RDA data capture and storage

Gordon Dunsire

Chair, RDA Steering Committee

Presented to Committee on Cataloging: Description and Access II (CC:DA) - ALCTS CaMMS

ALA Midwinter 2016, 11 January 2016, Boston, Mass.

RDA

Overview

- RDA for data management: a continuous process of development
- From here to the future, and what is on the way

R | D | A

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.

RDA

Recording relationships in RDA

RDA offers a choice of techniques for recording relationships between entities. The number of options varies depending on the type of entity:

- 4 techniques for relationships between works, expressions, manifestations, and items
- 3 techniques for primary relationships between works, expressions, manifestations, and items
- 2 techniques for relationships between persons, families, and corporate bodies.

RDA

The 4-fold path

a: Identifier
b: AAP
c: Description

c1: Structured
c2: Unstructured

Recording Relationships between Works, Expressions, Manifestations, and Items

Record the relationship between a work, expression, manifestation, or item and a related work, expression, manifestation, or item by using one or more of these techniques, as applicable:

a) identifier for the related work, expression, manifestation, or item (see 24.4.1 RDA)

b) authorized access point representing the related work or expression (see 24.4.2 RDA)

and/or

c) description of the related work, expression, manifestation, or item (see 24.4.3 RDA).

Record an appropriate relationship desig
the relationship (see 24.5 RDA).

b:Excludes manifestation and item

Description of the Related Work, Expression, Manifestation, or Item D-A-CH

Provide a description of the related work, expression, manifestation, or item by using either a structured or an unstructured description, as appropriate:

RDA

The 3-fold path

a: Identifier
b: AAP
c: Description

c1: Structured

semi-structured?

Techniques Used to Record Primary Relationships

Record primary relationships by using one or more of these techniques, as applicable:

a) identifier for the work, expression, manifestation, or item (see 17.4.2.1 RDA)

b) authorized access point representing the work or expression (see 17.4.2.2 RDA)

c) composite description (see 17.4.2.3 RDA).

Composite Description

Provide a composite description that combines one or more elements identifying the work and/or expression embodied in a manifestation with the description of that manifestation.

RDA

The 2-fold path

Recording Relationships between Persons, Families, and Corporate Bodies

Record the relationship between a person, family, or corporate body, and a related person, family, or corporate body by using one or both of these techniques:

a) identifier (see 29.4.1 RDA)

and/or

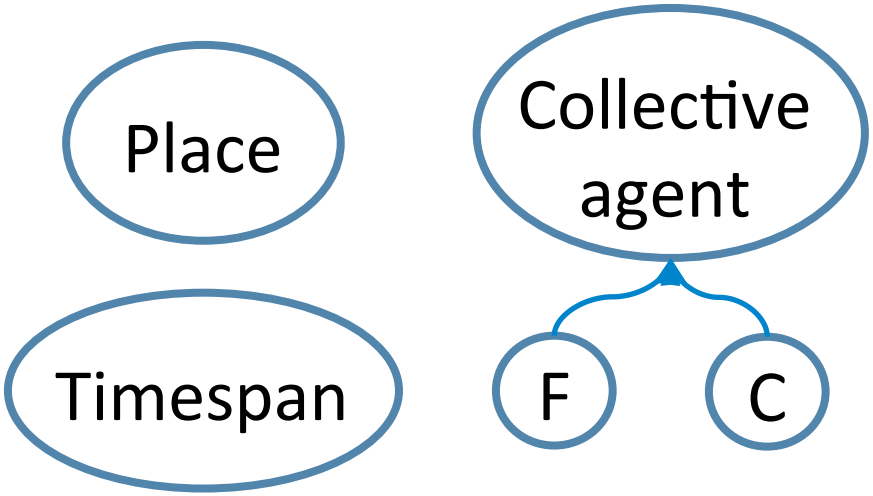
b) authorized access point (see 29.4.2 RDA).

