University of Strathclyde

RDA vocabularies and concepts

Gordon Dunsire
Depute Director, Centre for Digital Library Research
University of Strathclyde, Glasgow, Scotland

Presented at the IFLA Satellite Meeting on RDA
August 8, 2008, Québec City, Canada

RESOURCE DESCRIPTION AND ACCESS

CENTRE FOR DIGITAL LIBRARY RESEARCH

Overview

- History (when)
- Progress (what)
- Technology (how)
- Future

RDA and ONIX

- ONIX (Online Information Exchange)
  - Publishing industry metadata standard
- 2 day workshop, March 2006, British Library, London
  - RDA Editor, ONIX reps, facilitator
  - Followed up via email and tele-con
- RDA/ONIX framework for resource categorization, August 2006
  - Distinguishes content from carrier (at last!)
- Intention to extend framework
  - Status: Resources permitting

RDA and DCMI

- DCMI (Dublin Core Metadata Initiative)
- 2 day meeting, April/May 2007, British Library, London
  - RDA Editor, reps for RDA, DCMI and related Semantic Web communities
  - Established the DCMI RDA Task Group
  - Operates via wiki, email, tele-con, meetings at DC annual conferences
  - Charter: To define components of the draft standard "RDA - Resource Description and Access" as an RDF vocabulary for use in developing a Dublin Core application profile.
    - Status: Ongoing

RDA and FRBR

- FRBR Review Group, August 2007, WLIC (IFLA), Durban, South Africa
- New project: To define appropriate namespaces for FRBR (entity-relationship) in RDF and other appropriate syntaxes
  - Status: Report and recommendations to be discussed at WLIC, Québec City, Canada (next week)
- FRBR recently extended to Object-oriented FRBR (FRBRoo)
  - Based on CIDOC Conceptual Reference Model (CRM)

RDA/ONIX framework

- An ontology developed by RDA and the publishing community to improve metadata interoperability
- Set of low-level attributes for describing the content and carrier of a bibliographic resource
- Controlled vocabularies for some attributes
- Attributes combined to form high-level content and carrier types for RDA

RDA/ONIX framework example

- RDA content type “spoken word”
  - High-level label for a framework *base content category*
- Base category attributes
  - Character: Language
  - SensoryMode: Hearing
  - ImageDimensionality: not applicable
  - ImageMovement: not applicable
- User: what resources have content I can listen to?
  - = OPAC: what content types have SensoryMode: Hearing?
    - (“Spoken word”; “Performed music”; etc.)
  - then OPAC: list bib records with these content types!

Another framework example

- RDA carrier type “film reel”
  - High-level label for a framework *base carrier category*
- Base category attributes
  - StorageMediumFormat: roll
  - HousingFormat: reel
  - IntermediationTool: projector
- RDA media type “projected”
  - Based on IntermediationTool
    - Therefore technically redundant

RDA vocabularies in RDF

- RDF: Resource description framework
  - World-Wide Web Consortium (W3C) standard
  - Basic building block of the Semantic Web
- Two types of RDA vocabulary in development by DCMI/RDA
  - RDA metadata entities (elements, attributes)
    - E.g. “Title”, “Content type”
    - Represented as an RDF Schema (W3C)
  - RDA value vocabularies (terms)
    - E.g. “spoken word”, “microform” (media type)
    - Represented in Simple Knowledge Organization System (SKOS) (W3C) using RDF

Semantic Web foundations

- RDF
  - Statements about Web resources in the form of subject-predicate-object expressions, called triples
  - E.g. “This presentation” – “has creator” – “Gordon Dunsire”
- RDF Schema
  - Vocabulary description language of RDF
- SKOS
  - Expresses the basic structure and content of concept schemes such as thesauri and other types of controlled vocabularies
  - An RDF application
- OWL (Web Ontology Language)
  - Explicitly represents the meaning of terms in vocabularies and the relationships between them

