eduserv

ORE & SWAP: Composition & Complexity

14 October 2008

Metadata Infrastructures, MPG eScience Seminar, Berlin

Pete Johnston, Eduserv Foundation
pete.johnston@eduserv.org.uk
http://www.eduserv.org.uk/foundation

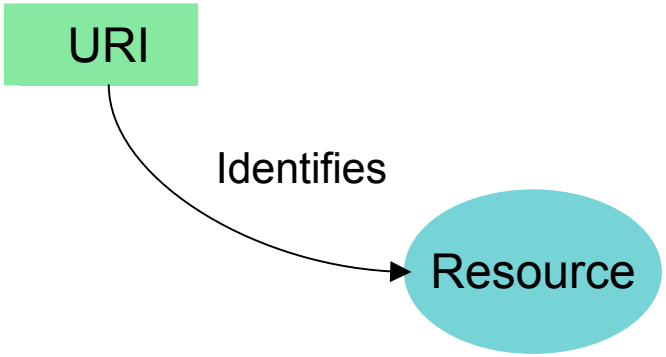
ORE & SWAP: Composition & Complexity

eduserv

- Background/Foundations
  - Web Architecture
  - Resource Description Framework
- OAI Object Reuse & Exchange (OAI ORE)
- More Background/Foundations
  - The DCMI Abstract Model
  - Dublin Core Application Profiles
- Scholarly Works DC Application Profile (SWAP)
- Conclusions

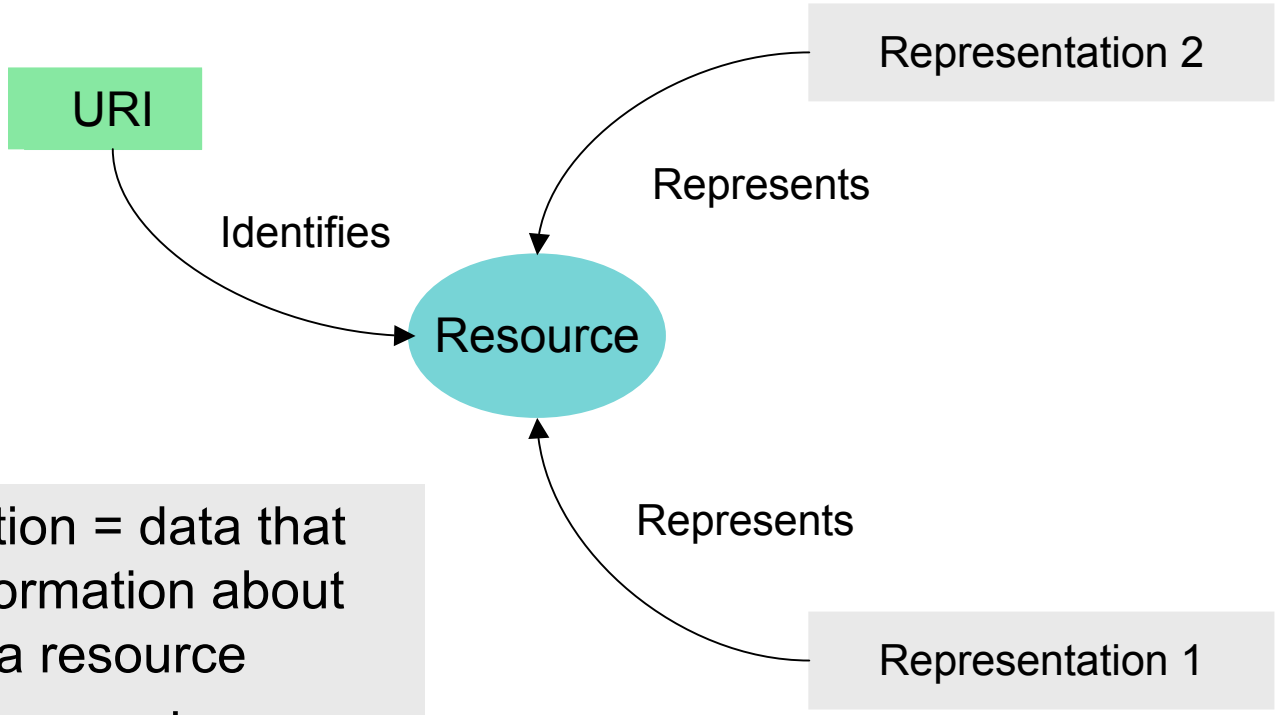
Background: Web Architecture, RDF & (briefly) Linked Data