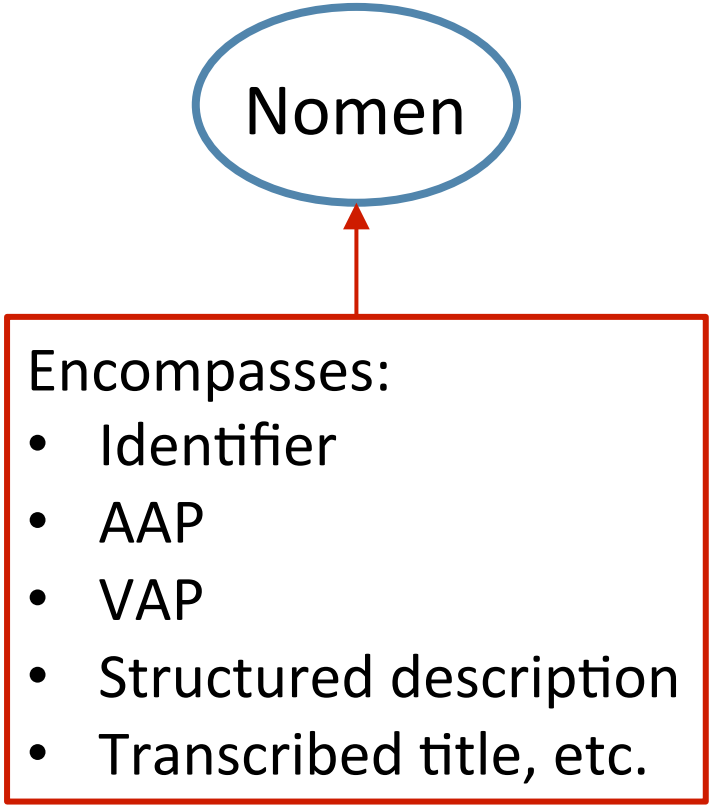
Record an appropriate relationship designator to specify the nature of the relationship (see 29.5 RDA).

a: Identifier
b: AAP

RDA

New FRBR-LRM entities





What techniques will apply to new RDA entities?

a: Identifier
b: AAP

c1: Structured?
c2: Unstructured?

RDA

Structured description

A full or partial description of the related resource using the same data that would be recorded in RDA elements for a description of that related resource presented in an order specified by a recognized display standard.

[Example: ISBD display pattern]
Title proper : other title information **/** statement of responsibility

How full?
A complete ISBD record with all of the data?

RDA : an introduction **/** by J. Smith

Title proper: “RDA”
Other title information: “an introduction”
Statement of responsibility: “by J. Smith”

RDA

Database implementation scenarios

0: Linked data	Fully linked (global)
1: Relational or object database	Fully linked (local)
2: Bibliographic and authority records	AAP/Identifier linked
3: Flat-file	Not linked

RDA

Techniques for obtaining data

Categorization of elements?

Recorded elements	
Sources	any (authoritative, recognized, etc.)
Tasks	all (Find, Identify, Select, Obtain, Explore)
Entities	all

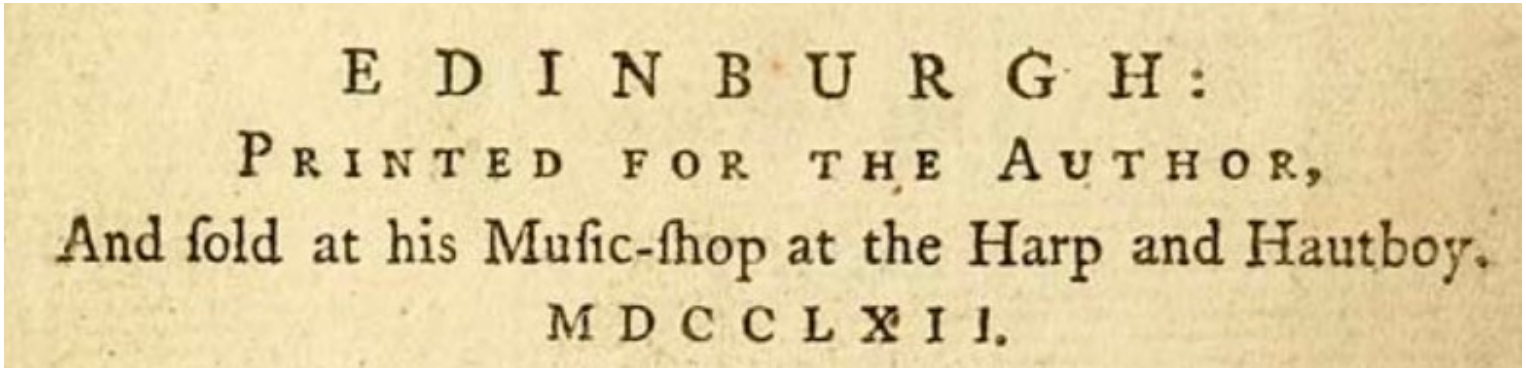
is form of? ↑

Transcribed elements	
Sources	Manifestation (Item in hand)
Tasks	Identify
Entities	Manifestation

RDA

Transcription

What you see is what you get?



