Linked Data, Libraries and the Semantic Web

Gordon Dunsire

Library science talk, Geneva-Bern

12 & 13 March 2012

Overview

❖ Evolution of library linked data

❖ Semantic Web and the Library Linked Data Incubator Group

❖ Current activity in library communities

In the beginning …

Lee, T. B.

Cataloguing has a future. - Audio disc (Spoken word). - Donated by the author.

    1. Metadata

… the catalogue card

From flat-file record ...

Bibliographic description

Author: Lee, T. B.

Title: Cataloguing has a future

Content type: Spoken word

Carrier type: Audio disc

Subject: Metadata

Provenance: Donated by the author

Name authority

Name:

Biography:

...

Subject authority

Term:

Definition:

...

... to relational record

From flat-file description …

Bibliographic description

Author:

Title: Cataloguing has a future

Work Content type: Spoken word

Author Carrier type: Audio disc

Subject:

Subject:

Expression Provenance: Donated by the author

Content type: Spoken word
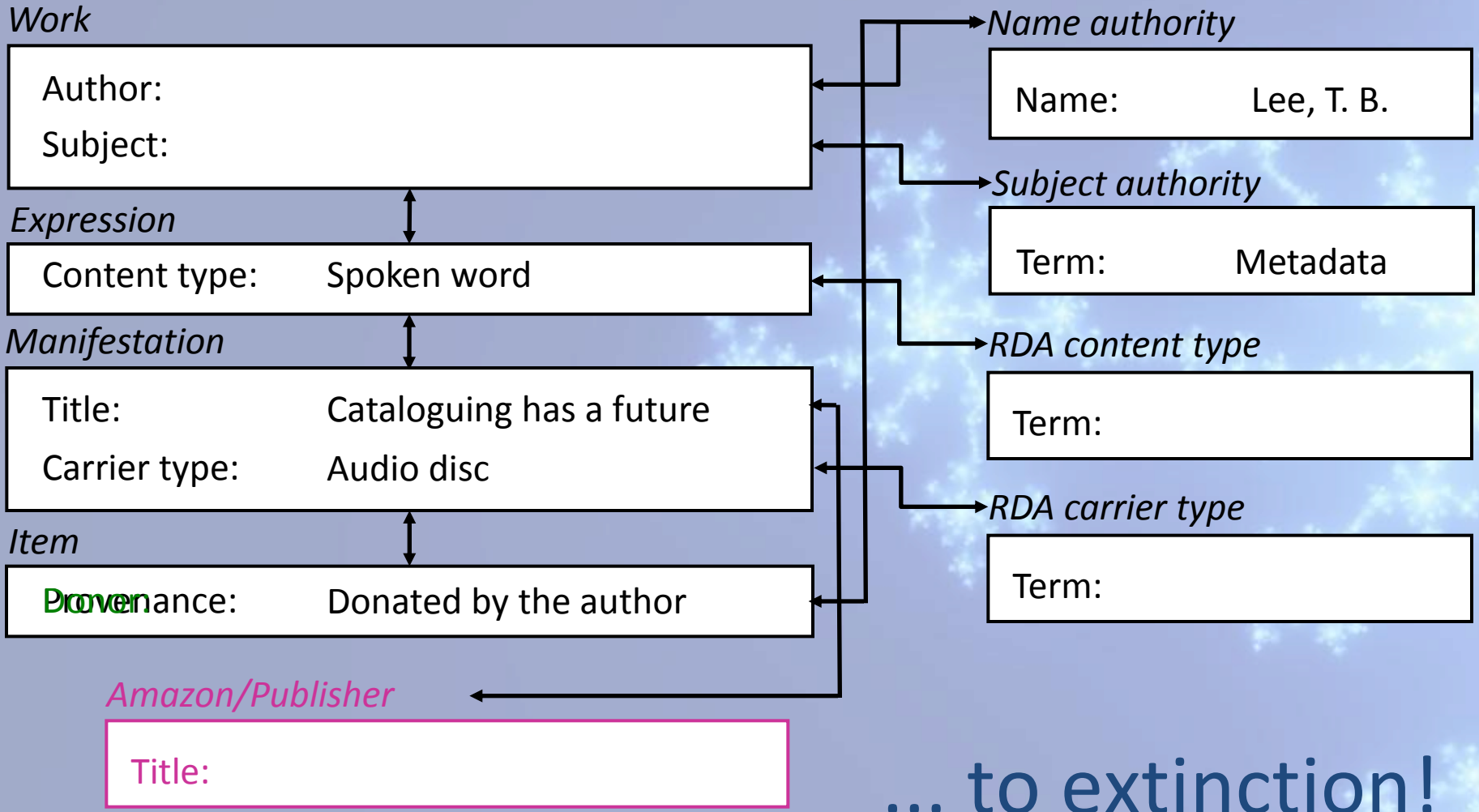
Manifestation

Item

Name authority

Name: Lee, T. B.