W3C Web Architecture: Identification



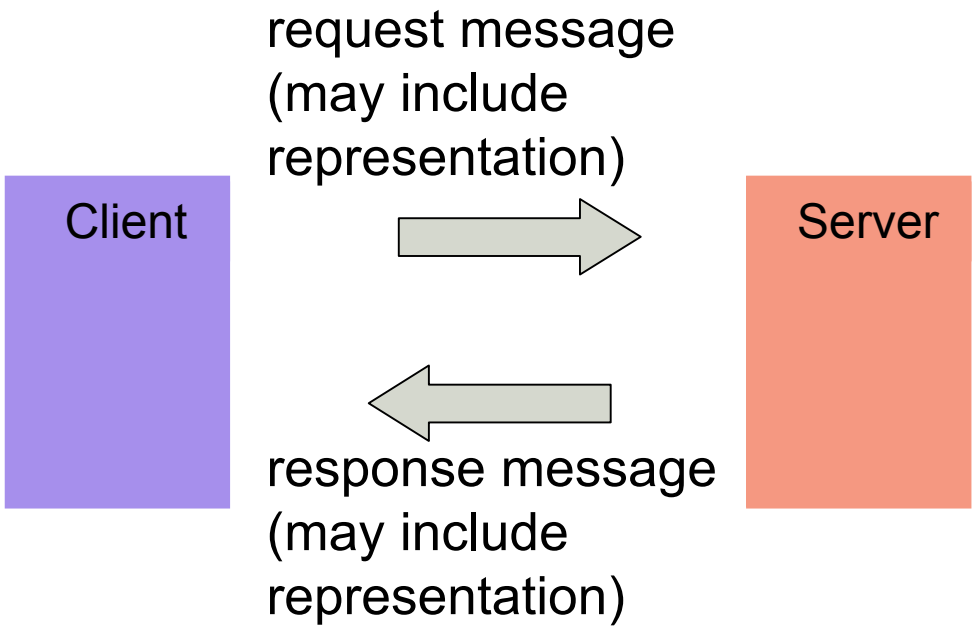
- Resource = anything of interest
- Resources are identified by URIs

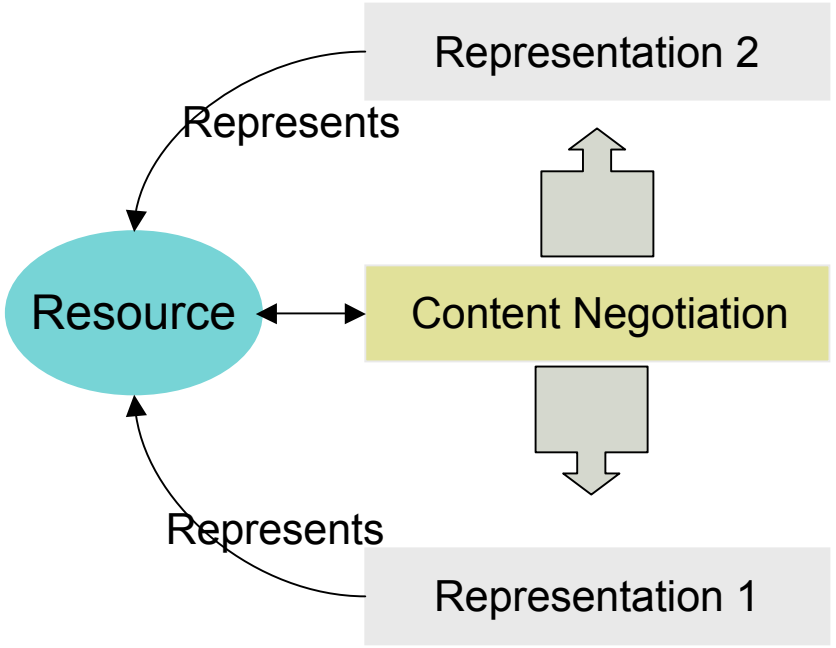
W3C Web Architecture: Representation



- Representation = data that encodes information about the state of a resource
- One resource may have multiple representations
- Representations may change over time