Digital image

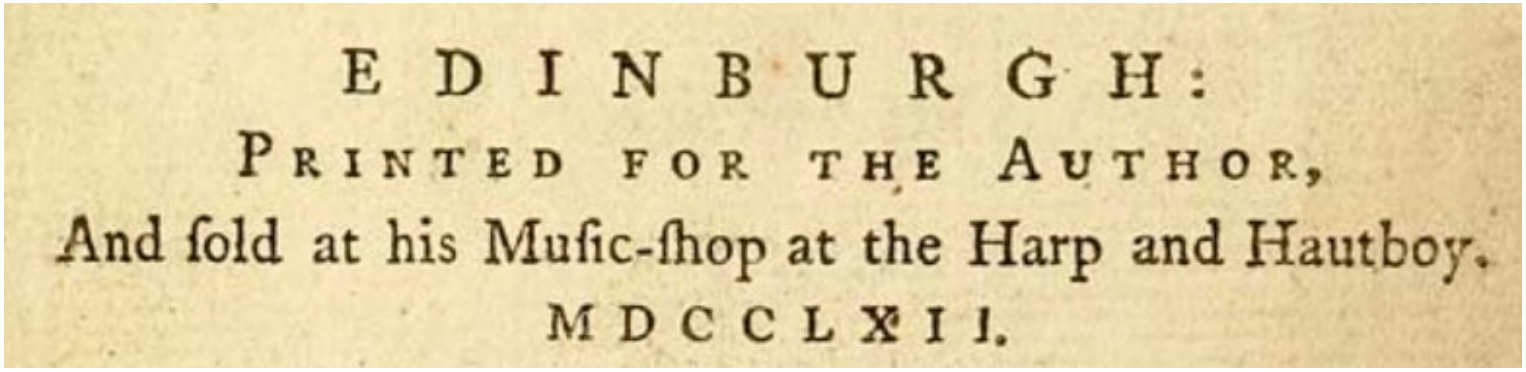
Optical Character Recognition transcription

EDINBURGH:
Printed for the Author,
And fold at his Mufic-fhop at the Harp and Hautboy,
M D C C L X I J.

RDA

Transcription

What you see is what you get?


EDINBURGH:
PRINTED FOR THE AUTHOR,
And ſold at his Muſic-ſhop at the Harp and Hautboy.
MDCCLXII.

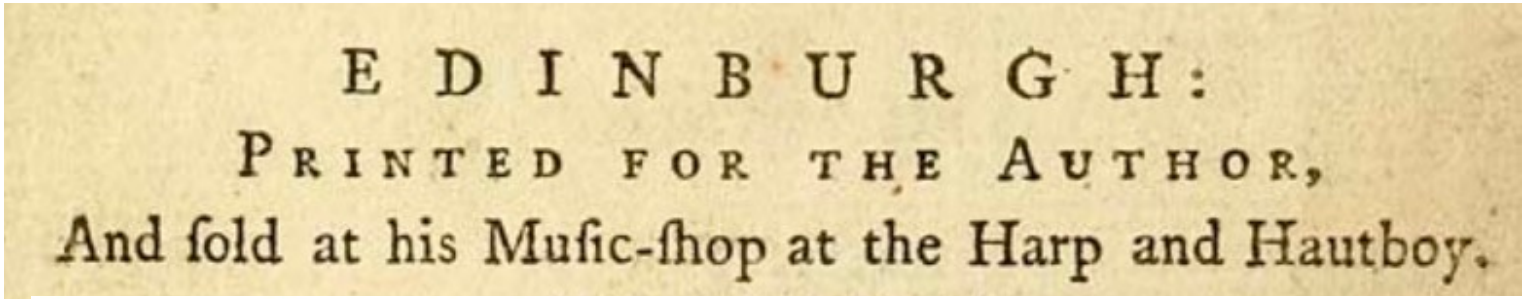
"User" transcription

EDINBURGH:
PRINTED FOR THE AUTHOR,
And fold at his Music-fhop at the Harp and Hautboy.
MDCCLXII.

