

Resource Description: Metadata/Dublin Core

CS 431 – February 21, 2007

Carl Lagoze – Cornell University

But first

A bit more XML

Referencing a name within the schema

- “ref” attribute
- Example:

http://www.cs.cornell.edu/courses/CS431/2007sp/examples/xml_schema/business_card.xsd

Type Reuse

- Include – add definitions in same namespace
- Redefine – extend or restrict definitions in same namespace
- Import – Create new namespace and use definitions from another namespace

Import Example

- Address schema
 - http://www.cs.cornell.edu/courses/CS431/2007sp/examples/xml_schema/address.xsd
- Person schema
 - http://www.cs.cornell.edu/courses/CS431/2007sp/examples/xml_schema/person.xsd
- Instance document
 - http://www.cs.cornell.edu/courses/CS431/2007sp/examples/xml_schema/person.xml

Include Example

- Address schema
 - http://www.cs.cornell.edu/courses/CS431/2007sp/examples/xml_schema/address.xsd
- Person schema
 - http://www.cs.cornell.edu/courses/CS431/2007sp/examples/xml_schema/person2.xsd
- Instance document
 - http://www.cs.cornell.edu/courses/CS431/2007sp/examples/xml_schema/person2.xml

Redefine Example

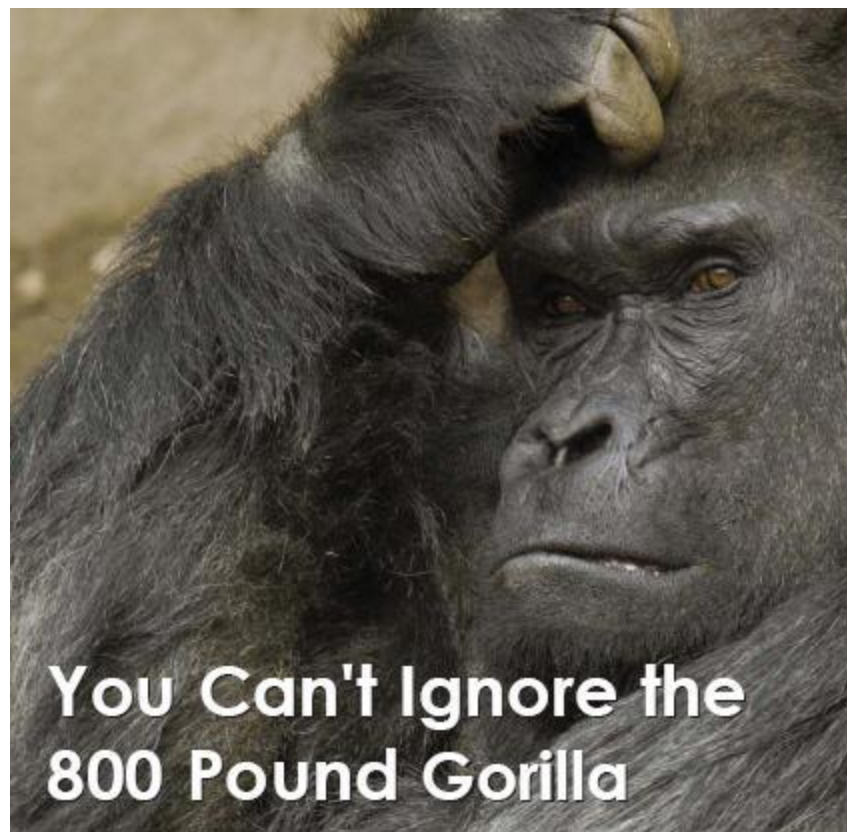
- Address schema
 - http://www.cs.cornell.edu/courses/CS431/2007sp/examples/xml_schema/address.xsd
- Person schema
 - http://www.cs.cornell.edu/courses/CS431/2007sp/examples/xml_schema/person3.xsd
- Instance document
 - http://www.cs.cornell.edu/courses/CS431/2007sp/examples/xml_schema/person3.xml

Substitution Groups

- Provide alternates for a sub-tree
- Example
 - http://www.cs.cornell.edu/courses/CS431/2007sp/examples/xml_schema/products.xsd
 - http://www.cs.cornell.edu/courses/CS431/2007sp/examples/xml_schema/products.xml

Acknowledgments

- Andy Powell, Head of Development, Eduserv Foundation, UK
- Tom Baker, Dublin Core Metadata Initiative
- Diane Hillmann, Cornell University
- Erik Wilde, UC Berkeley School of Information



**You Can't Ignore the
800 Pound Gorilla**



Dublin Core Metadata Initiative®

Making it easier to find information.

Dublin Core

- Origins at 1994 Web Conference
 - Metadata was necessary for finding things on the web
 - Simple cross-domain vocabulary (15 elements) describing “document-like” objects
- 2004 ISO standard elements
 - <http://dublincore.org/documents/dces/>

The fifteen Dublin Core Elements

Creator	Title	Subject
Contributor	Date	Description
Publisher	Type	Format
Coverage	Rights	Relation
Source	Language	Identifier

<http://dublincore.org/documents/dces/>

Dublin Core **Qualifiers**

- From loose semantics to more specific description
- Model of “graceful degradation”
 - Support both simplicity and specificity
 - Intra-domain and inter-domain semantics
- Informally three class of qualification
 - Element refinement – from “date” to “date published”, from “contributor” to “illustrator”
 - Value encoding schemes – from “subject” to “LCSH subject”
 - Language