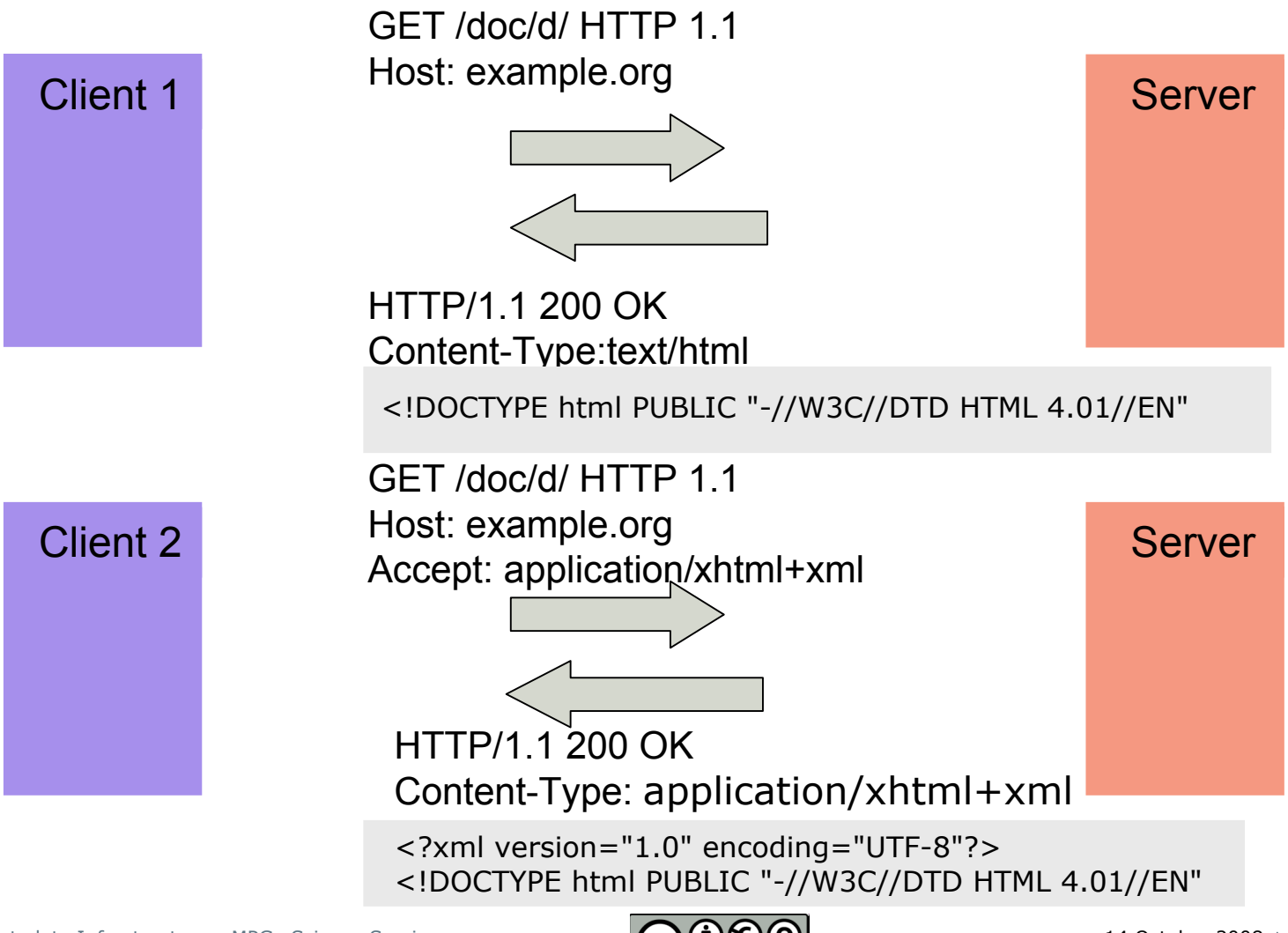
W3C Web Architecture: Interaction





- Resources manipulated through exchange of representations
- Messages exchanged using network protocols
- Resource is always “hidden”