Biography:

…

Subject authority

Term: Metadata

Definition:

…

… to FRBR record

From FRBR record …

Work

Author:

Subject:

Expression

Content type: Spoken word

Manifestation

Title: Cataloguing has a future

Carrier type: Audio disc

Item

Provenance: Donated by the author

Amazon/Publisher

Title:

Name authority

Name: Lee, T. B.

Subject authority

Term: Metadata

RDA content type

Term:

RDA carrier type

Term:
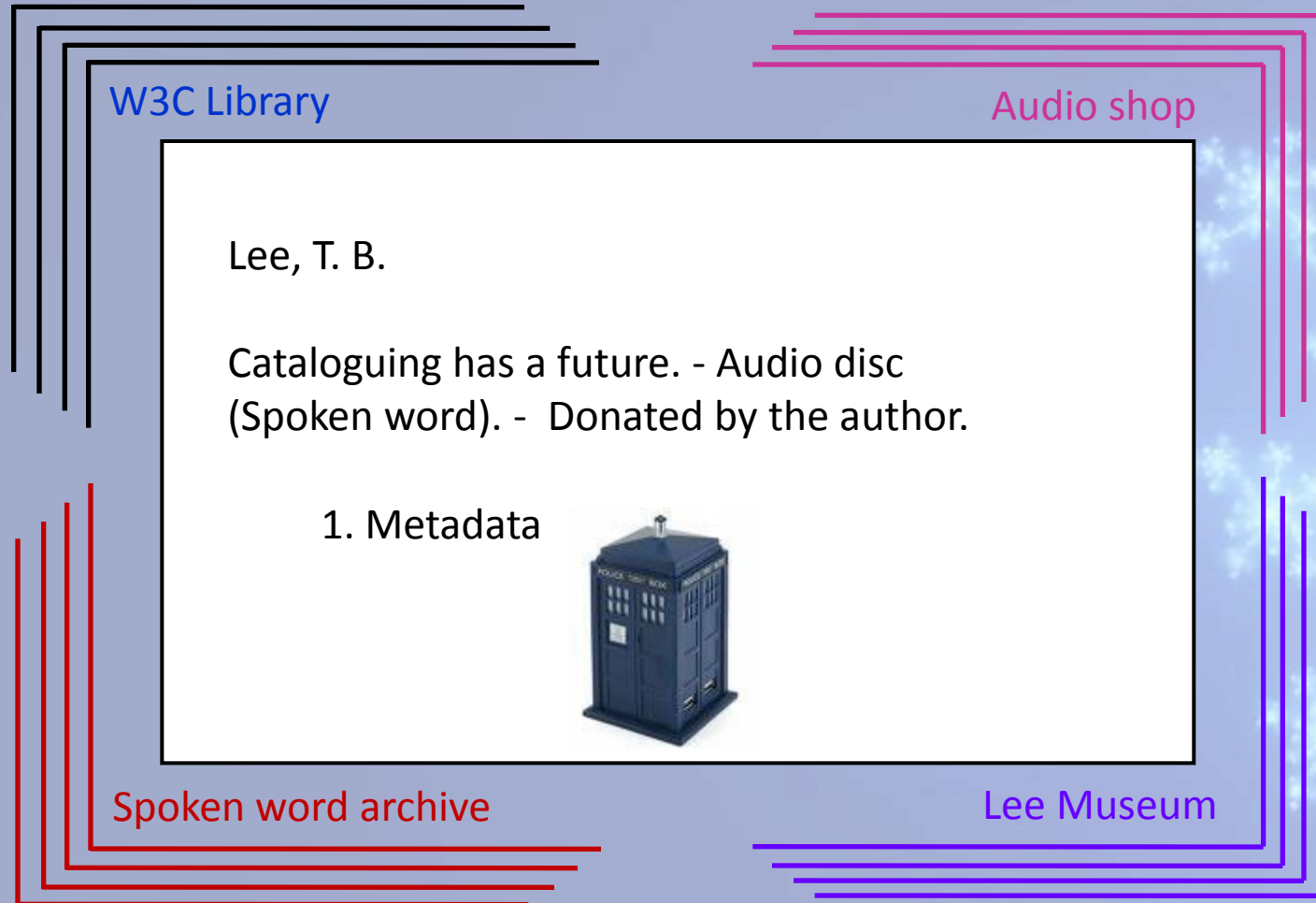
… to extinction!