The Dublin Core Vocabulary

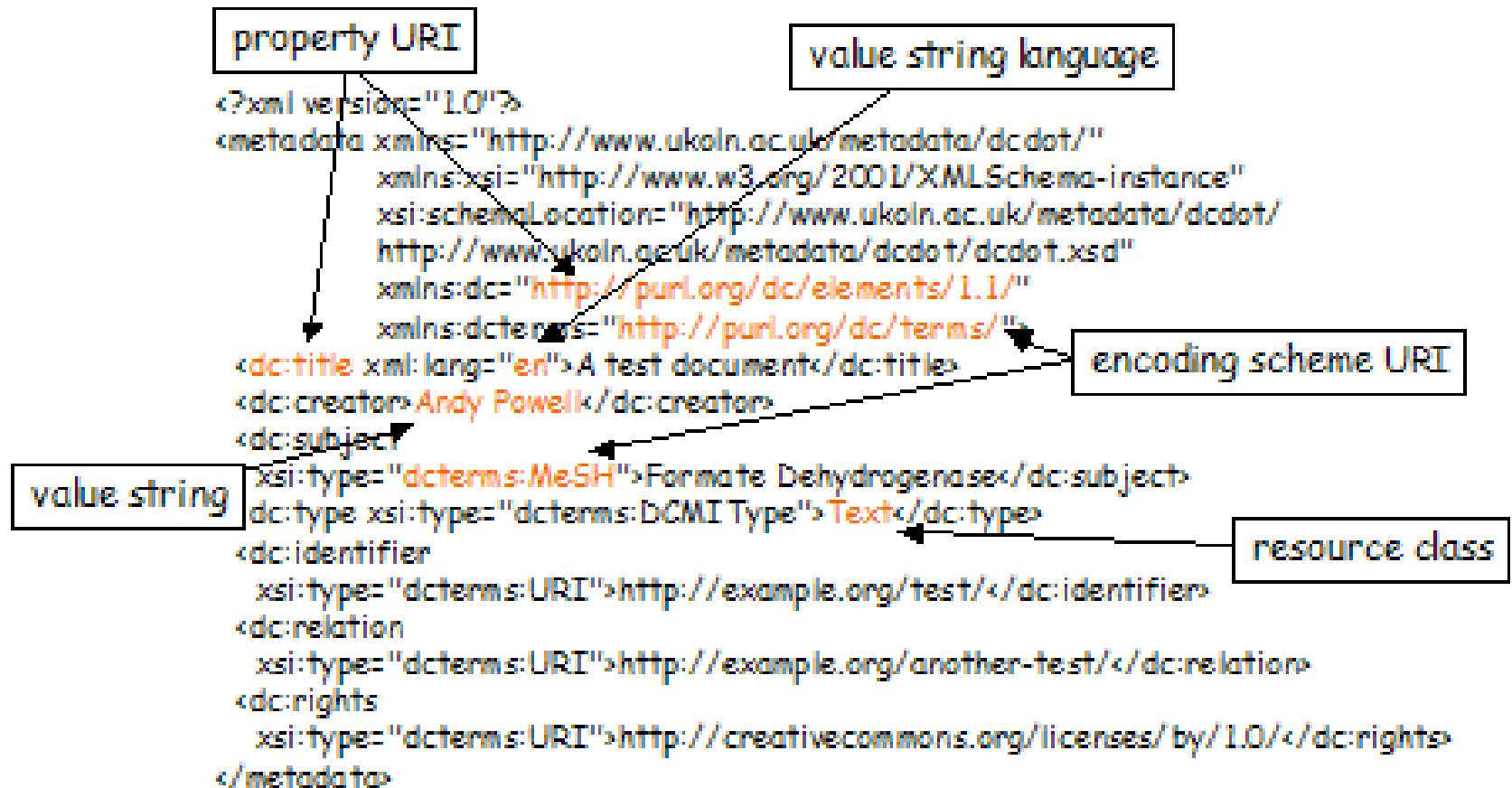
<http://dublincore.org/documents/dcmi-terms/>

Elements	Refinements		Schemes	Types
1. Identifier	Abstract	Is referenced by	Box	Collection
2. Title	Access rights	Is replaced by	DCMIType	Dataset
3. Creator	Alternative	Is required by	DDC	Event
4. Contributor	Audience	Issued	IMT	Image
5. Publisher	Available	Is version of	ISO3166	Interactive
6. Subject	Bibliographic citation	License	ISO639-2	Resource
7. Description	Conforms to	Mediator	LCC	Moving Image
8. Coverage	Created	Medium	LCSH	Physical Object
9. Format	Date accepted	Modified	MESH	Service
10. Type	Date copyrighted	Provenance	Period	Software
