGill, Catherine, |d 1945-**

670   Cataloging for beginners, 1995 |b t.p. (Catherine McGill)

Authority Record

150   **Cataloging**

- Our current library data
  - Has a **record** structure that makes relationships between *parts* of different records difficult to exploit

So…

- Our current library data
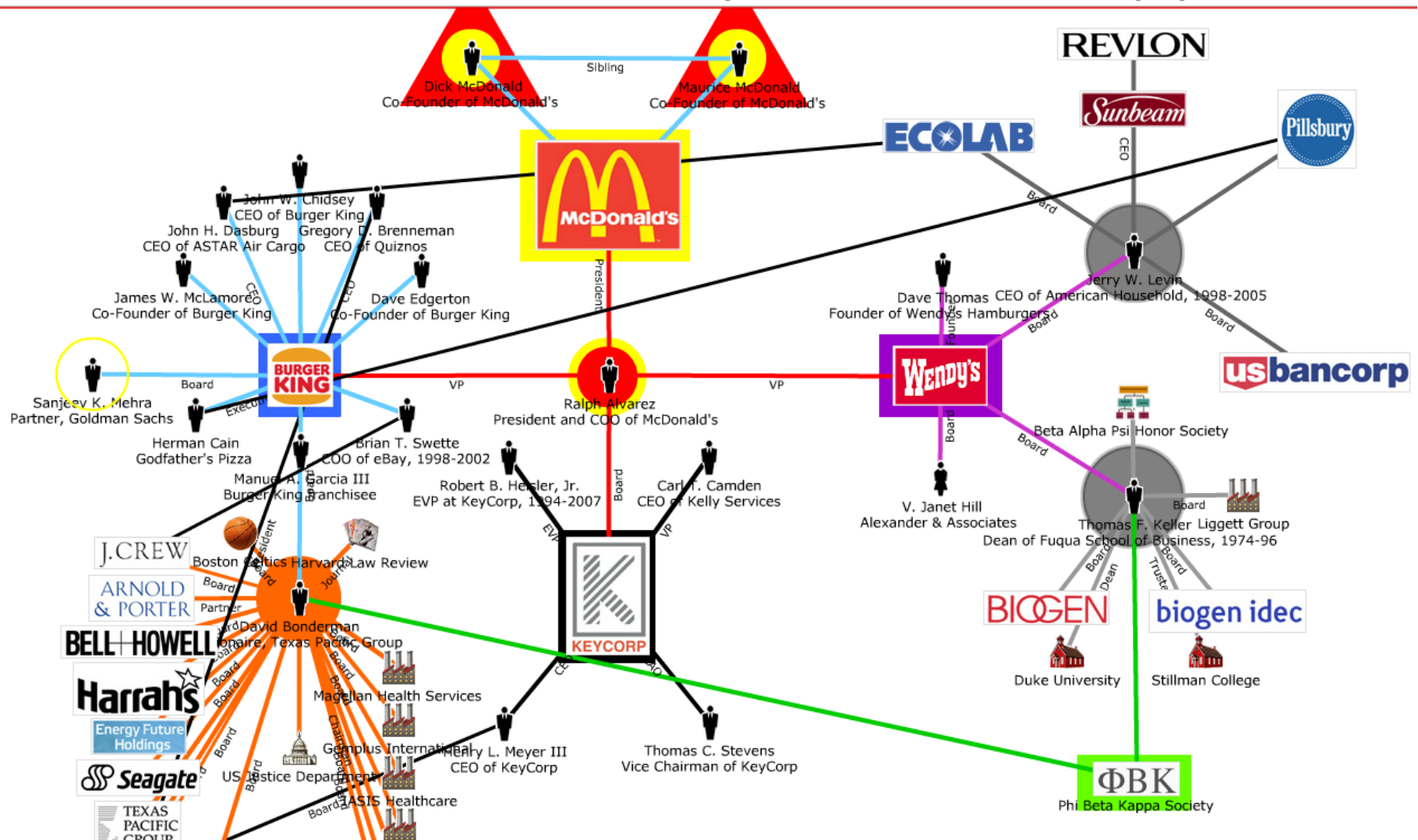 - Is not web-ready