Semantic Web building blocks

- Each component of an RDF statement (triple) is a “resource”
- 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:
  - Object “This presentation”
    - e.g. its electronic location (URL):
    - http://cdlr.strath.ac.uk/pubs/dunsireg/QuebecRDA.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 (real soon now?)
- Declaring vocabularies/values in SKOS and OWL provides URIs
  - Without such identifiers, the Web will never become Semantic

RDA RDF vocabularies

- Being added to the National Science Digital Library metadata registry
  - Stored in a database
  - Output as RDF(S)/SKOS
  - Automatic creation of a URI for each entry
- Base domain: http://RDVocab.info
  - First part of every RDA vocabulary URI
  - Identifies the “namespace” or collection/set of terms

DCMI/RDA progress

- 246 RDA elements
  - http://metadataregistry.org/schema/show/id/1.html
- Carrier vocabularies
  - Media type (8 entries)
  - Font size (2)
  - Reduction ratio (5)
  - Etc.
- All provisional
  - Awaiting final draft of RDA for terms, definitions, scope, etc.

RDA RDF vocabulary 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>
```

Namespaces used to declare the RDA namespace – everything must be defined explicitly to the machine!

Overall base domain

Vocabulary URI

TermURI

Term definition

RDA content type “spoken word”

The term “spoken word” can be referenced as the value of the field “content type” in any metadata record using RDF/XML (Semantic Web):

…

xmlns:rdvct =
http://RDAVocab.info/termList/RDAContentype”

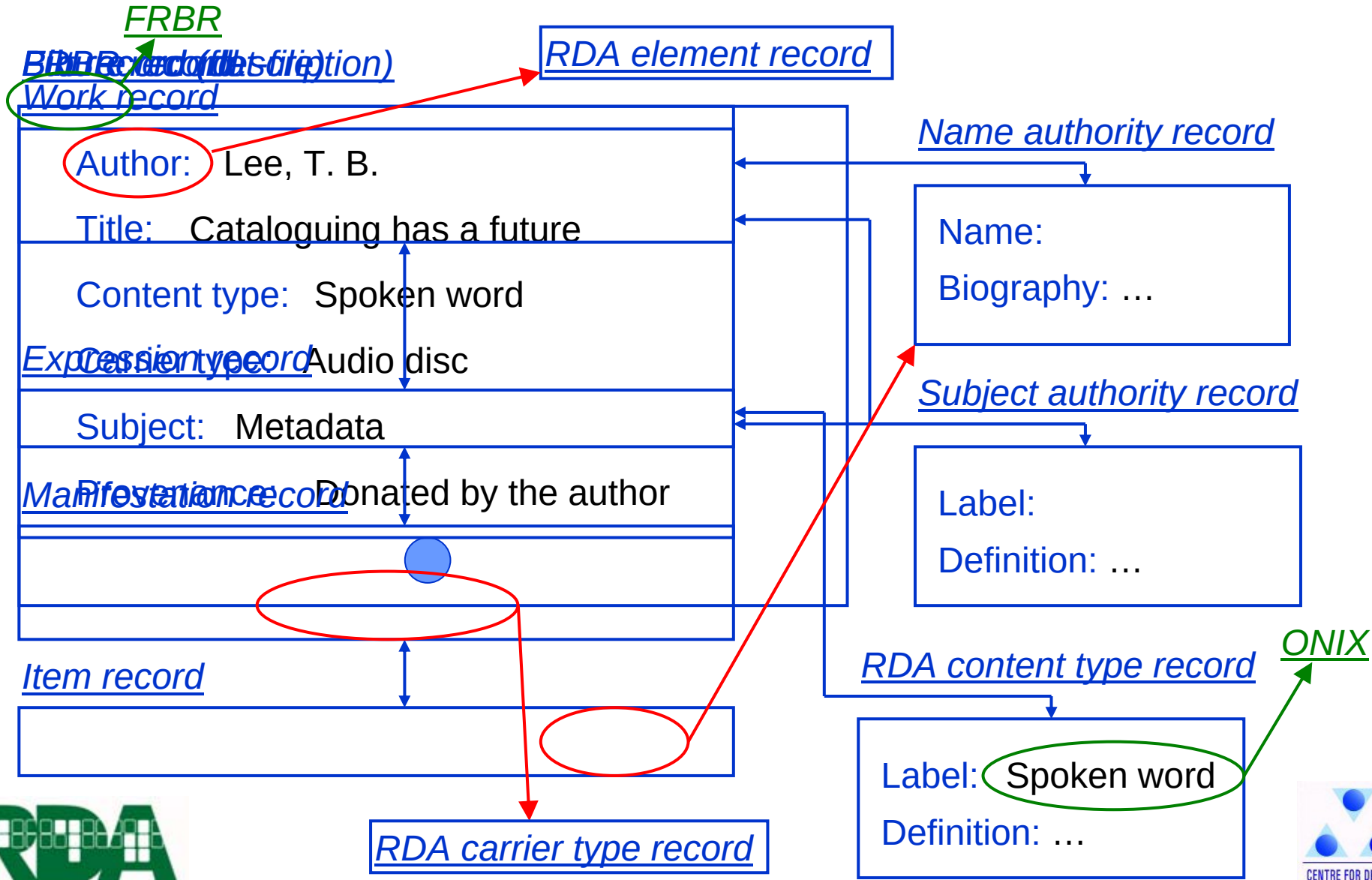
…

<… rdvct:1001 …>

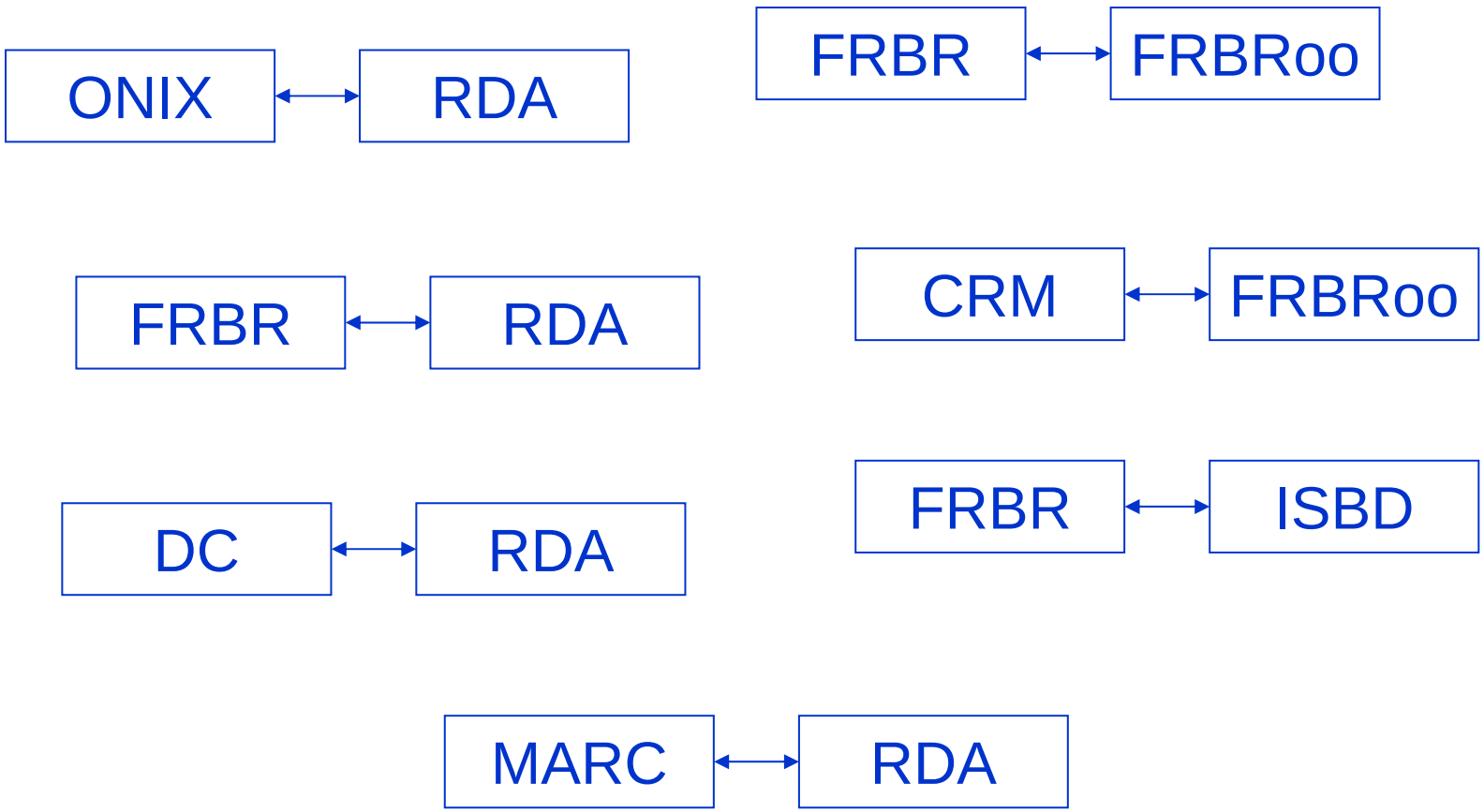
…

The field/attribute/element “content type” can be referenced in a similar way to the RDF Schema for RDA elements being developed by DCMI/RDA

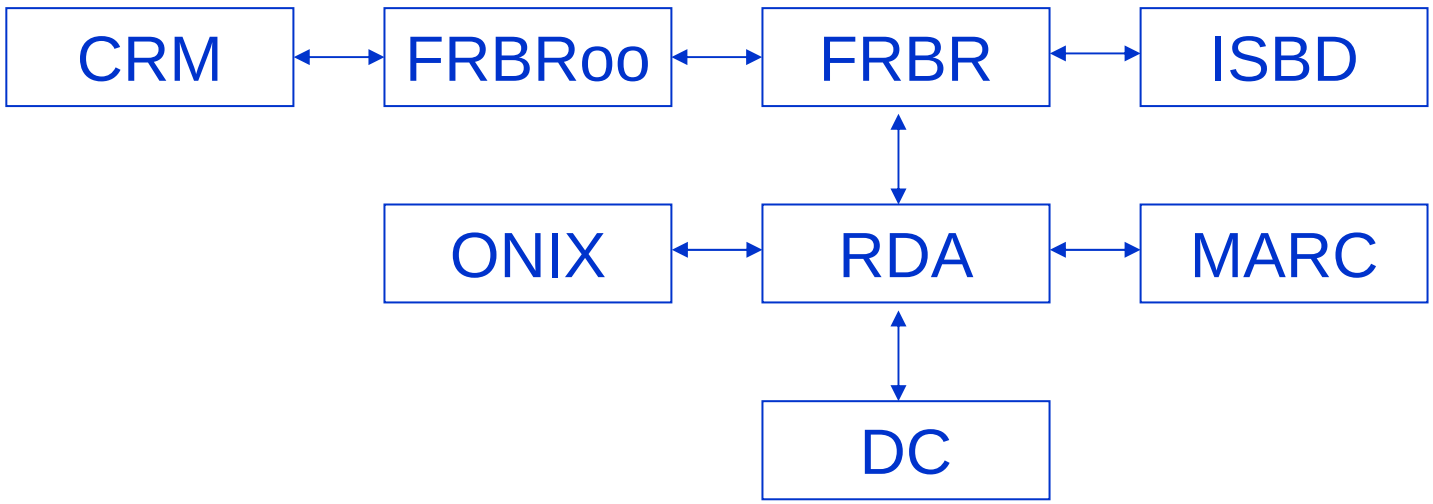
Database/format scenarios



Linking communities



Everything is connected



… at the community (human) and technical (Semantic Web) levels

Thank you

- Another identifier:
  - g.dunsire@strath.ac.uk
- See the handout for acronyms and links …