11. Date	Date submitted	References	Point	Sound
12. Relation	Education level	Replaces	RFC1766	Still Image
13. Source	Extent	Requires	RFC3066	Text
14. Rights	Has format	Rights holder	TGN	
15. Language	Has part	Spatial	UDC	
	Has version	Table of contents	URI	
	Is format of	Temporal	W3CTDF	
	Is part of	Valid		

Dumb-down

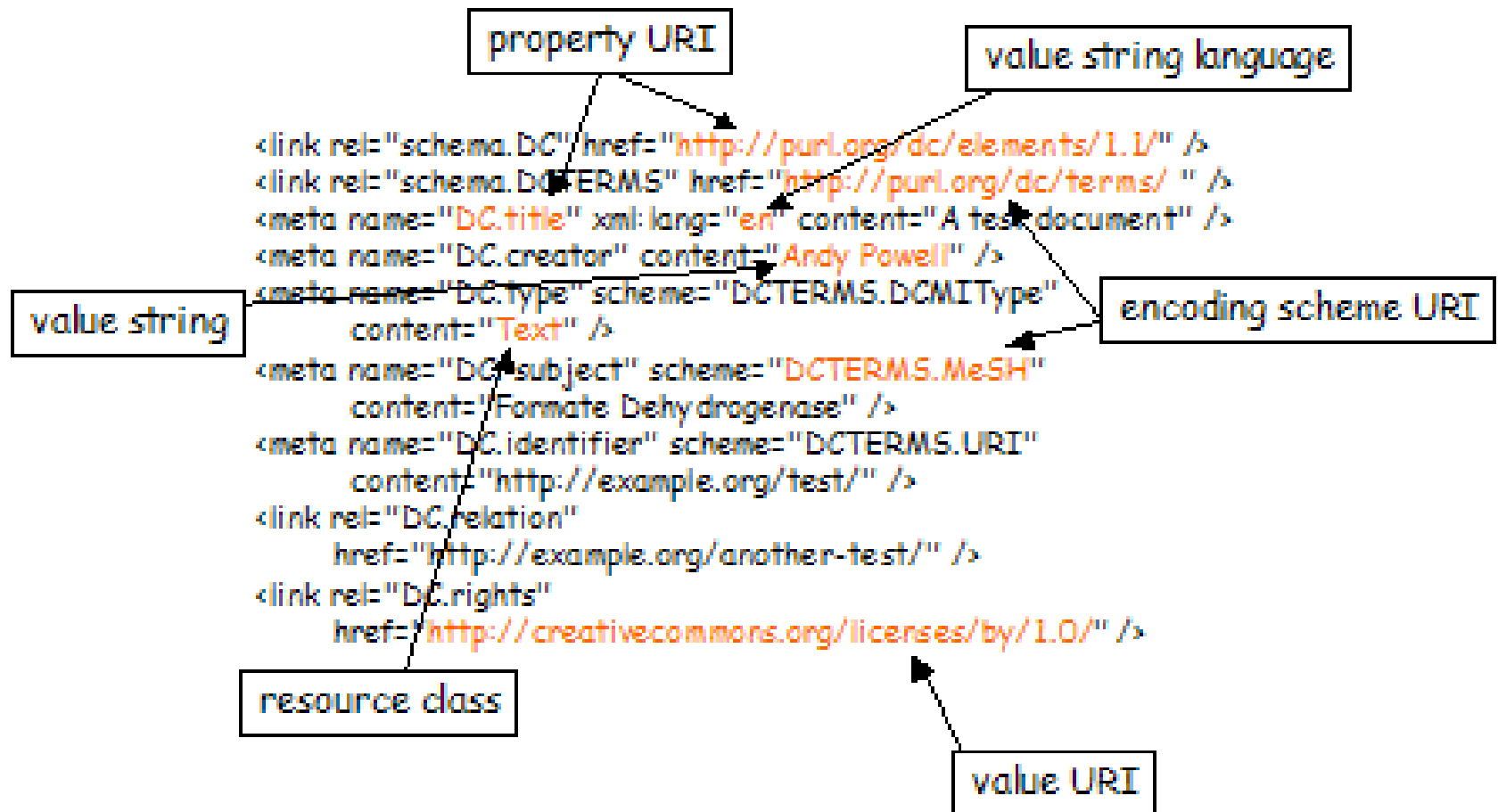
- the process of translating a qualified DC metadata record into a simple DC metadata record is normally referred to as 'dumbing-down'
- can be separated into two parts:
 - Property – from refinement to core element
 - Value – from encoding to basic string

