The Semantic Web and expert metadata: pull apart then bring together

Presented at 12.seminar Arhivi, Knjižnice, Muzeji

26-28 Nov 2008, Poreč, Croatia

Gordon Dunsire

A problem

- Humans are very good at processing information
  - Creation, analysis, synthesis, communication
    - Some say this is what defines us
- We have invented machines to process data
  - Faster, globally, non-stop
- The result is the information eruption
  - The Web: a continual explosion
- Information professionals cannot keep up
- We need our machines to process metadata

Semantic Web

- "… an evolving extension of the [WWW] in which the semantics of information and services on the web is defined."
  - Wikipedia, English, 19.50 30 Aug 2008
- The basic building block is Resource Description Framework (RDF)

University of Strathclyde

CENTRE FOR DIGITAL LIBRARY RESEARCH

Resource Description Framework (RDF)

- Simple metadata statements in the form of subject-predicate-object expressions, called triples
  - E.g. “This presentation” – “has creator” – “Gordon Dunsire”
- “presentation” and “creator” are metadata structure terms
  - Classes and properties
- “this ...” and “Gordon Dunsire” are metadata content terms
  - Instances or values

Machine-processing

- RDF is about making machine-processable statements, requiring
  - A machine-processable language for representing RDF statements
    - Extensible Markup Language (XML) ✓
  - A system of machine-processable identifiers for resources (subjects, predicates, objects)
    - Uniform Resource Identifier (URI) ✓
  - For full machine-processing, an RDF statement is a set of three URIs

Identifiers

- Things requiring identification (a URI):
  - Subject “This presentation”
    - e.g. its electronic location (URL): http://cdlr.strath.ac.uk/pubs/dunsireg/AKM2008.pps
  - Predicate “has creator”
    - e.g. http://purl.org/dc/terms/creator
  - Object “Gordon Dunsire”
    - e.g. URI of entry in Library of Congress Name Authority File: http://errol.oclc.org/laf/nb2001-72552.html
- Declaring vocabularies/values as “namespaces” in Semantic Web applications provides URIs

University of Strathclyde

CENTRE FOR DIGITAL LIBRARY RESEARCH

Semantic Web applications

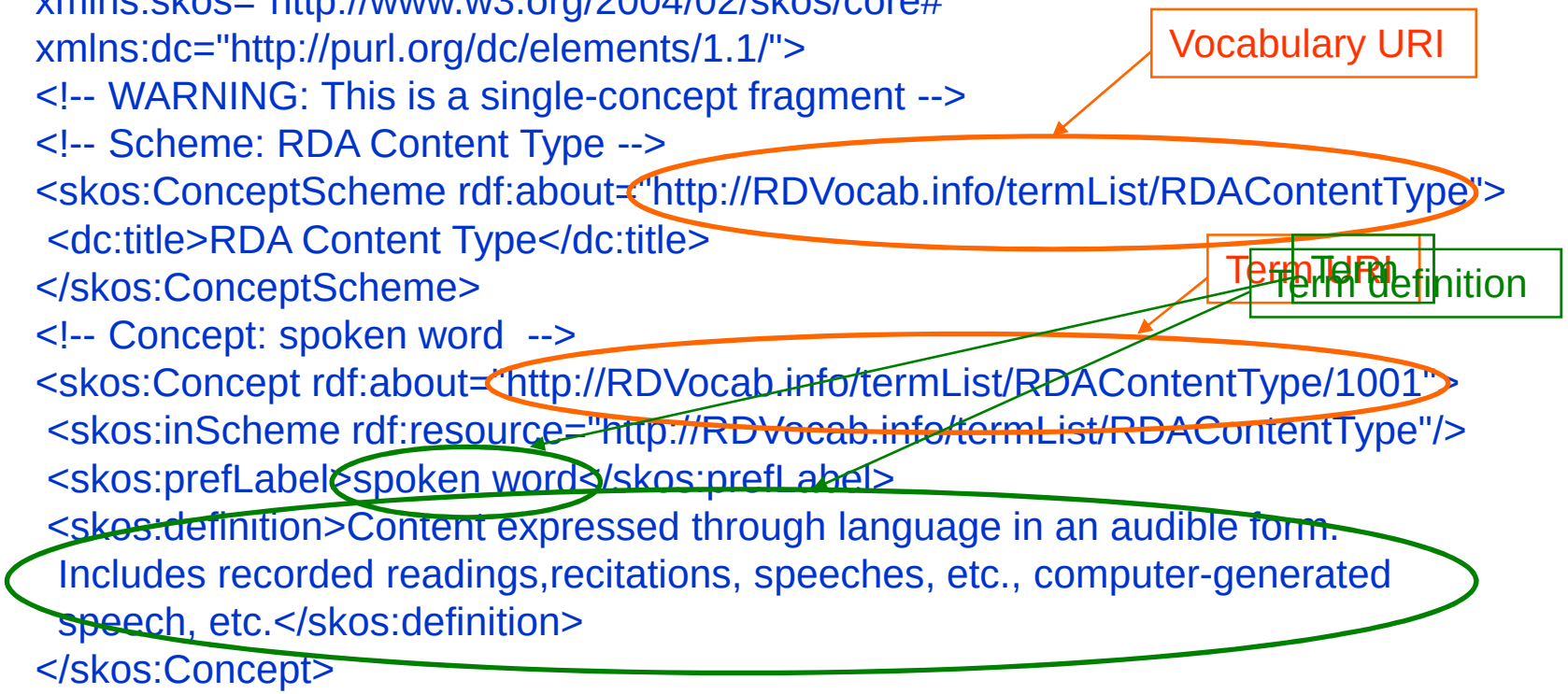
- RDF Schema (RDFS)
  - Expresses the structure of metadata classes and properties
- Simple Knowledge Organization System (SKOS)
  - Expresses the basic structure and content of concept schemes such as thesauri and other types of controlled vocabularies
- Web Ontology Language (OWL)
  - Explicitly represents the meaning of terms in vocabularies and the relationships between them (scope, etc.)