RDA

Transcription

What you see is what you get?


EDINBURGH:
PRINTED FOR THE AUTHOR,
And ſold at his Muſic-ſhop at the Harp and Hautboy.

Recording Publication Statements 2014/04

Record a publication statement or statements for a published resource.

Transcribe places of publication and publishers' names as they appear on the source of information (see 1.7 RDA).

Record dates of publication as they appear on the source of information. Apply the general guidelines on transcription for words that are not numbers (see 1.7 RDA). Apply the general guidelines on numbers expressed as numerals or as words (see 1.8 RDA).

Edinburgh: Printed for the author, and sold at his music-shop at the Harp and Hautboy, 1762

RDA

Transcription for “Identify” task

Digital image is:

- Quickest and cheapest
- Easiest for user with item/image in hand

21st century!
Web of machines!

App + Camera + Touch-screen +
Image matching software service

Transcription string for item citation:

- User must know transcription rules
- Is OCR good enough?
  - Feedback capture = Crowdsourcing

RDA

Recording for user tasks

If data is not transcribed, it is recorded

Recording excludes (more or less):

- Typos
- Deliberate errors
- Fictitious entities

Some of the recorded data support the Find, Identify, Select, Obtain, or Explore user tasks
How can the data best be accommodated in RDA?

RDA

N-fold path

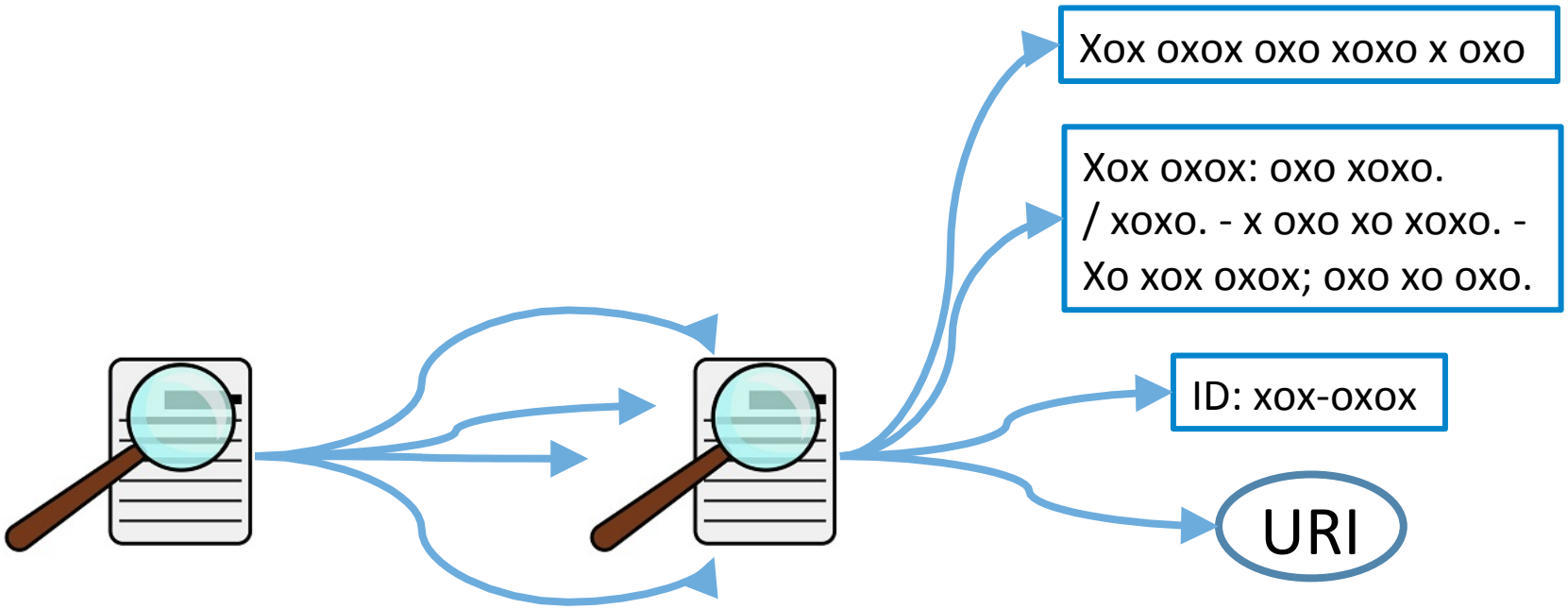
1. Unstructured string.
    1. Exact transcription (OCR or born digital).
    2. Transcription using the RDA guidelines.
    3. Data recorded from another source.
2. Structured string of delimited sub-values.
    1. Access point.
    2. Structured description.
3. Structured string.
    1. Identifier
4. URI of entity, including Nomen.
    1. URI/URL of digital image.