Encoding DC - XML



<http://dublincore.org/documents/2002/12/02/dc-xml-guidelines/>

Encoding DC - XHTML

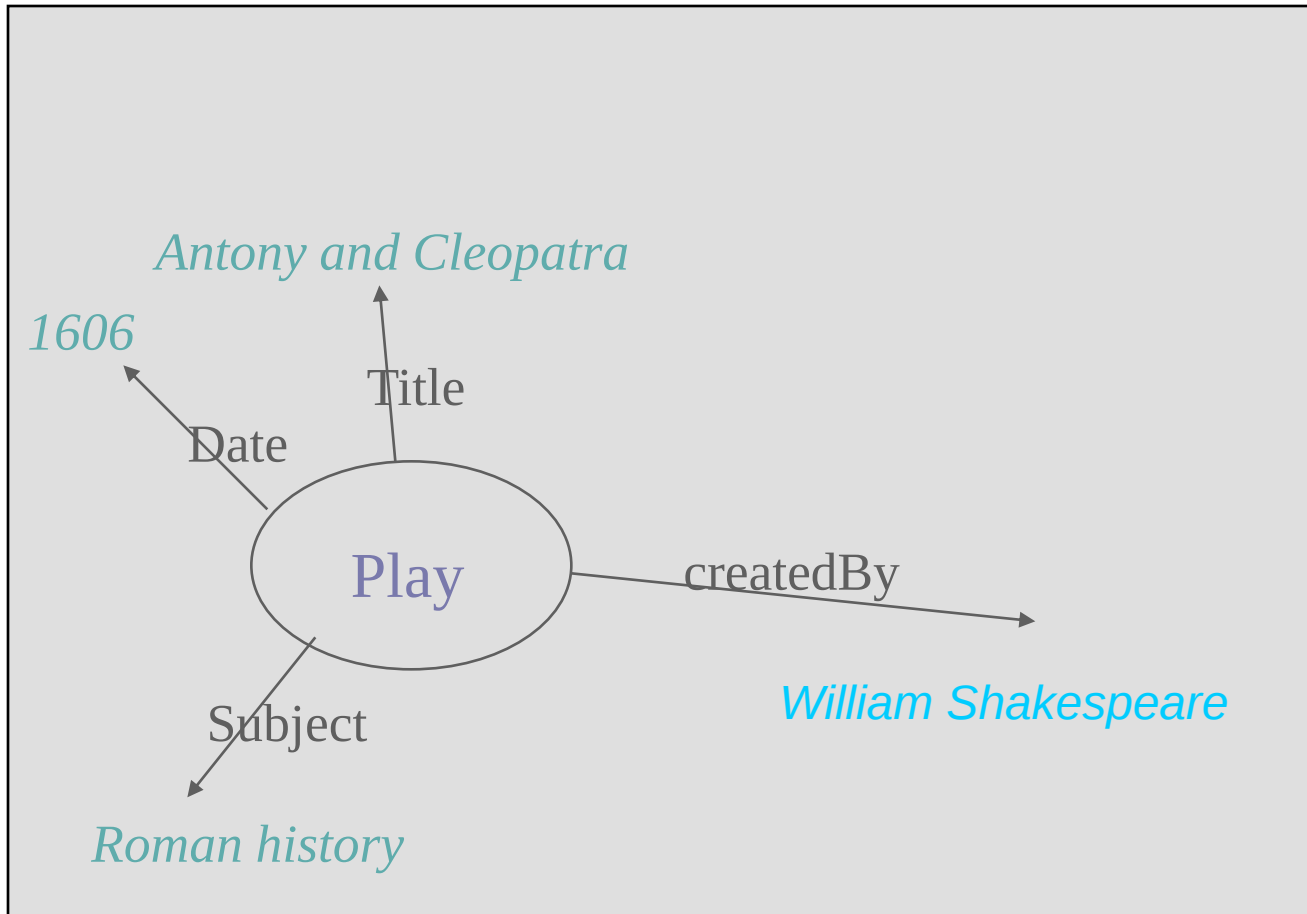


<http://dublincore.org/documents/dcq-html/>

DC Vocabulary in Context: Model and modularity

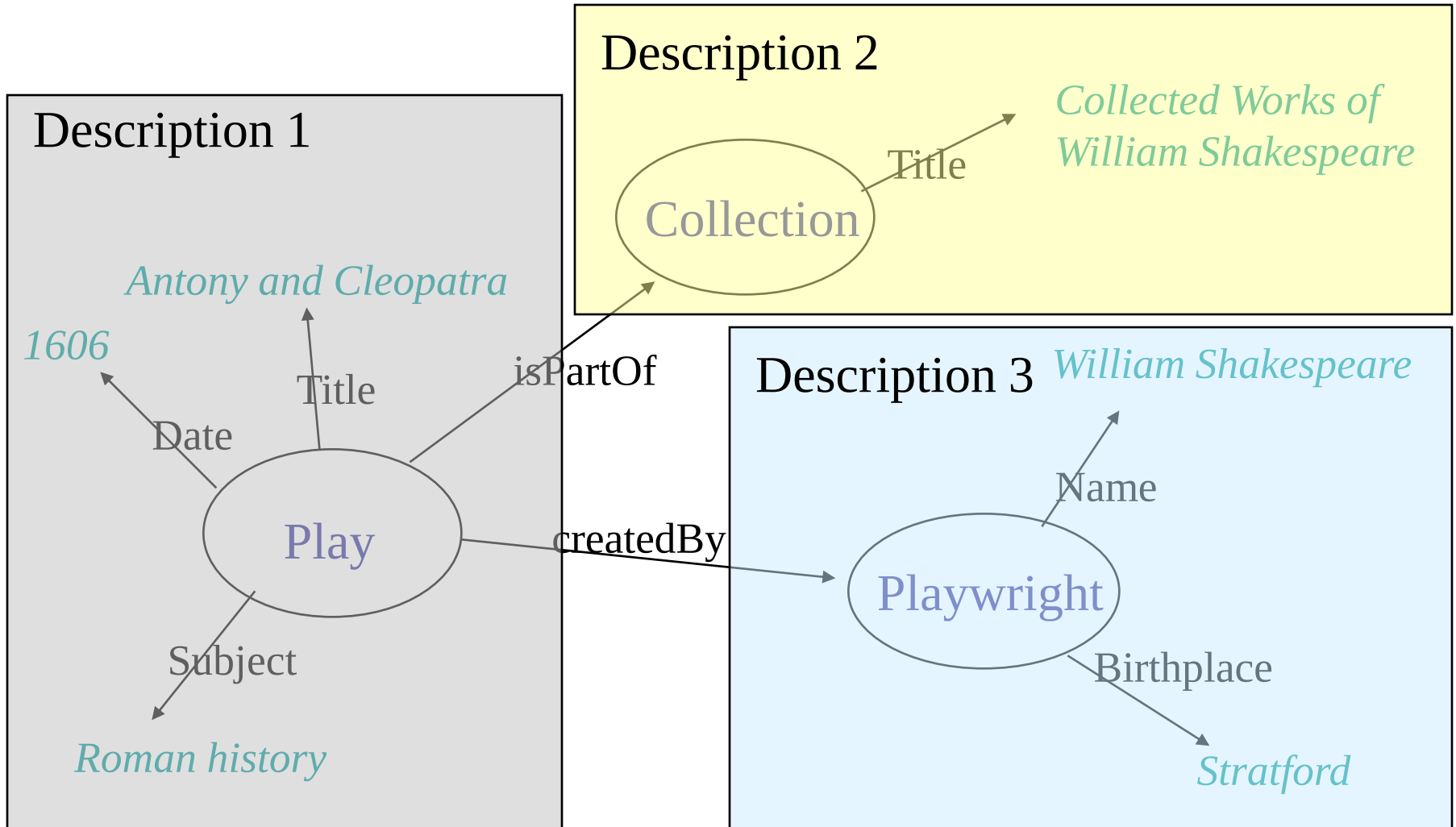
- Resources are related to each other
- There are many vocabularies