Library namespaces

- Resource Description and Access
  - Successor to Anglo-American Cataloguing Rules
- Two types of vocabularies
  - Metadata entities (elements, attributes)
    - E.g. “Title”, “Content type”
    - Represented as RDF Schema
  - Metadata content (controlled terms)
    - E.g. “spoken word” (instance of content type), “microform” (instance of media type)
    - Represented in SKOS

RDA SKOS example (fake)

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
 xmlns="http://www.w3.org/2004/02/skos/core#"
 xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
 xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
 xmlns:skos="http://www.w3.org/2004/02/skos/core#"
 xmlns:dc="http://purl.org/dc/elements/1.1/">
 <!-- WARNING: This is a single-concept fragment -->
 <!-- Scheme: RDA Content Type -->
 <skos:ConceptScheme rdf:about="http://RDVocab.info/termList/RDAContentType">
  <dc:title>RDA Content Type</dc:title>
 </skos:ConceptScheme>
 <!-- Concept: spoken word  -->
 <skos:Concept rdf:about="http://RDVocab.info/termList/RDAContentType/1001">
  <skos:inScheme rdf:resource="http://RDVocab.info/termList/RDAContentType"/>
  <skos:prefLabel>spoken word</skos:prefLabel>
  <skos:definition>Content expressed through language in an audible form.
   Includes recorded readings,recitations, speeches, etc., computer-generated
   speech, etc.</skos:definition>
 </skos:Concept>
</rdf:RDF>
```



More library namespaces

- IFLA bibliographic control standards
  - Discussions during WLIC 2008, Québec City
- RDF Schema for entities and relationships from Functional Requirements for Bibliographic Records (FRBR)
  - E.g. “Work”, “has Expression” / ”is Expression of”
- Others are likely to follow:
  - Functional Requirements for Authority Data (FRAD)
  - International Standard Bibliographic Description (ISBD)
  - Functional Requirements for Subject Authority Records (FRSAR)
  - UNIMARC
- Library of Congress taking a similar approach with MARC21

University of Strathclyde

