

RDA data and context

Gordon Dunsire

Presented at "Desarrollo de RDA y casos de
implementación en Europa"

National Library of Spain, Madrid, 27 Oct 2017



The digital deluge

Libraries are facing a flood of digital resources:

- "Born digital" publications
- Digitization of publishers' back catalogues
- Mass digitization of library collections
- Self-published resources:
 - Blogs
 - Email
 - Twitter



Costs of cataloguing

Libraries must acquire digital resources 😊:

- Electronic legal deposit
- National bibliographies
- Special collections
- Digitization for access and preservation

But there is no extra money for professional metadata creation 😞

Just use Google!*?**



Sources of metadata

Professional cataloguers

Smart, standard, and expensive

Amateur cataloguers (crowd-sourcing)

Smart, non-standard, and voluntary

Publishers

Smart, non-standard, and reluctant

They have costs too

Machines

Dumb, cheap, and slaves

24/7, no food, shelter, etc.



The metadata maelstrom

Libraries must obtain metadata from a variety of sources

Get it when you can; take it when offered

Metadata can be standard, non-standard, smart, dumb, incomplete, contextual, and non-contextual
Created for non-library applications

But it must be processed for coherent and consistent information retrieval applications



(Meta)Data values

Data value = "descriptive string"

⇒ Attribute element

Data value = identified thing

⇒ Relationship element (local)

Linked data value = URI

⇒ Relationship element (global)

IFLA Library Reference Model has string/thing duality



RDA database scenarios

Flat-file data

Catalogue cards, printed bibliographies, etc.

Description and access

Bibliographic and authority records

Relational databases

Linked data in the Semantic Web



RDA development contexts (2-3 years)

Governance transition to international representation

Strategic communities: international, cultural heritage, linked data

IFLA Library Reference Model (LRM) and other international standards (IFLA, ISSN, ...)

Toolkit review and re-organization (3R Project)



RDA data

“RDA is a package of data elements, guidelines, and instructions for creating **library and cultural heritage** resource metadata that are well-formed according to **international models** for **user-focussed linked data** applications.”

RDA Toolkit provides the user-focussed elements, guidelines, and instructions.

RDA Registry provides the infrastructure for well-formed, linked, RDA data applications. Open Metadata Registry (OMR) provides linked data representation of RDA elements.



RDA data recording methods

Unstructured description

Structured description

Identifier

Internationalized Resource Identifier (IRI-URI)



Unstructured description

Transcribed data: what you see is what you get

Free text notes

Uncontrolled terminology for description

Non-standard metadata from non-professional sources or lacking provenance (authority)



Structured description

Data recorded in regular, standard, structured formats for human consumers

Sources of data have "authority"

Authorized and variant access points, or controlled terminology

Data from authority files, vocabulary encoding schemes, and knowledge organization systems



Identifiers

Coded labels intended for machine identification

Identification and disambiguation within a local domain

Authority control numbers, standard identifier schemes, machine-readable database keys, terminology and vocabulary notations



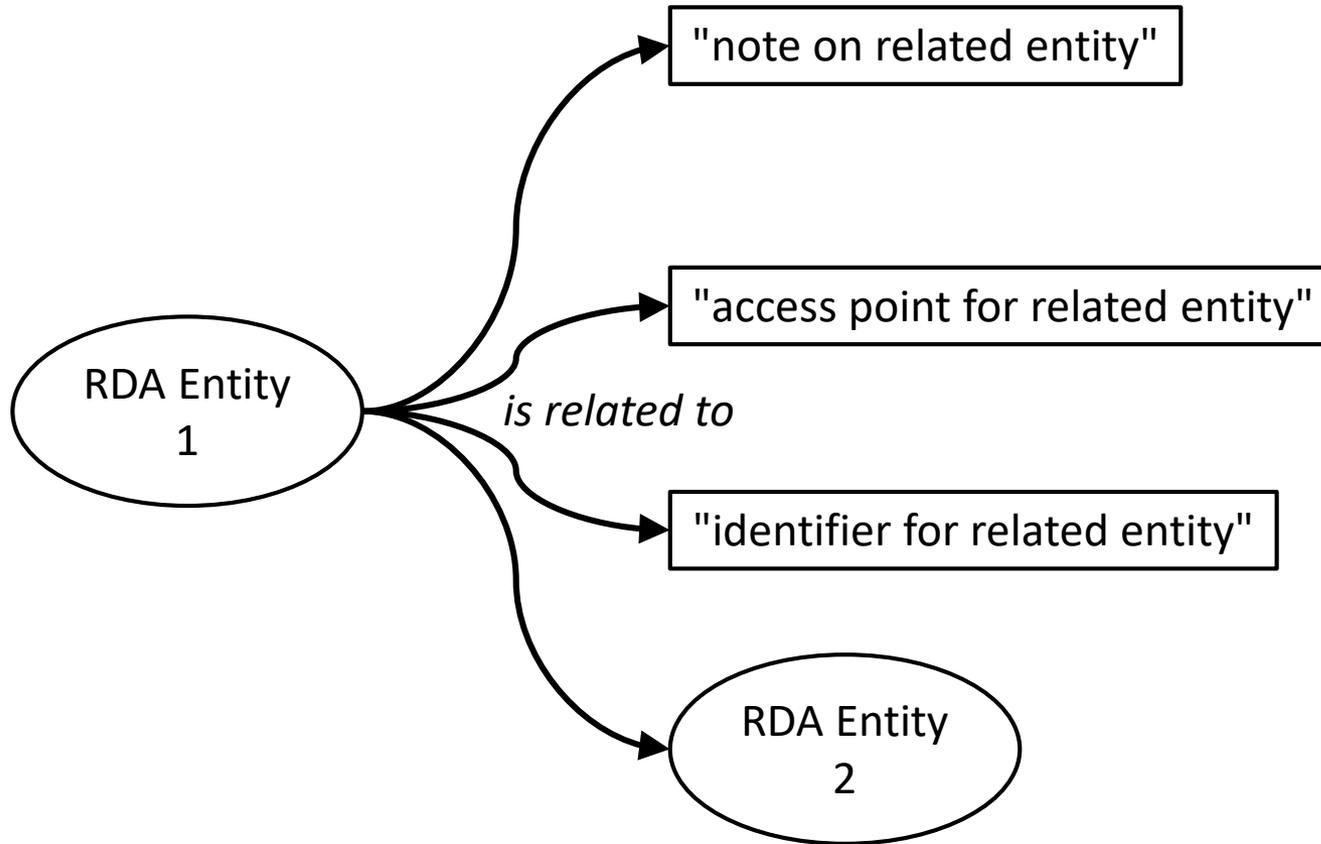
IRI/URI

Internationalized Resource Identifier > Uniform Resource Identifier

Identification within a global domain: the Semantic Web of linked data

Data for "smart" machine applications

Recording methods for related data





RDA as data integration framework

An RDA data element can (usually) accommodate all four types of data value

Multiple types can be assigned to a single element

Metadata from multiple sources can be used in integrated catalogues and other applications



Getting the best out of data

Unstructured: keyword extraction and indexing

Structured: authority and terminology control

Identifier: relational database applications

IRI: Semantic web and linked open data

RDA is for a smart future, not a dumb past

Thank you!