  - Stored in bibliographic records
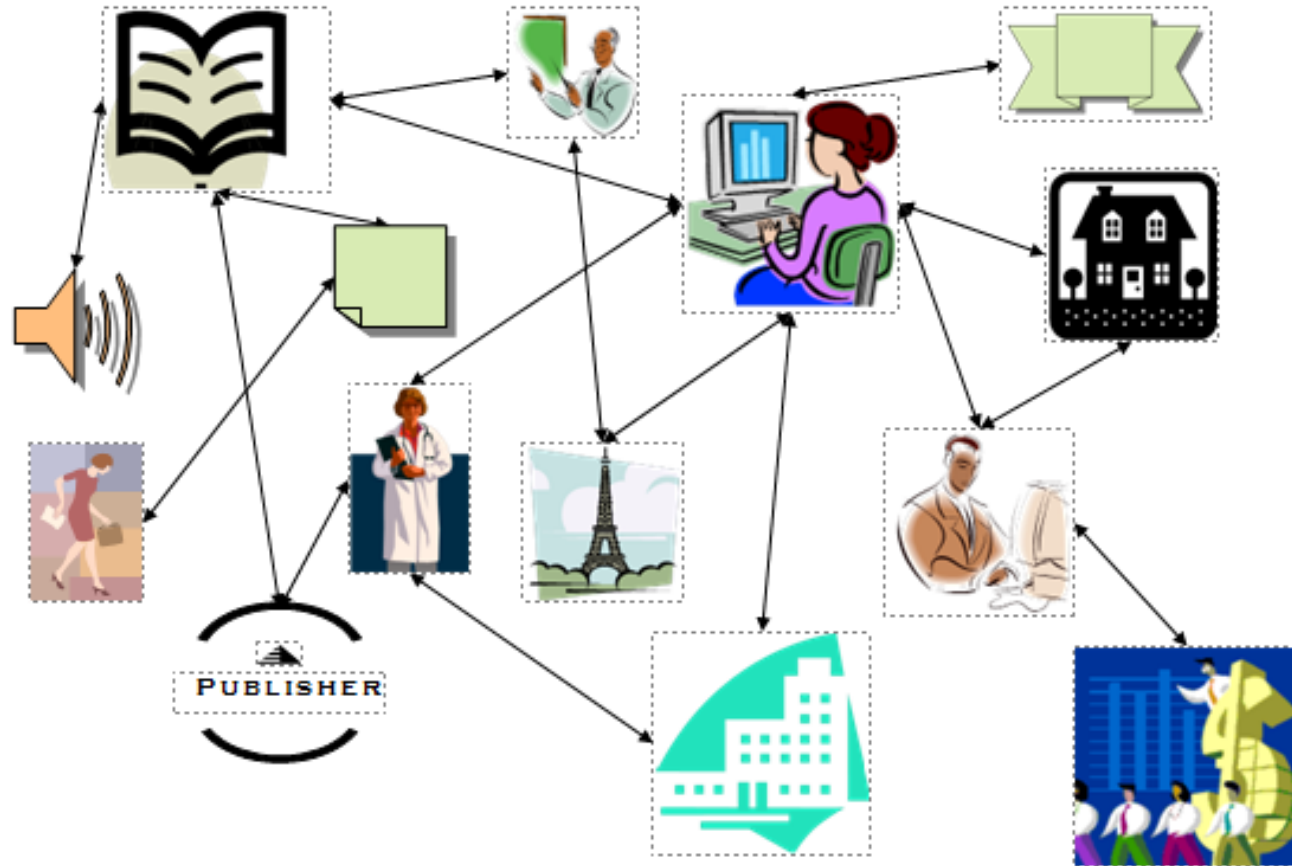 - Stored as text strings

McDonald's, the corporate body

- Lots of information to be recorded in a record about McDonald's (founder, date of incorporation, sales, number of employees, executive board, etc.)
- http://www.nndb.com/company/744/000044612/
- If content is stored as data that's machine actionable, it's possible to show a lot of interesting relationships...

