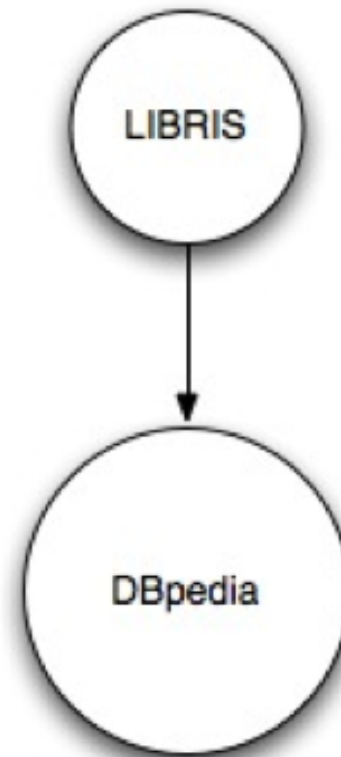# Introducing W3C Library Linked Data Incubator Group (LLD XG)

Marcia Zeng, Jon Phipps, Ed Summers
ASIST 2010, October 22-27, 2010, Pittsburgh, PA.

Based on the PPT prepared by Antoine Isaac
Presented at Talis Open Day: Linked Data and Libraries, London, July 21st 2010
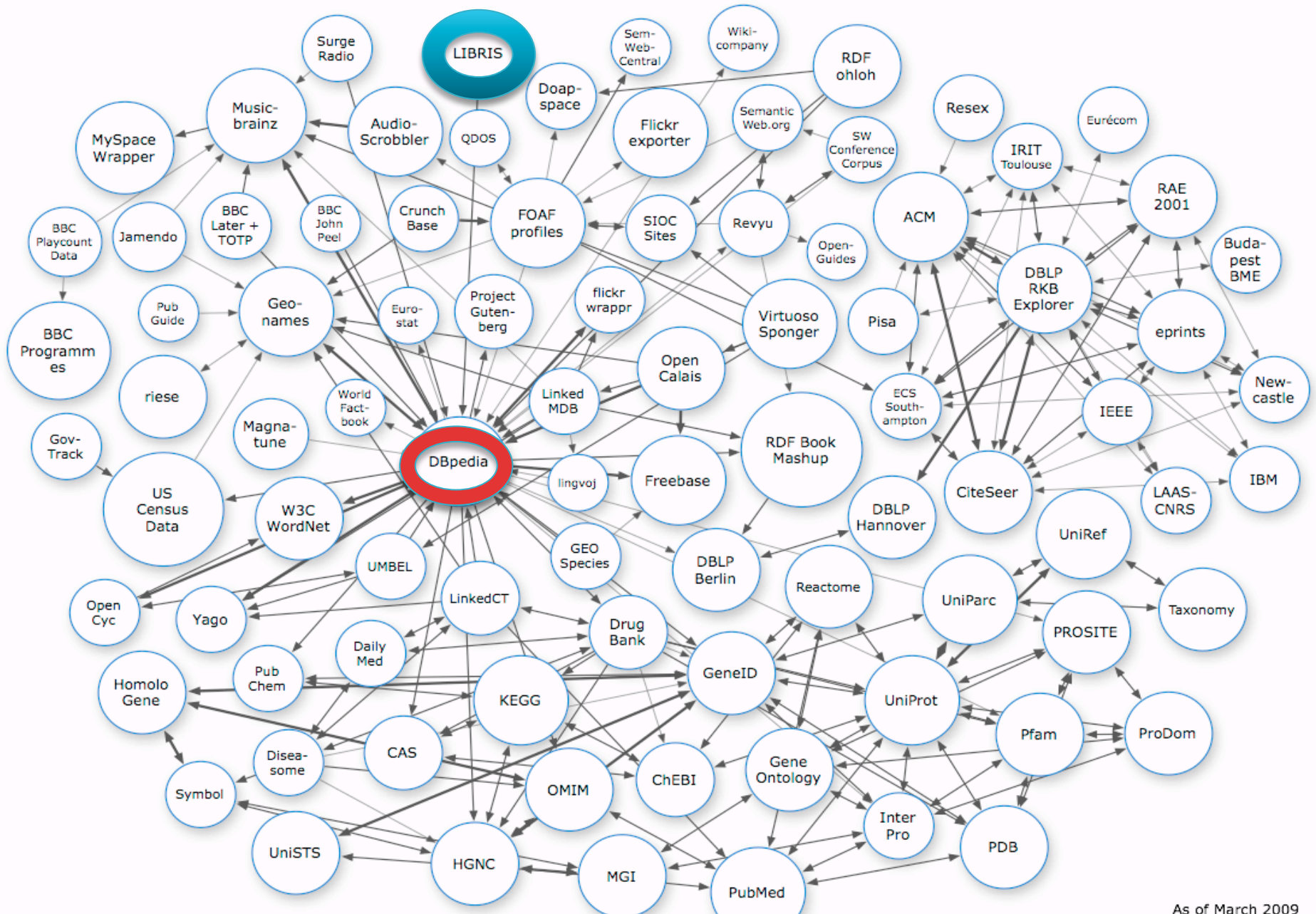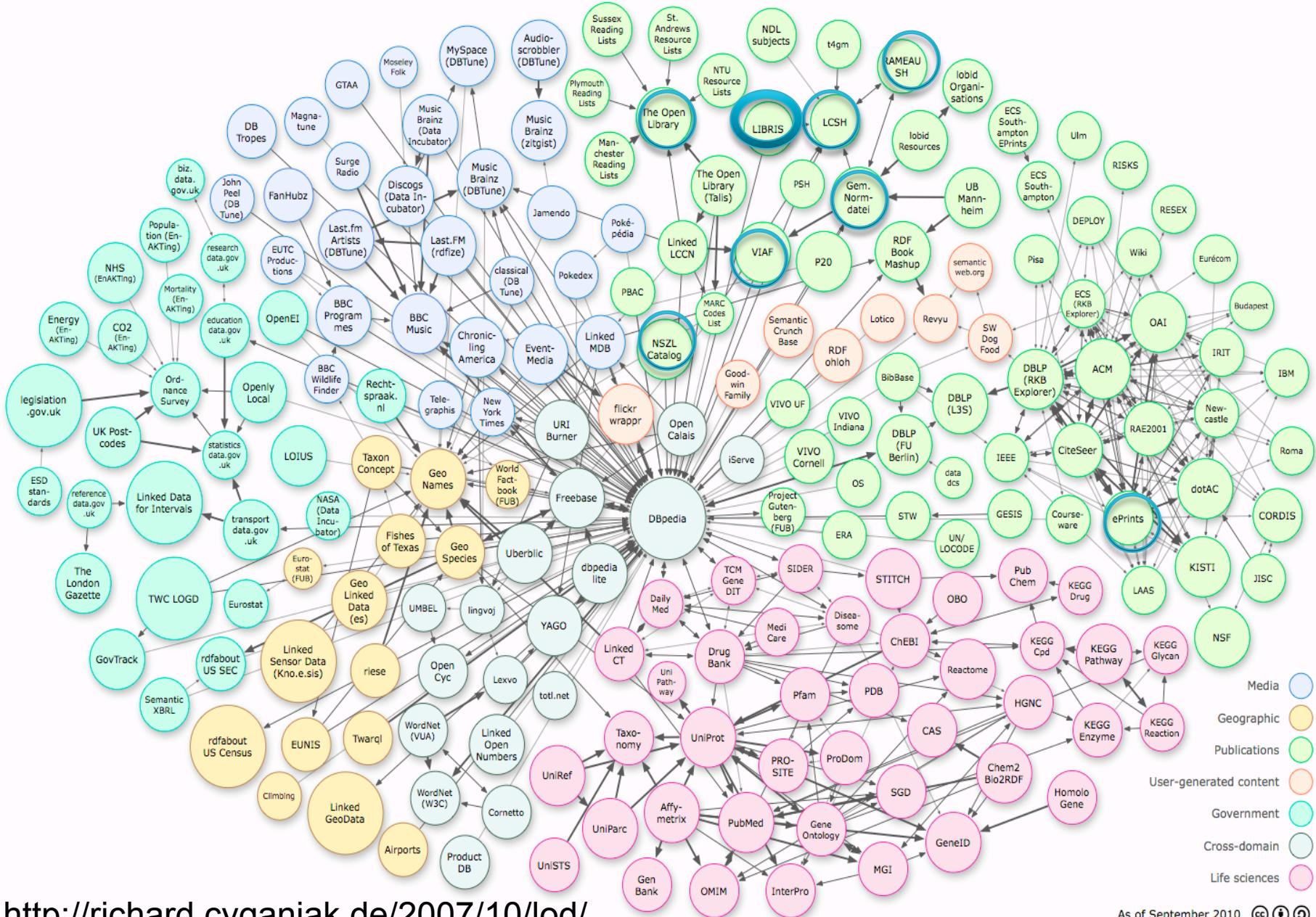
Linked Library Cloud beginning 2008