Where is the record?

❖ Implicit, not explicit

  ❖ Everywhere and nowhere

❖ A semantic Web will allow machines to create the record just-in-time

  ❖ We will not have to maintain records just-in-case

❖ The user will have control over the presentation

  ❖ I want to see an archive or library or museum or Amazon or Google or Flickr or ? display

❖ And by avoiding duplication, we can all get on with describing new stuff …

The hyperdimensional (Tardis) card

W3C Library

Audio shop

Lee, T. B.

Cataloguing has a future. - Audio disc
(Spoken word). - Donated by the author.

1. Metadata
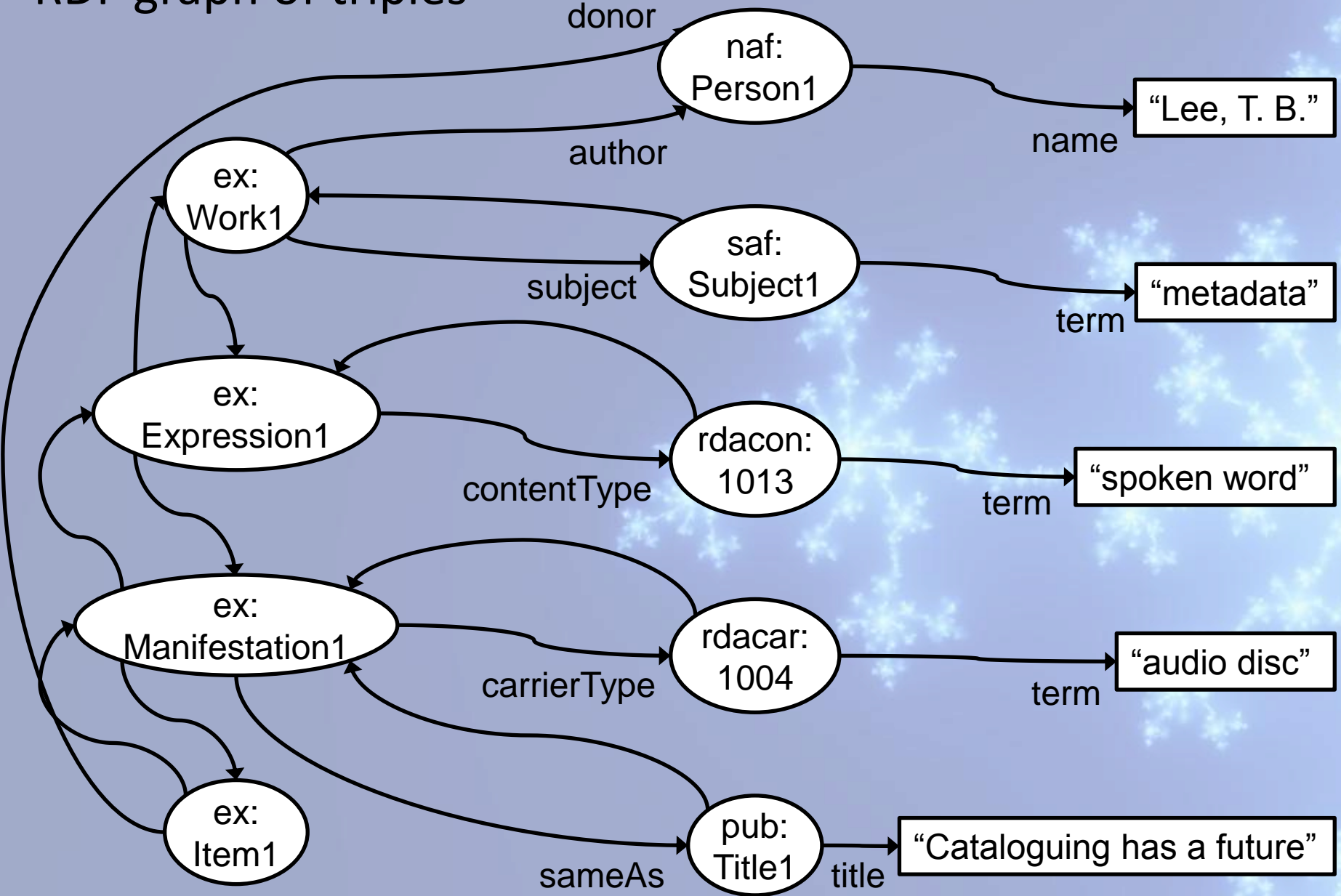

Spoken word archive

Lee Museum

"TARDIS four port USB hub, for office-bound Time Lords:
Open a time vortex on your desk" – Pocket-lint