McDonald's family (NNDB Mapper)

A book has a similar web of relationships

How do we get

From this…   ## To this…

Bibliographic Record

100 1	**McGill, Catherine, |d 1945-**	
245 10	Cataloging for beginners / c by Catherine McGill.	
260	Manhattan, KS : |b Space Pr., |c 1995.	
300	136 p. : |b ill. ; |c 28 cm.	
650 0	**Cataloging.**	

Current structure

Authority Record

100 1	**McGill, Catherine, |d 1945-**
670	Cataloging for beginners, 1995 |b t.p. (Catherine McGill)

Authority Record

150	**Cataloging**

Enter… the FR-family and relationships

- FRBR (Functional Requirements for Bibliographic Data)

- FRAD (Functional Requirements for Authority Data)

- FRSAD (Functional Requirements for Subject Authority Data)

FR family: entities

Group 1 Entities	Group 2 Entities	Group 3 Entities
Works Expressions Manifestations Items	Persons Corporate Bodies Families	Group 1 Group 2 Concept Object Event Place

Some of the relationships

- Works   ←————————————→   Persons
- Expressions   ←————————————→   Subjects
- Manifestations   ←————————————→   Corporate bodies
- Items   ←————————————→   Families
- Persons   ←————————————→   Corporate bodies

Some attributes

Enter... RDA

- Resource Description and Access (RDA)
  - New content standard (replacing AACR2 in March 2013)
  - Based on FR-Family

http://www.voalamo.com/wp-content/uploads/2010/01/gumby-11.jpg

RDA

- Based on FRBR so…
  - Attributes – as data elements
  - Relationships are clearly identified

http://kraina-zabaw.pl/wp-content/uploads/2011/12/twister.jpg

One glitch…

… to function as planned, RDA should ideally be marked up in a way that is more **flexible** than what MARC allows

- Welcome to the LC Bibliographic Framework Transition Initiative!

- Well… there is another glitch, too
  - System design – we need to work with vendors to ensure that our systems will realize what we are trying to accomplish

Sharable metadata

Next step will be to break our metadata out of bibliographic records

020   1607507161

ISBN:1607507161<hasAuthor>http://viaf.org/viaf/81847951

100 1   Tran, Duc Thanh,|d1981

ISBN:1607507161<hasTitle>"Process-Oriented Semantic Web search"

245 10 Process-oriented Semantic Web search

ISBN:1607507161<hasPlaceOfPublication>http://viaf.org/viaf/144258029

260   Heidelberg, Germany :|b.AKA,|c2011.

ISBN:1607507161<hasPublisher>"AKA"

300   xii, 221 p. ;|c25 cm.

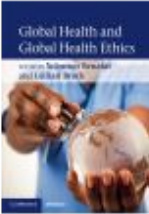ISBN:1607507161<hasSubject>http://id.loc.gov/authorities/subjects/sh2002000569

650 0 Semantic Web.

Web documents

- Have already been linking out to content that is relevant
  - Direct links to content, not to entire documents
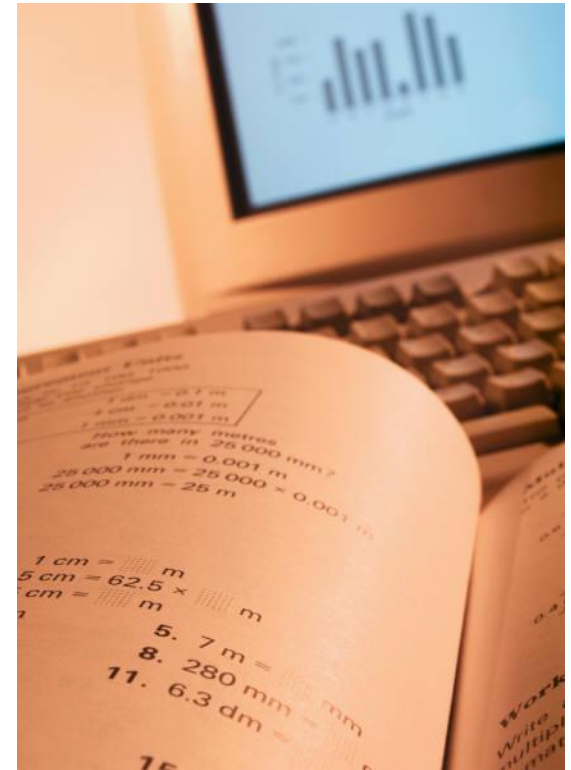