Linking Open Data -- dataset cloud (as of 2009-03)

Linking Open Data -- dataset cloud (as of 2010-09, dataset: 203)

http://richard.cyganiak.de/2007/10/lod/

Media
Geographic
Publications
User-generated content
Government
Cross-domain
Life sciences

As of September 2010

Landmark Linked Data implementations by library actors

- Swedish National Library – Libris catalogue and thesaurus — libris.kb.se/
- Library of Congress – vocabularies, including LCSH — id.loc.gov/
- German National Library (DNB) – Gemeinsame Normdatei and authority data — d-nb.info/gnd/
- National Library of France (BnF) – RAMEAU subject headings — stitch.cs.vu.nl/rameau/
- OCLC's DDC classification — dewey.info/
- OCLC and LC– VIAF (— viaf.org/
  (The *Virtual International Authority File*)
- National Library of Hungary – catalogue and thesauri — oszkdk.oszk.hu/resource/DRJ/404

Also relevant!

- Agrovoc — aims.fao.org/
- New York Times subject headings — data.nytimes.com/
- Scientific publications — (among others) dblp.rkbexplorer.com/
- … other thesauri

Linked Library Cloud mid-2010
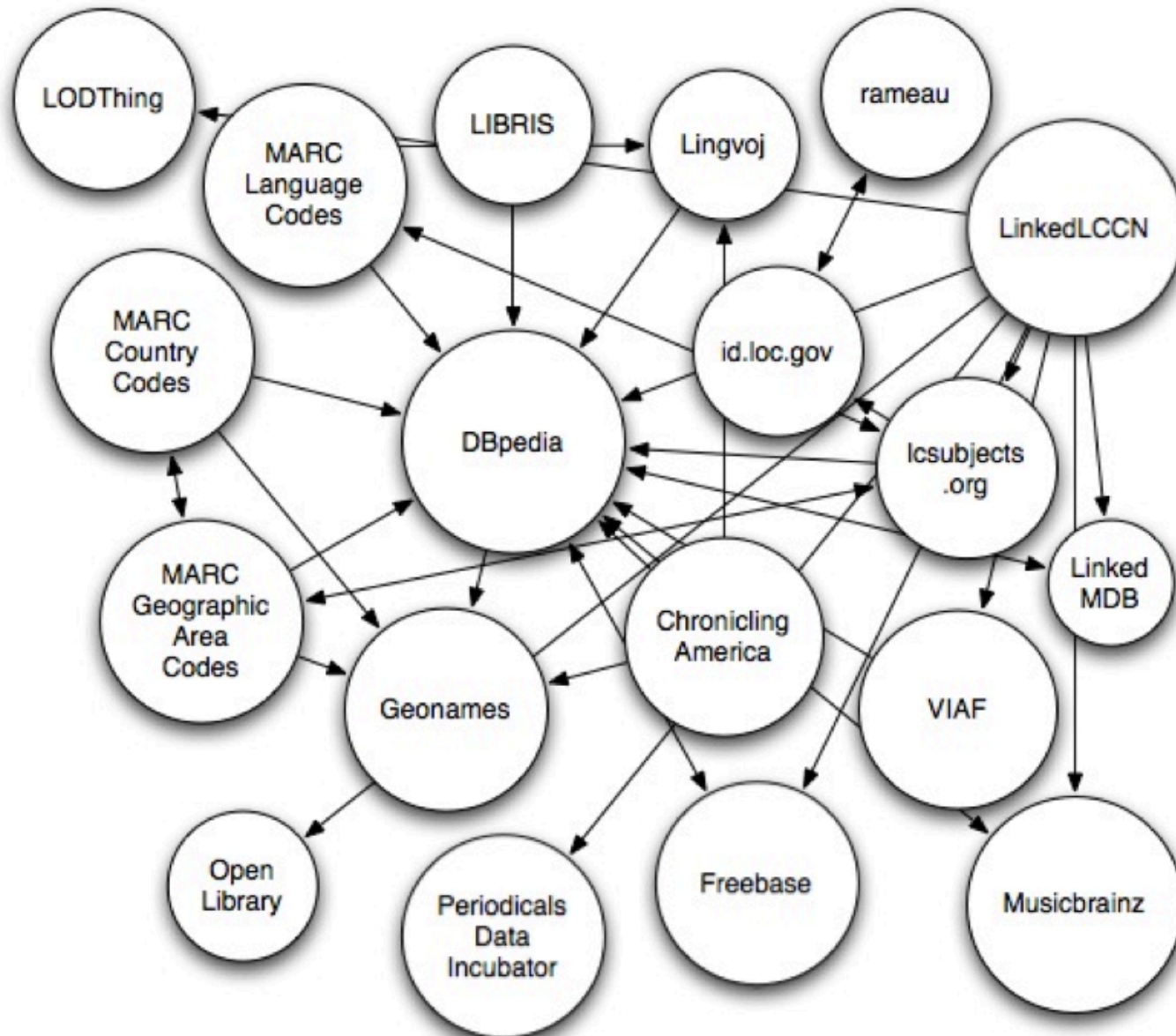
Plus:
- Germany NL
- Hungary NL
- STW
- GEMET
- NYT
- Agrovoc
- …