One resource, on description



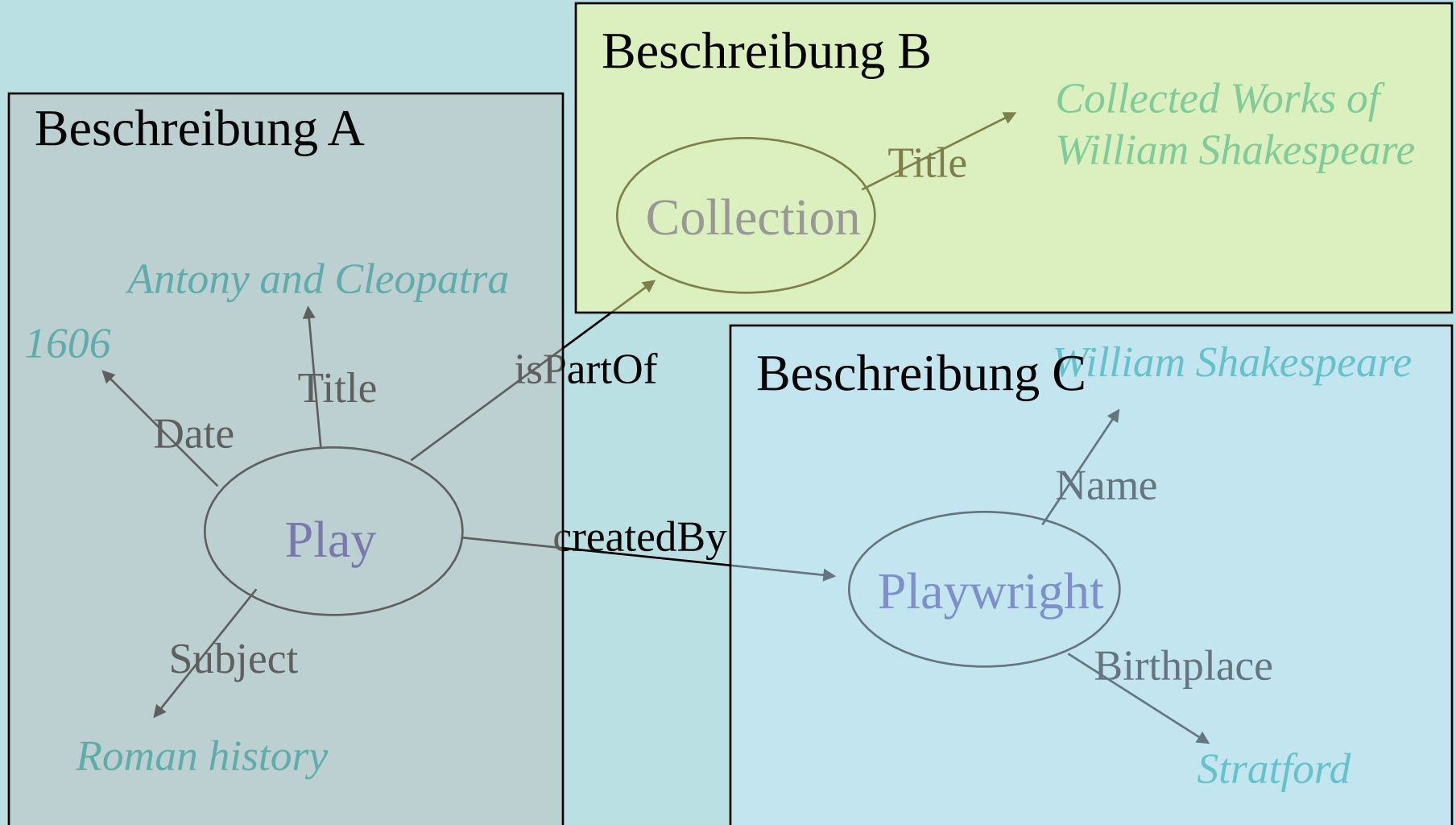
Relationship among many resources

One-to-one principle



...in one record

Description Package



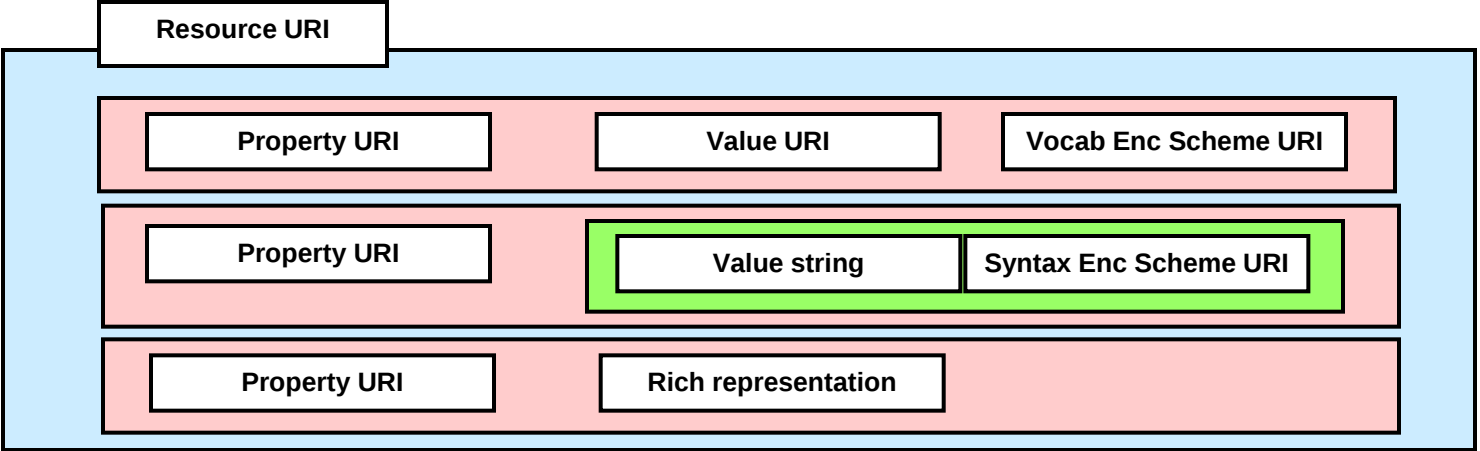
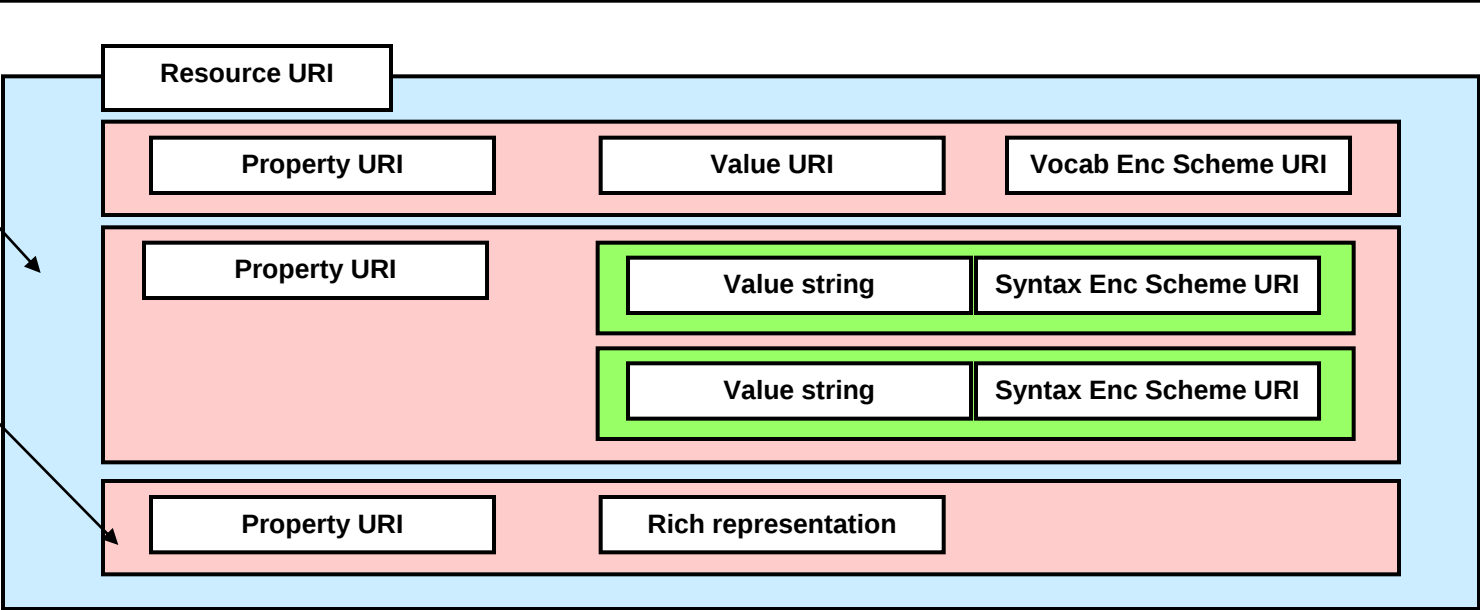
Dublin Core Abstract Model