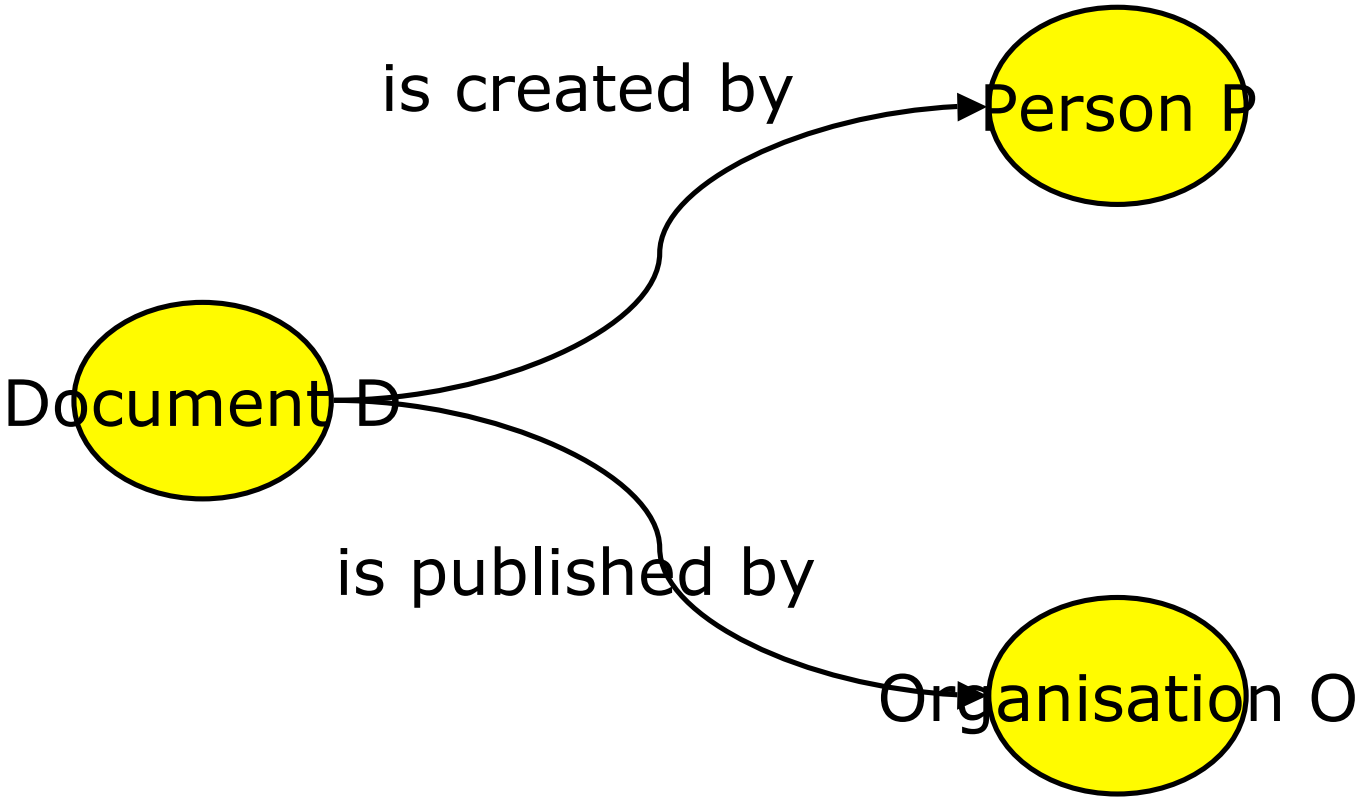
W3C Web Architecture: Interaction



Resource Description Framework (RDF)

- Set of specifications from W3C
- A simple data model...
- ...for making assertions about relationships between resources
  - "Thing-X is-related-in-some-way-to Thing-Y"
- Types of relationship are properties
- Assertions made in the form of triples
  - Subject, Predicate, Object
- Sets of triples represented as graphs
- Serialisation in multiple formats, including RDF/XML