Useful tooling

Available
- Dublin Core
- SKOS
- BIBO
- OAI-ORE

…

In progress
- RDA vocabularies
- FRBRer@IFLA

dublincore.org/

www.w3.org/2004/02/skos/

bibliontology.com/

www.openarchives.org/ore/

metadataregistry.org/rdabrowse.htm
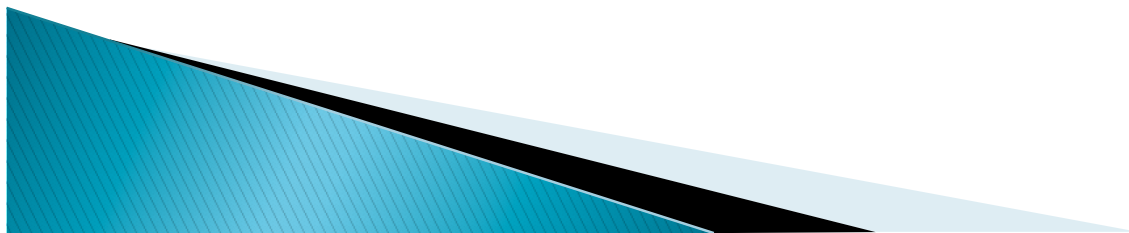
Libraries and LD, the perfect match?

- Libraries have been producing metadata **for ages**
- Libraries (often) produce **high-quality** metadata

- Library metadata is still **locked** in records
- While it does maintain links to the outside world
  - Bibliographic and web references
  - Shared vocabularies
  - Same books!
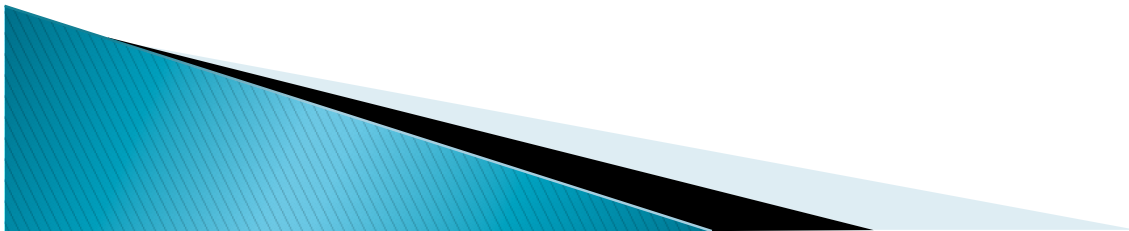
But there are obstacles

Emerging best practices?