A short history of the evolution of the library catalogue record

CENTRE FOR DIGITAL LIBRARY RESEARCH

In the beginning ...

Lee, T. B.

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

1. Metadata

... the catalogue card

University of Strathclyde

CENTRE FOR DIGITAL LIBRARY RESEARCH

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

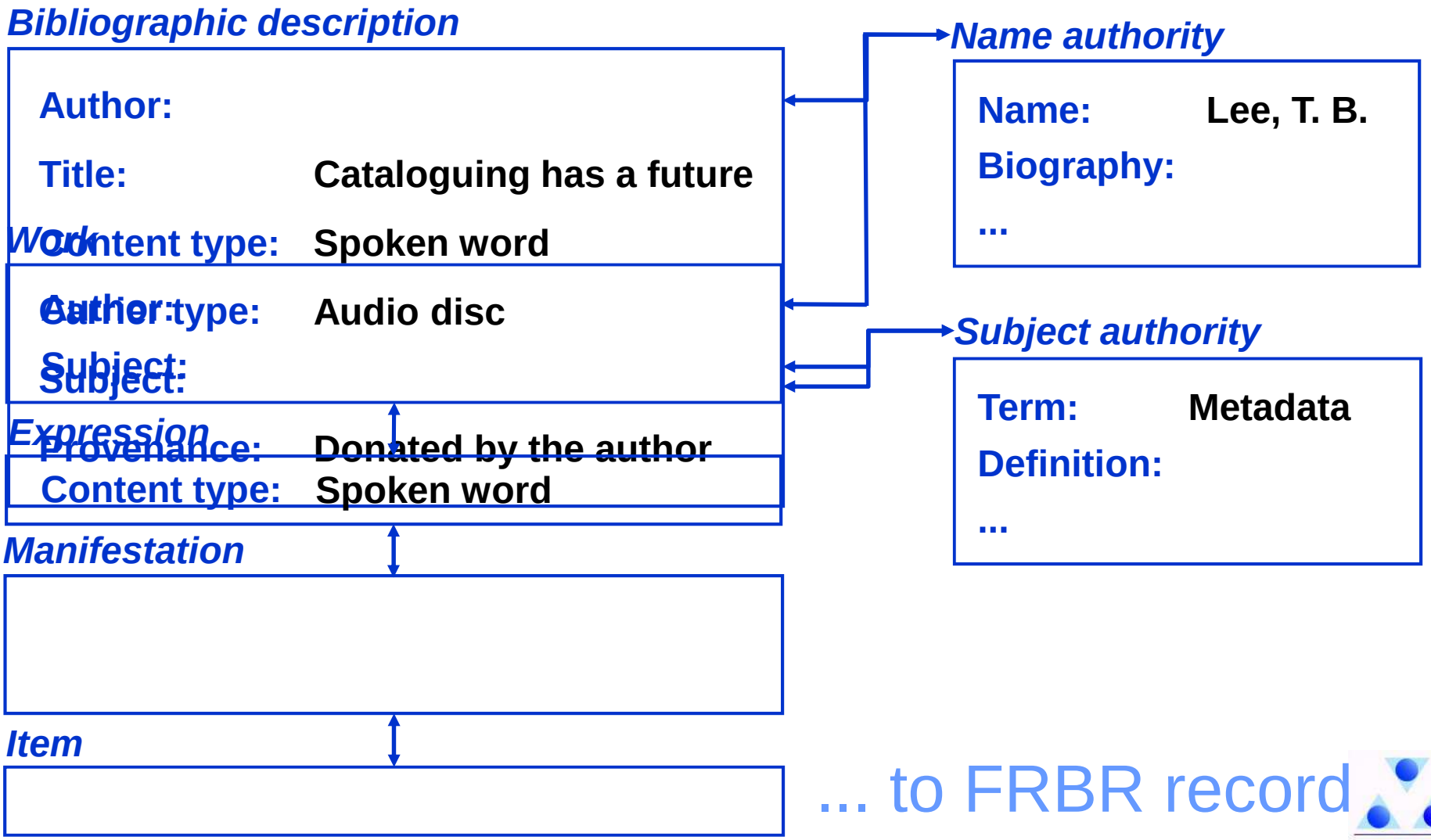
Name:
Biography:
...

Subject authority

Term:
Definition:
...

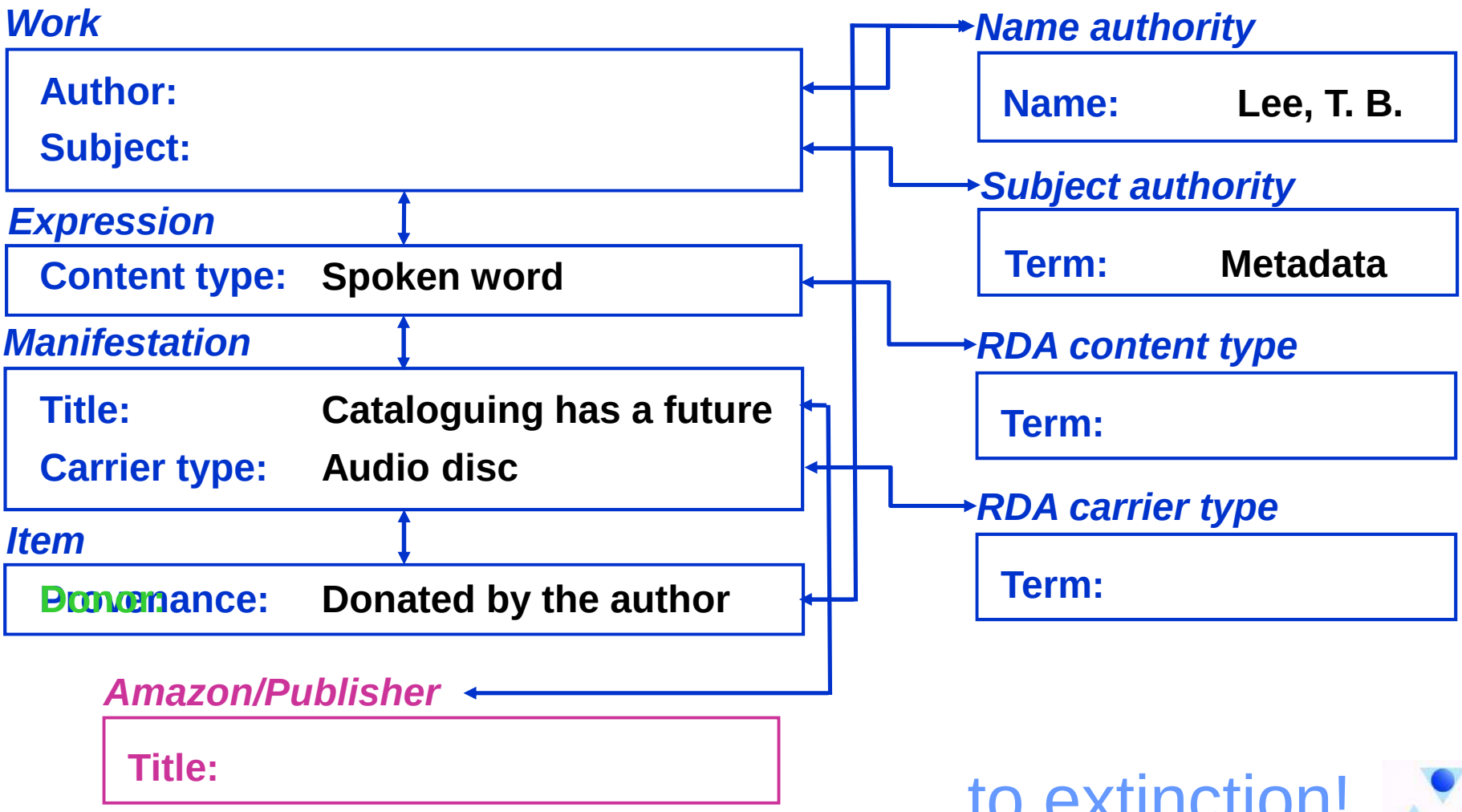
... to relational record

From flat-file description ...



... to FRBR record

From FRBR record ...



... 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 ...

University of Strathclyde

CENTRE FOR DIGITAL LIBRARY RESEARCH

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

University of Strathclyde

CENTRE FOR DIGITAL LIBRARY RESEARCH

Thank you

- Another identifier
  - g.dunsire@strath.ac.uk
  - owl:sameAs
    - http://errol.oclc.org/laf/nb2001-72552.html

University of Strathclyde

CENTRE FOR DIGITAL LIBRARY RESEARCH