Semantic Web

- ❖ "provides a common framework that allows **data** to be shared and reused across application, enterprise, and community boundaries."
  - ❖ "a Web of data" – W3C Semantic Web FAQ
- ❖ Uniform Resource Identifiers (URIs)
  - ❖ Globally unique identifiers
- ❖ Resource Description Framework (RDF)
  - ❖ Specifies relationships between things
    - ❖ Used for navigating between, or integrating, information from multiple sources
  - ❖ 3-part statement: thing-relationship-thing = "triple"

RDF graph of triples

Library Linked Data Incubator Group (LLDXG)

❖ **W3C (World-Wide Web Consortium)**
   ❖ "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."

❖ **International group of experts**
   ❖ Active June 2010-August 2011

❖ **Final report October 2011**

Key recommendations: 1

❖ That **library leaders** identify sets of data as possible candidates for early exposure as Linked Data and foster a discussion about Open Data and rights;

Library data (Open license)

❖ Datasets

   ❖ British National Bibliography; Deutsche Nationalbibliografie; etc.

❖ Values

   ❖ Library of Congress Subject Headings; Rameau; Virtual International Authority File; etc.

❖ Metadata elements

   ❖ International Standard Bibliographic Description; Functional Requirements for Bibliographic Records; etc.

❖ Each set of data stored and managed in a "namespace"

Key recommendations: 2

❖That **library standards bodies** increase library participation in Semantic Web standardization, develop library data standards that are compatible with Linked Data, and disseminate best-practice design patterns tailored to library Linked Data;

Participation

- International Federation of Library Associations and Institutions (IFLA)
  - Namespaces Technical Group reporting to Committee on Standards
- Joint Steering Committee for Development of RDA [resource description and access] (JSC)
  - DCMI/RDA Task Group
    - Dublin Core Metadata Initiative
- DCMI Bibliographic Metadata Task Group
  - DCMI Vocabulary Management Community

Key recommendations: 3

❖That **data and systems designers** design enhanced user services based on Linked Data capabilities, create URIs for the items in library datasets, develop policies for managing RDF vocabularies and their URIs, and express library data by re-using or mapping to existing Linked Data vocabularies;

User services

- At a very early stage of development
  - Datasets are experimental
    - URIs not stable
  - Datasets and values may be incomplete
    - Tension between open licenses and income generation
- Initiatives to align and link metadata elements
  - ISBD/RDA project
  - Preliminary analysis of UNIMARC/ISBD
  - Consolidation of Functional Requirements family

Key recommendations: 4

❖ That **librarians and archivists** preserve Linked Data element sets and value vocabularies and apply library experience in curation and long-term preservation to Linked Data datasets.

Preservation

❖ Linked data is forever

  ❖ A triple is immutable: it should not be amended or deleted, but "deprecated"

❖ Identifiers should be permanent

  ❖ Identifiers are the links!