▸ What vocabularies are being used, and is there emerging consensus about which to use?

▸ What licenses (if any) are associated with the data?

▸ How much linking and interlinking is going on?

▸ What sorts of mechanisms does the publisher offer for getting the data: sitemap, feeds, SPARQL, bulk download?

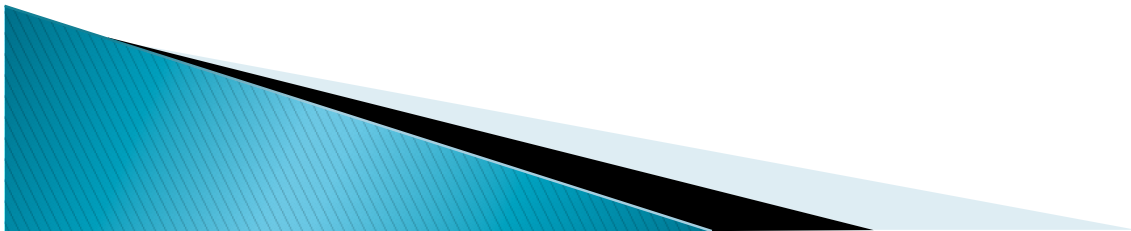▸ What is the quality of the data: granularity, link integrity, vocabulary usage.

▸ What approaches to identifiers for "real world things" have publishers taken: hash, slash, 303, PURLs, reuse of traditional identifiers, etc.

▸ What are the relative sizes of the pools of library linked data?

▸ How are updates being managed?

Ed Summers
http://inkdroid.org/journal/2010/04/18/research-ideas-for-library-linked-data/

W3C incubator (XG) activity

▸ Short-lived working groups: 1 year

▸ Light administration burden

▸ **Not** W3C Recommendations, but "innovative ideas for specifications, guidelines, and applications that are not (or not yet) clear candidates as Web standards"
Deliverables are, but XGs can trigger further W3C work

http://www.w3.org/2005/Incubator/

Library Linked Data Incubator Group (LLD XG)

Initial chairs: Tom Baker, Emmanuelle Bermès, Antoine Isaac

Launched May 21st 2010; weekly teleconference

- 43 participants
- 20 W3C member organizations
- 10+ invited experts

▸ Weekly teleconfrences, F2F Oct. 23, 24, Pittsburgh

Charter

To **help** increase global interoperability of library data on the Web, by

▸ **bringing together people** involved in Semantic Web activities—focusing on Linked Data—in the library community and beyond,

▸ **building on existing initiatives**, and

▸ **identifying collaboration tracks for the future.**

http://www.w3.org/2005/Incubator/lld/charter

Activities

▸ Gathering use cases and case studies demonstrating successful implementation of Semantic Web technologies in libraries and related sectors

▸ Fostering collaboration among actors (libraries, museums, archives, publishers) interested in porting cultural assets to the Linked Data Web

▸ Identifying relevant data models, vocabularies and ontologies and ways to build or improve interoperability among them

▸ Identifying the need for the elaboration of new standards, guidelines & best practices

▸ Identifying the areas of (Semantic) Web technology that could benefit from the expertise of the communities represented in the Group

▸ Proposing a relevant scope and organization for work that follows on the initial effort carried by the Group.

http://www.w3.org/2005/Incubator/lld/charter

Planned deliverables

Report presenting the landscape of Linked data development in the library domain and related sectors, including:

▸ A use-case document that describes a number of real-world use cases, case studies, outreach and dissemination initiatives targeted to the library community and related sectors

▸ A document that describes relevant technology pieces, including vocabularies and ontologies (e.g., SKOS), with the intended goal to identify extension or interoperability requirements, and help determine what other standards may be needed.

http://www.w3.org/2005/Incubator/lld/charter

1. Use Cases/Case Studies collected

▶ 43 use cases collected by Aug. 20
▶ Clusters:
  ◦ Bibliographic data
  ◦ Authority data
  ◦ Vocabulary alignment
  ◦ Archives and heterogeneous data
  ◦ Citations
  ◦ Digital objects
  ◦ Not yet clustered

http://www.w3.org/2005/Incubator/lld/wiki/UseCases

Examples of Use Cases/Case Studies

- Authority Data Enrichment
- Digital Preservation
- Bibliographic Network
- Publishing 20th Century Press Archives
- Subject Search
- Open Library Data
- Europeana
- Language Technology
- National Digital Newspaper Program (NDNP)
- Vocabulary Merging
- Bridging OWL and UML