RDA

The path starts here

Paths are available for describing related entities

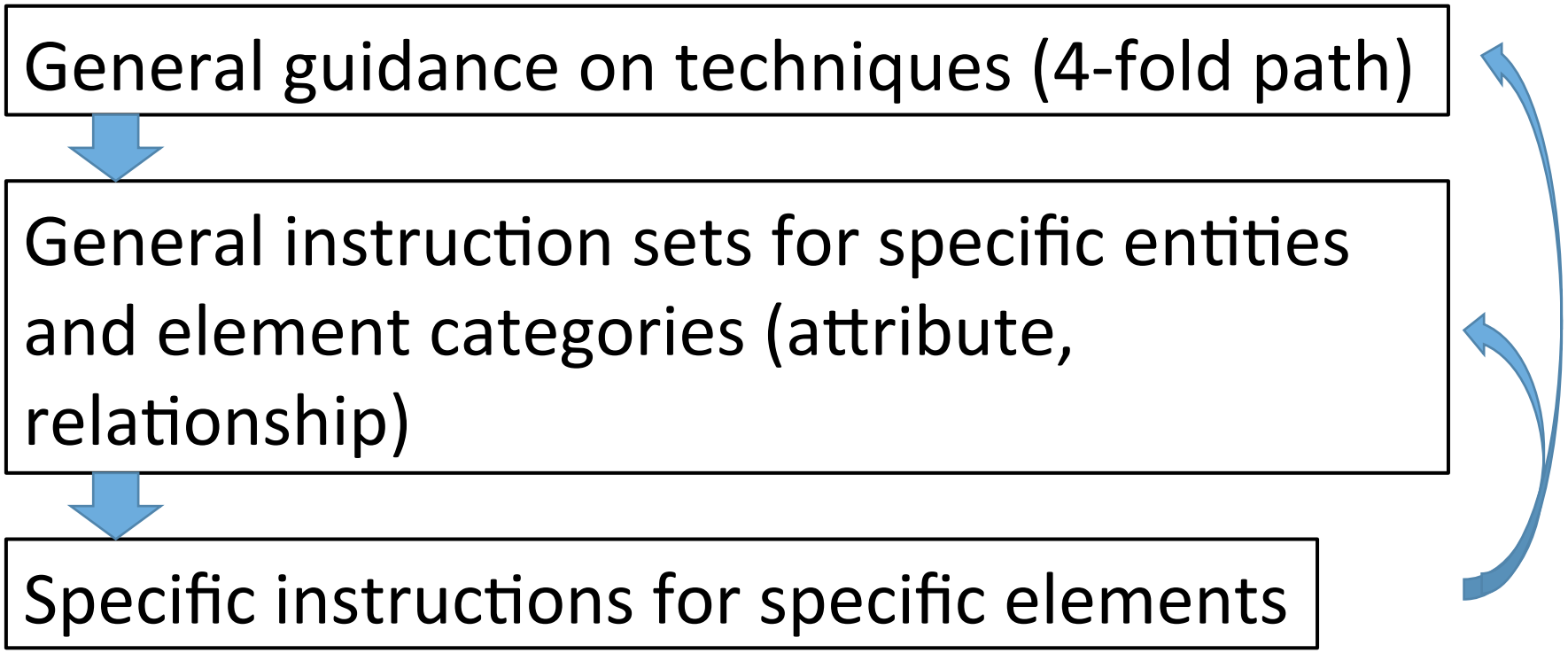
The same paths describe the entity in focus