SLU Catalog Record

Cloud Computing will allow libraries to link up and share data

Cloud computing = computing power and storage "out there" on the Internet

More on machines that… read?

- The **Semantic Web** was an idea put forth by Tim Berners-Lee
  - Web 3.0
  - The Intelligent Web
- The primary instantiation of this currently being discussed involves **linked data**
  - Content encoded so that machines know what it is
  - Machines can also act on this data, search for contents, highlight relationships
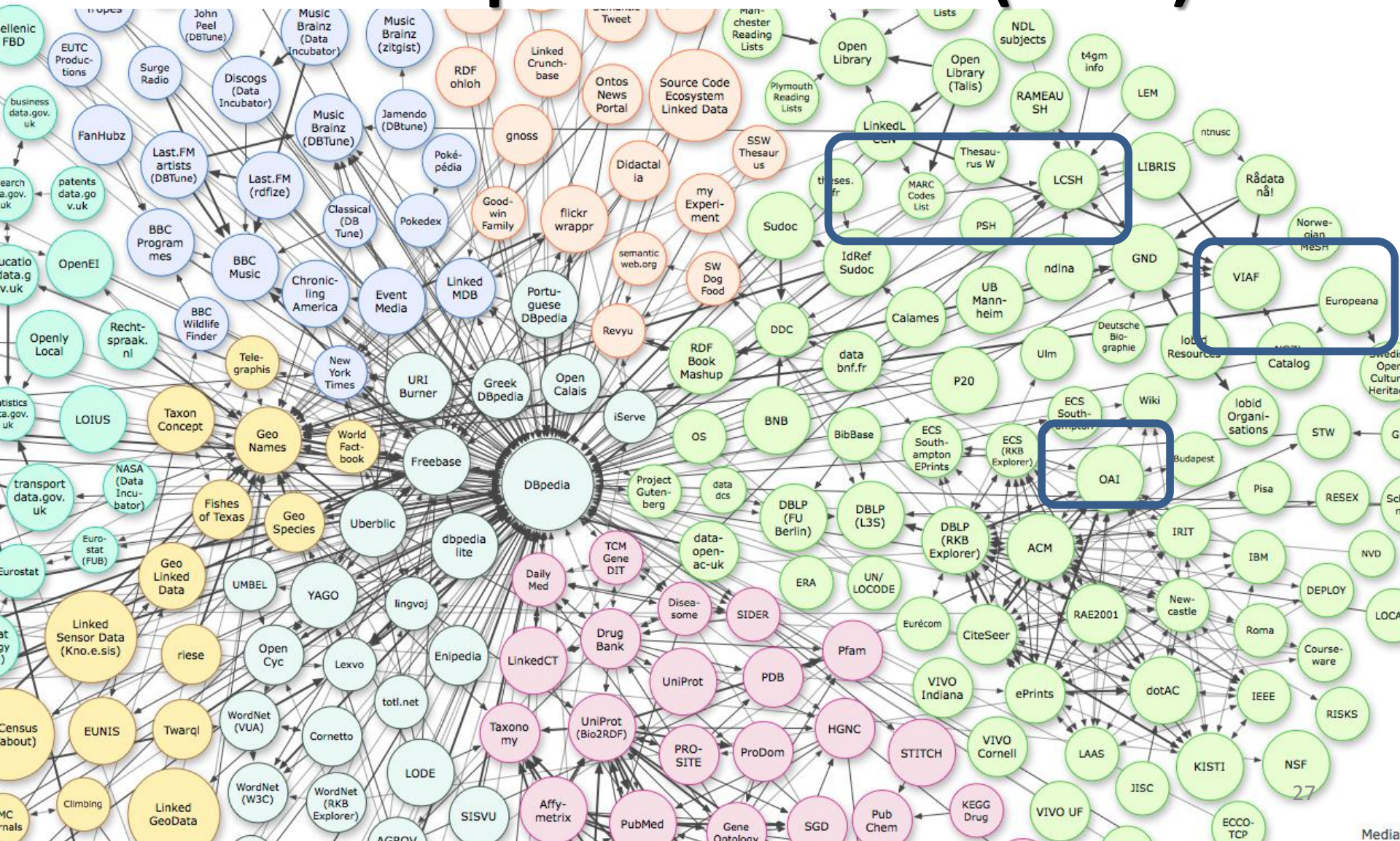
Linked data