Packaging multiple descriptions
and vocabularies together

Description Set

Description

Statement



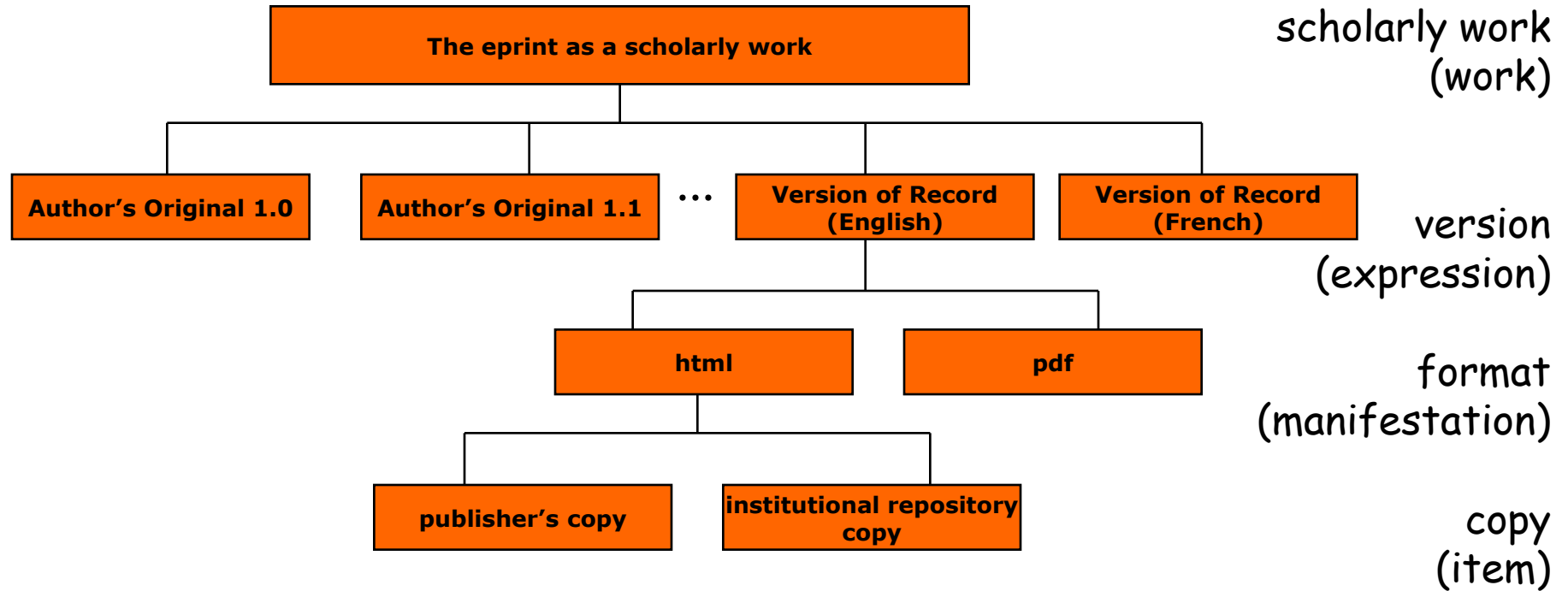
Packaging a Complex Object

```
<descriptionSet>
  <description resourceURI=http://eprints.gla.ac.uk/503/>
    <statement propertyURI=dc:title> <valueString>Attempts to detect
    retrotransposition and de novo deletion of Alus and other dispersed repeats at
    specific loci in the human genome </valueString> </statement>
    <statement propertyURI=eprint:isExpressedAs valueRef=expression1 />
  </description>
  <description resourceId=expression1 >
    <statement propertyURI=eprint:isManifestedAs valueRef=pdfmanifestation />
  </description>
  <description resourceId=pdfmanifestation >
    <statement propertyURI=eprint:isAvailableAs
      valueURI=http://eprints.gla.ac.uk/503/01/Eu_J._Hum_Gen.9(2)143_.pdf />
    <statement propertyURI=eprint:isAvailableAs
      valueURI=http://www.nature.com/ejhg/journal/v9/n2/pdf/5200590a.pdf />
  </description>
  <!-- descriptions of the two copies here -->
</descriptionSet>
```

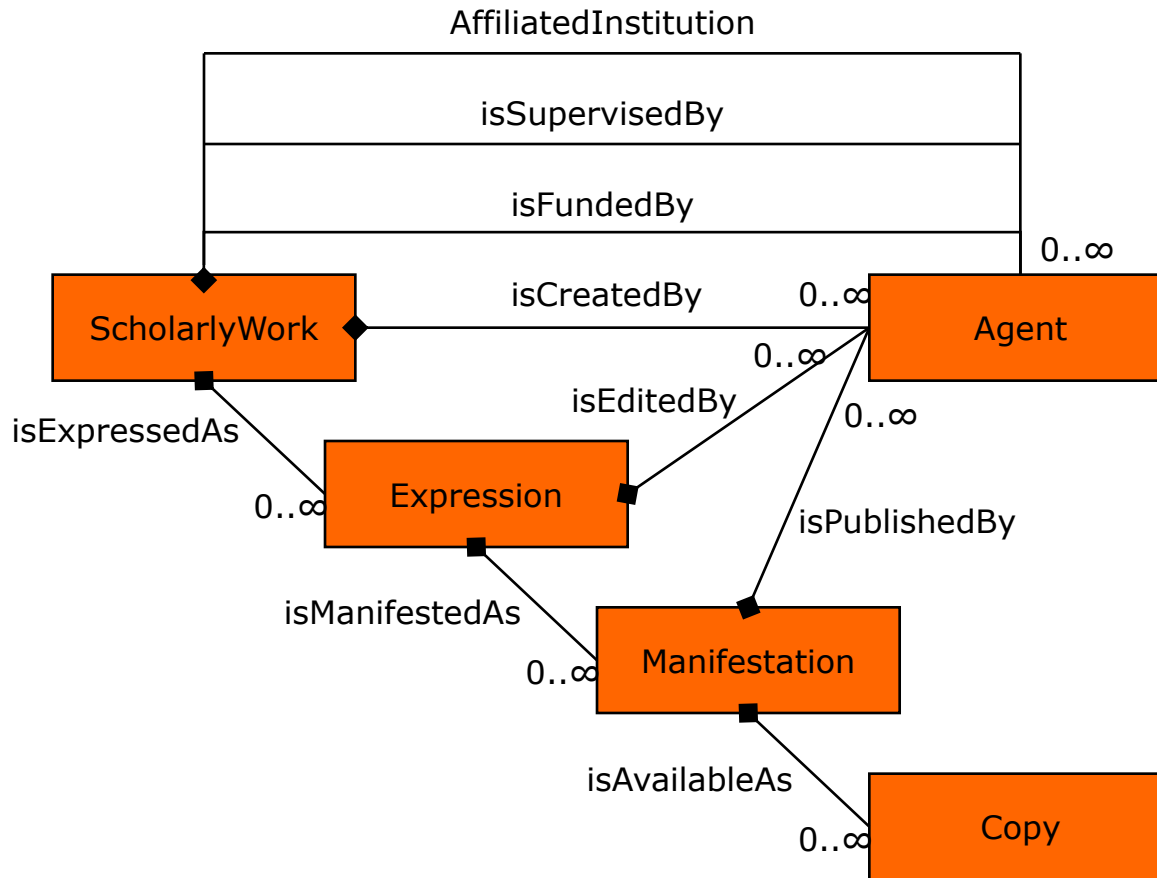
Applying this model in the context of scholarly communication