RDA

Developing Toolkit guidance and instructions

Methods of recording RDA data



RDA

Developing RDA Registry for applications

Elements for storage of RDA (linked) data

Element domain = parent Entity (constrained)

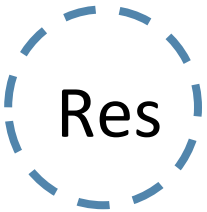
Element range = type of path (not currently specified)

Sub-properties (sub-types) of each element have 2 types of range to accommodate 4-fold path: literal and object

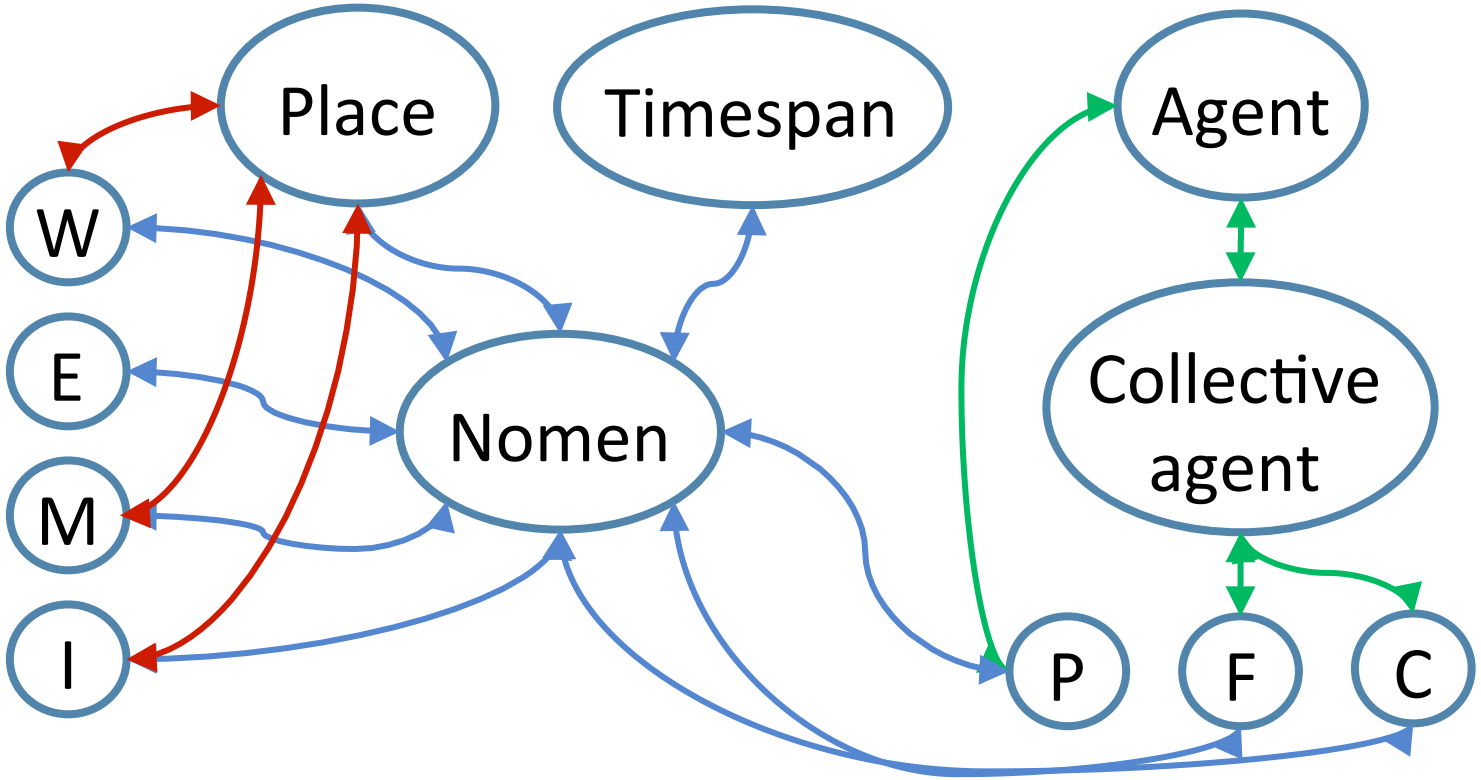
Element range	Path
Literal	Unstructured
Literal (associated with construction encoding scheme)	Structured/AP/Identifier
Object	URI