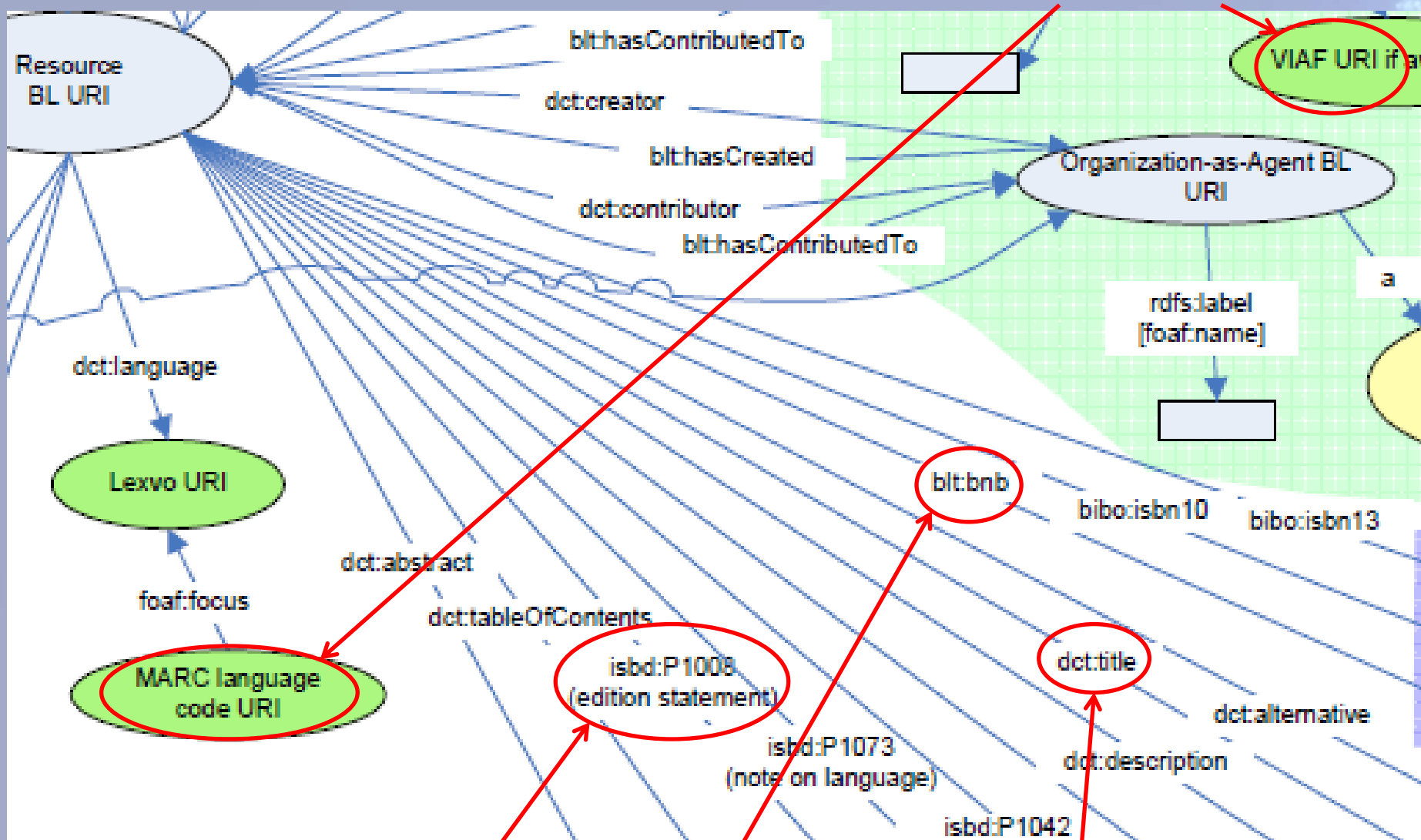❖ Standard labels (or names, titles, headings, etc.) are cultural heritage artefacts

Provenance

❖ Provenance is important
  - ❖ Anyone can say Anything about Any thing (AAA)
  - ❖ No intrinsic test of truth – only inconsistency
  - ❖ "Who said that?"
    - ❖ Competing data from many different sources: social networks, publishers and sellers, governments, propagandists, etc.

❖ Library data generally of higher quality
  - ❖ Ethos of trust, neutrality, etc.

Publishing linked data

❖ Datasets based on legacy records (billions?)

  ❖ MARC and other formats

❖ Can use combinations of attribute and relationship designators in metadata element sets

❖ Linked to identifiers for authority headings from value sets

❖ Example: British National Bibliography

British Library Data Model   Values: MARC VIAF etc.

Elements: ISBD   Local (BL)   Dublin Core   etc.

Questions?

❖gordon@gordondunsire.com

Links (1)

Library Linked Data XG final report:

http://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/

Library Linked Data XG wiki:

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

Deutsche Nationalbibliothek:

https://wiki.d-nb.de/display/LDS/
Dokumentation+des+Linked+Data+Services+der+DNB

British Library:

http://www.bl.uk/bibliographic/datafree.html

Links (2)

IFLA Namespaces Task Group:

http://www.ifla.org/en/node/5353/

DCMI Bibliographic Metadata Task Group:

http://wiki.dublincore.org/index.php/
Bibliographic_Metadata_Task_Group

DCMI Vocabulary Management Community:

http://dublincore.org/groups/vocabulary-management/