- Linked data was also described by Tim Berners-Lee (inventor of the web)
  - Uses special URLs (called URIs) to represent real-world objects
  - Encodes data in a way that is machine actionable and that can be web-friendly (RDF)
  - Semantics can allows for short "sentences" that show relationships
  - When made openly available, linked open data can be:
    - Shared
    - Reused

Repurposed
Integrated

Some library data in the Linked Open Data cloud (2012)

Machine-actionable library data

- Some library databases are already machine-actionable and in the cloud
  - VIAF (Virtual International Authority File) from OCLC
    - http://viaf.org/
  - Library of Congress Authorities and Vocabularies
    - http://id.loc.gov/
  - These are tools for **creating** library records (controlled vocabularies, codes for MARC records), but are not actually the library records
    - The are also not machine-*understandable* once pasted into library records! They remain character strings with underlying controlled vocabularies.

Identifiers for improved linking

Other projects within the library and information community …
- **ISNI (International Standard Name Identifier)**
  - Part of ISO standards; administered by OCLC
  - Will include authors, performers, and publishers
  - Entries will be linked to and from VIAF
- **ORCID (Open Researcher & Contributor ID)**
  - Grew out of Thompson-Reuters' Researcher ID system
  - Run by a not-for-profit organization
  - Includes authors, especially of scholarly books and articles
- **EZID**
  - University of California Curation Center
  - Identifiers for research material, including datasets

Other encoding schema on the web

- The ultimate in web-ready encoding is HTML
  - HTML5 is the current standard for webpages
  - Microdata allows for semantics to be embedded in web documents, encouraging retrieval
- Microformats also encourage retrieval, but are meant to be human-friendly too
  - Use HTML5
  - Are the basis for the emerging search engine standard schema.org

To every silver lining its cloud?