RDA

New entities



New high-level relationship elements



New relationship designators (cross-entity)

RDA

Toolkit Entity views

Proposed development to provide a focus for each RDA entity and its elements

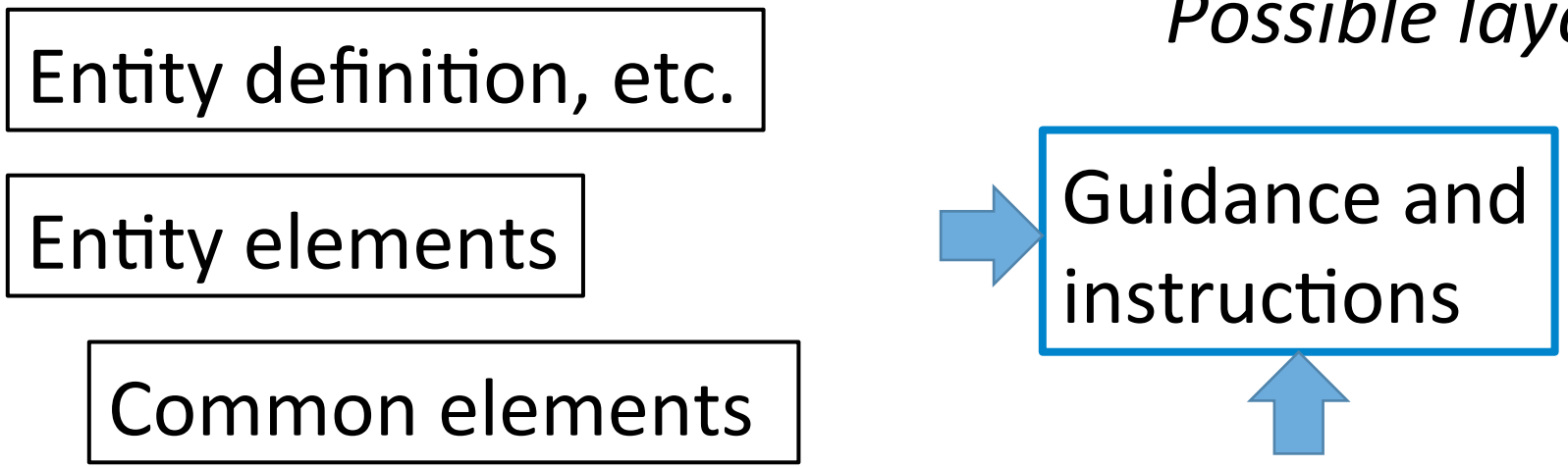
Replaces out-of-date Element set views

Acts as a ready-reference to all elements and instructions associated with the entity

RDA

Entity view: a dictionary/reference for RDA

Possible layout



With n-fold path:

- literal range + associated structure
- object (Entity) range

RDA

Re-organizing the Toolkit

- Appendices and tabs
- Vocabulary Encoding Schemes
  - Sharing, extending, linking (RDA and other communities)
- RDA Reference (entities, elements, terms)
- Glossary
  - How far beyond entities, elements, and vocabulary terms?
- Translations
- Policy statements and application profiles
- Entity views, Relationship designators, etc.

RDA

Some issues

- Needs of international, cultural heritage, and linked data communities
- Primary (WEMI) vs Secondary (PFC …) entities
  - Reciprocal relationships/links/designators
  - Elements other than those for access points?
- Structure in descriptions
  - How much specification?
    - International communities use different structures
- Nomen control (a kind of authority control?)
- Relationship designators
  - Cross-entity, and many more (labels, definitions?)

RDA

Thank you!

- rscchair@rdatoolkit.org
- http://access.rdatoolkit.org/
- http://www.rdaregistry.info/
- http://www.rda-rsc.org/

R | D | A