- Increasing availability of scholarly research in open access repositories – e.g., arXiv
 - Mirrored
 - Multi-format (pdf, LaTeX)
 - Co-exist in journal published form and ePrint form
- FRBR is a model for representing these relationships.

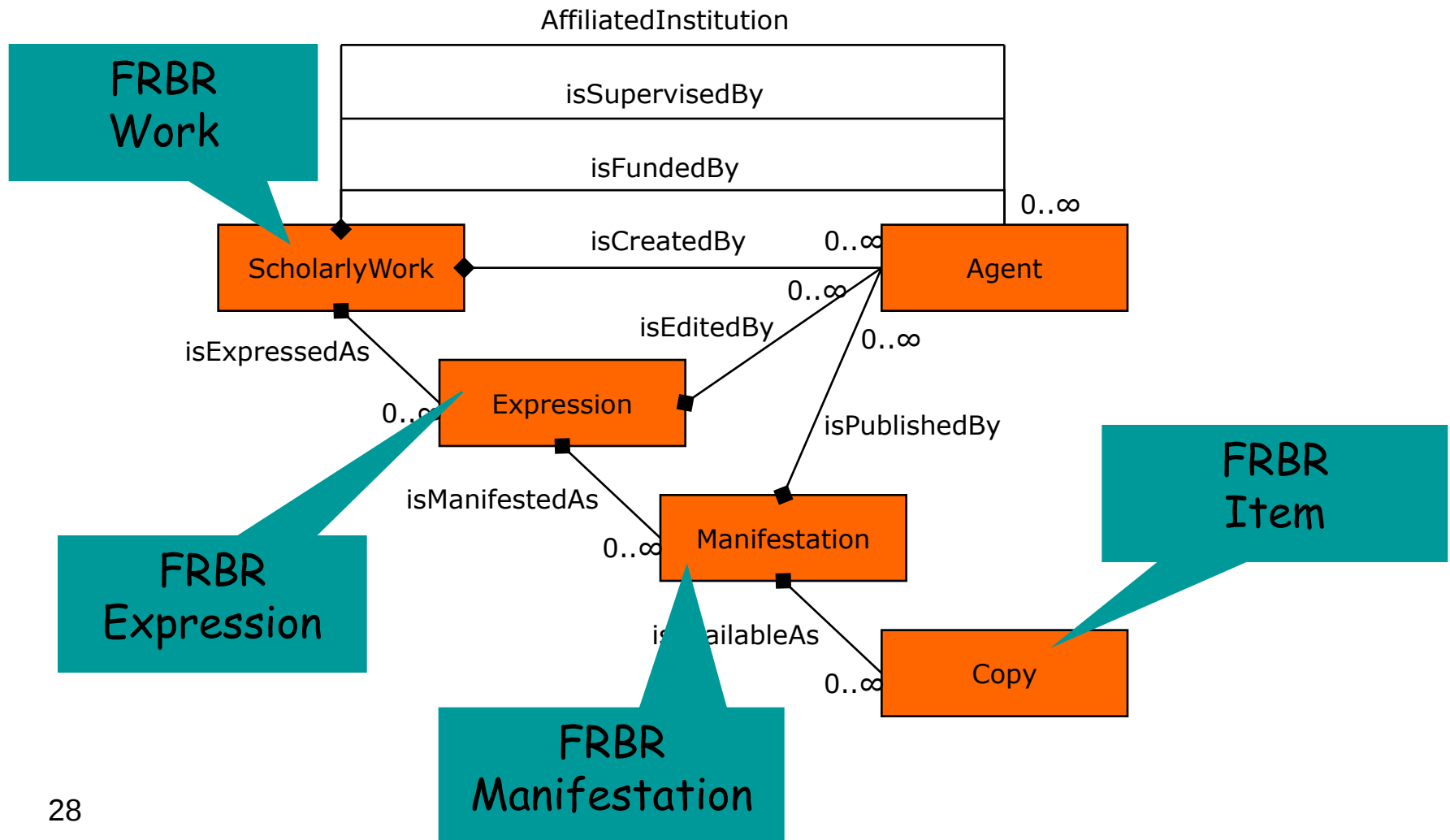
FRBR for eprints



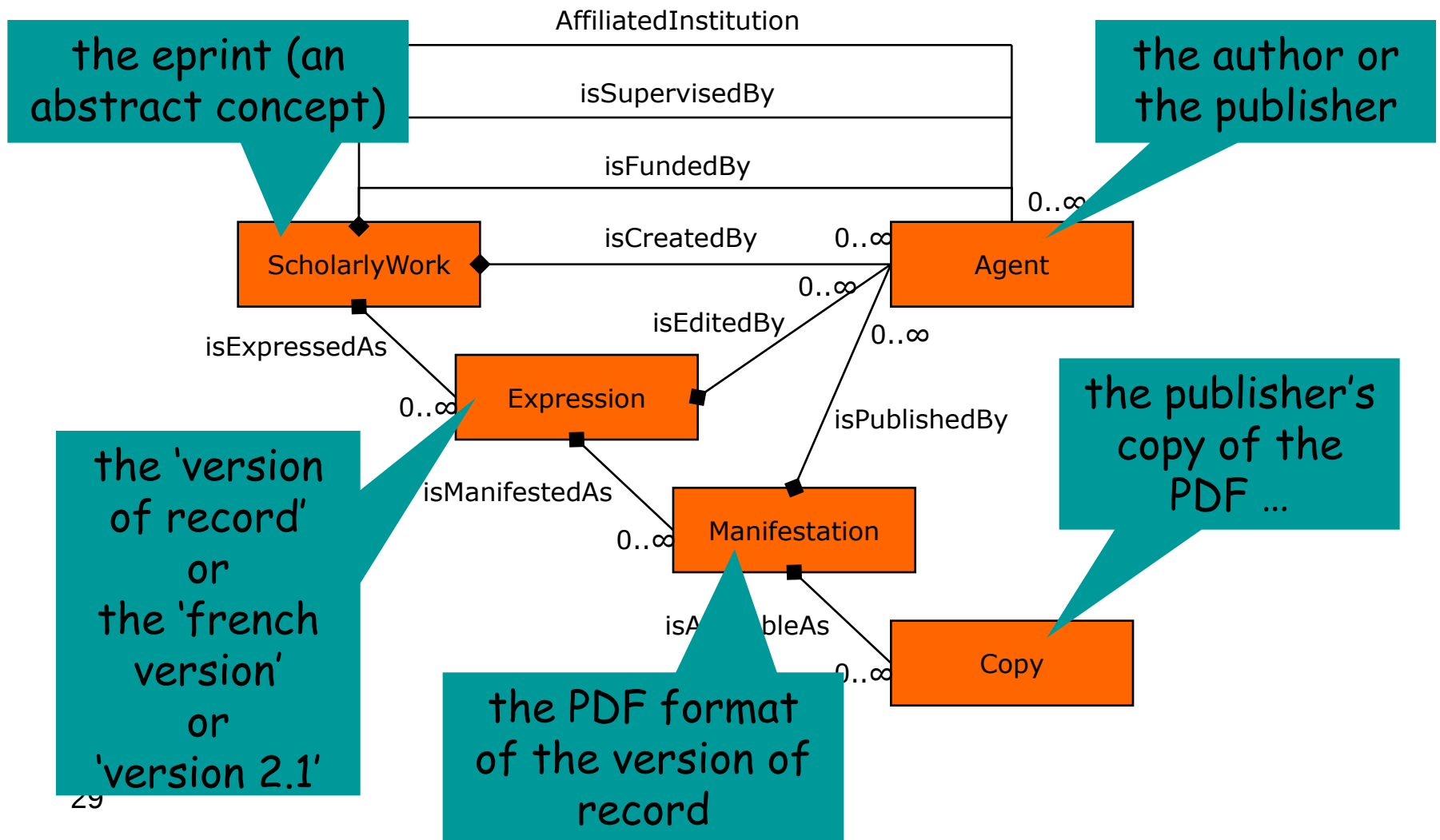
Eprints application model



Eprints model and FRBR



Eprints model and FRBR



Attributes

- the application model defines the entities and relationships
- each entity needs to be described using an agreed set of attributes

Example attributes

ScholarlyWork:

title
subject
abstract
affiliated institution
identifier

Expression:

title
date available
status
version number
language
genre / type
copyright holder
bibliographic citation
identifier

Agent:

name
type of agent
date of birth
mailbox
homepage
identifier

Manifestation:

format
date modified

Copy:

date available
access rights
licence
identifier



How is this complexity captured in DC?

- the DC Abstract Model provides the notion of ‘description sets’
- i.e. groups of related ‘descriptions’
- where each ‘description’ is about an instance of one of the entities in the model
- relationships and attributes are instantiated as metadata properties

```
<ep:description
  ep:resourceURI="http://eprints.gla.ac.uk/503/" >
  <ep:statement
    ep:propertyURI="&dc:type"
    ep:valueURI="&epentType;ScholarlyWork" />

  <!-- more statements describing Scholarly Work -->

  <ep:statement
    ep:propertyURI="&epprint;isExpressedAs"
    ep:valueRef="expression1" />
  </ep:description>
```

Description of
Scholarly Work

```
<ep:description
  ep:resourceId="expression1" >
  <ep:statement
    ep:propertyURI="&dc:type"