- Virtual Int. Authority File (VIAF)
- Component Vocabularies
- Digital Text Repository
- Identification and deduplication of lib records
- NEP: New Economics Papers
- Photo museum case
- Browsing and searching in data repositories annotated with different thesauri
- Radio Station Archive
- FAO Authority system and AGROVOC
- … … (43 use cases)

http://www.w3.org/2005/Incubator/lld/wiki/UseCases

2. Use Cases vs. Topics

Conceptual Models
- The FR's: FRBR, FRAD, FRSAR
- Types of library data other than bibliographic and authority
- "Application profiles"

Applying SemWeb Technology to Library Data (Implementation)
- Legacy data
- Vocabularies
- Use of Identifiers for and in LLD

Semantic Web environmental issues
- Identifiers
- Linking across datasets
- The Linked Data paradigm and the Metadata Record paradigm

Data management and distribution
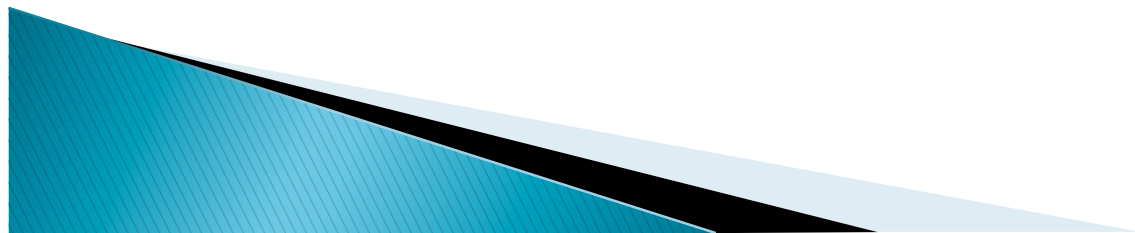- Models for packaging Linked Data
- Provenance

Community and management issues
- Rights, Strategic guidance, Community-building, education and outreach

http://www.w3.org/2005/Incubator/lld/wiki/UseCasesVsTopics
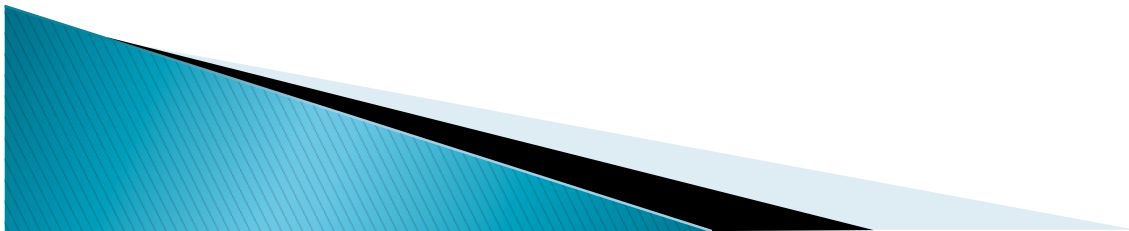
F2F meetings, Pittsburgh, October 2010

1. October 22: Joint meeting LLD XG with DCMI Architecture Forum
   - "Application Profiles for Linked Data: models and requirements"
     - Review of DCMI Abstract Model, brainstorming on requirements
     - Emerging models and new uses for application profiles

2. October 23-24: LLD XG F2F meeting
- Use case discussion
- Vocabularies
- Discussion on deliverables
- Use cases & Topics

- Publicly readable LLD XG wiki

  http://www.w3.org/2005/Incubator/lld/wiki/

- Publicly readable LLD XG mailing list

  http://lists.w3.org/Archives/Public/public-xg-lld/

- LLD community wiki

  http://www.w3.org/2001/sw/wiki/LLD

- LLD community mailing list

  http://lists.w3.org/Archives/Public/public-lld/

- Twitter hashtag
  #lldata