- Replacing MARC and making RDA fully functional will be a challenge
  - Systems will be slow in making the most of new encoding formats
- Legacy (i.e. current) library data will be challenging to update and link
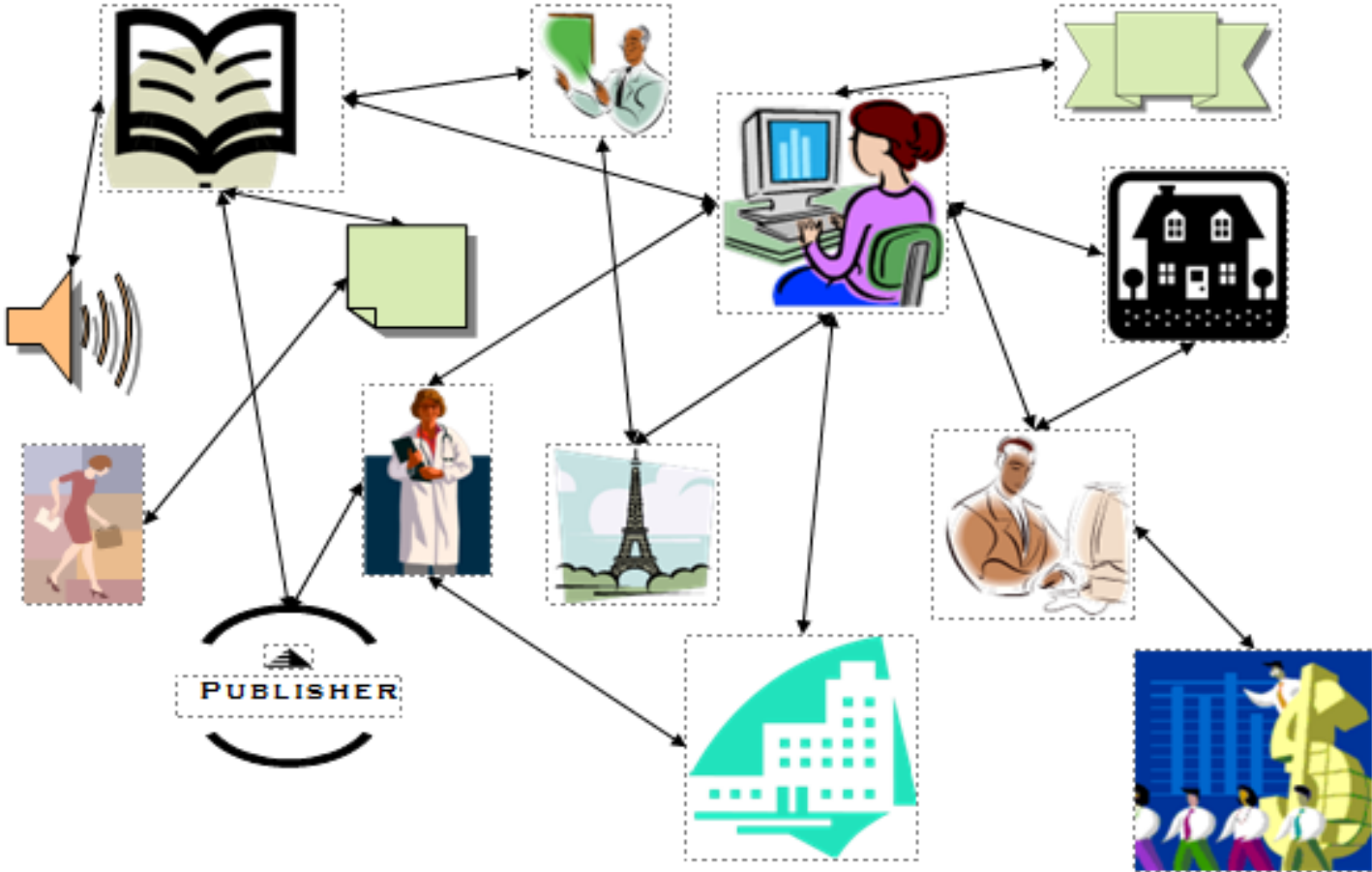
Libraries and Webscale

Providers that deliver Webscale, such as Amazon, Google, eBay and Facebook, have some common attributes. They:

– **Create value based on massive aggregations of data**

– **Connect users** to create large-scale, engaged communities

– Support the community and the data through **scalable, shared infrastructure**   (p. 17)

Big collaboration in the information ecosystem will come not from broader collaboration across libraries, library groups, consortia and cooperatives, but increasingly through **new, innovative alliances and partnerships across the broader knowledge community**—across researchers, publishers, commercial vendors and Webscale providers such as Google, Amazon and Facebook. (p. 31)

(Libraries at Webscale: A Discussion Document. OCLC, 2012.)

Library users will benefit

The future of the ILS is so bright…

- With all of the advances in organization for retrieval and systems, we can hope for future systems that are:
    - Interoperable
    - User-friendly
    - Accurate
    - Promoting a dynamic aggregation of data to meet user needs

THANK YOU!

Questions?
Comments?