    ep:valueURI="&epentType;Expression" />

  <!-- more statements describing Expression -->

  <ep:statement
    ep:propertyURI="&epprint;isManifestedAs"
    ep:valueRef="pdfmanifestation" />
  </ep:description>
```

Description of
Expression

```
<ep:description
  ep:resourceId="pdfmanifestation" >
  <ep:statement
    ep:propertyURI="&dc:type"
    ep:valueURI="&epentType;Manifestation" />

  <!-- more statements describing Manifestation -->

  <ep:statement
    ep:propertyURI="&epprint;isAvailableAs"
    ep:valueURI="http://eprints.gla.ac.uk/503/01/Eu_J._Hum_Gen.9(2)143_.pdf" />
  <ep:statement
    ep:propertyURI="&epprint;isAvailableAs"
    ep:valueURI="http://www.nature.com/ejhg/journal/v9/n2/pdf/5200590a.pdf" />
  </ep:description>
```

Description of
Manifestation

```
<ep:description
  ep:resourceId="http://eprints.gla.ac.uk/503/01/Eu_J._Hum_Gen.9(2)143_.pdf" >
  <ep:statement
    ep:propertyURI="&dc:type"
    ep:valueURI="&epentType;Copy" />

  <!-- more statements describing Copy 1 -->

  </ep:description>
```

Description of
Copy 1

```
<ep:description
  ep:resourceId="http://www.nature.com/ejhg/journal/v9/n2/pdf/5200590a.pdf" >
  <ep:statement
    ep:propertyURI="&dc:type"
    ep:valueURI="&epentType;Copy" />

  <!-- more statements describing Copy 2 -->

  </ep:description>
```

Description of
Copy 2

```
</ep:descriptionSet>
```

Resources

- DCMI Abstract Model
 - <http://dublincore.org/documents/abstract-model/>
- Eprints Application Profile
 - http://www.ukoln.ac.uk/repositories/digirep/index/Eprints_Application_Profile
- Eprints DC XML
 - http://www.ukoln.ac.uk/repositories/digirep/index/Eprints_DC_XML
- Eprints DC XML/Instances
 - http://www.ukoln.ac.uk/repositories/digirep/index/Eprints_DC_XML/Instances