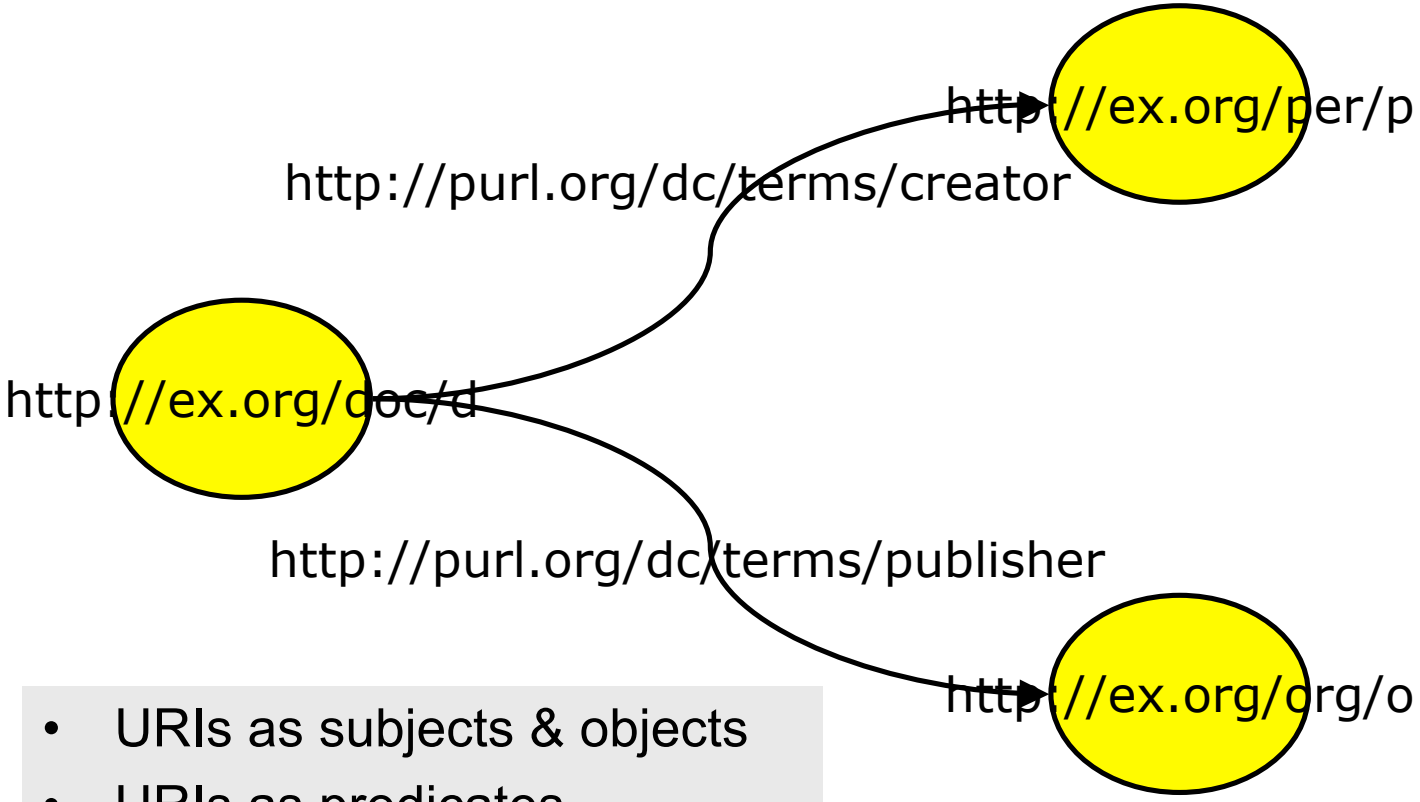
Resource Description Framework (RDF)



Resource Description Framework (RDF)

- RDF designed for use on Web
- Global context, global naming: URIs
  - URIs as names for things described (subject, object)
  - URIs as names of relationships between things (predicate)
  - URI-based extensibility
- Also allows for
  - "unnamed" things: "blank nodes" (subject, object)
  - "literals" = text strings (object only)
- Formal semantics defines rules for
  - comparing & merging sets of triples
  - logical inferencing

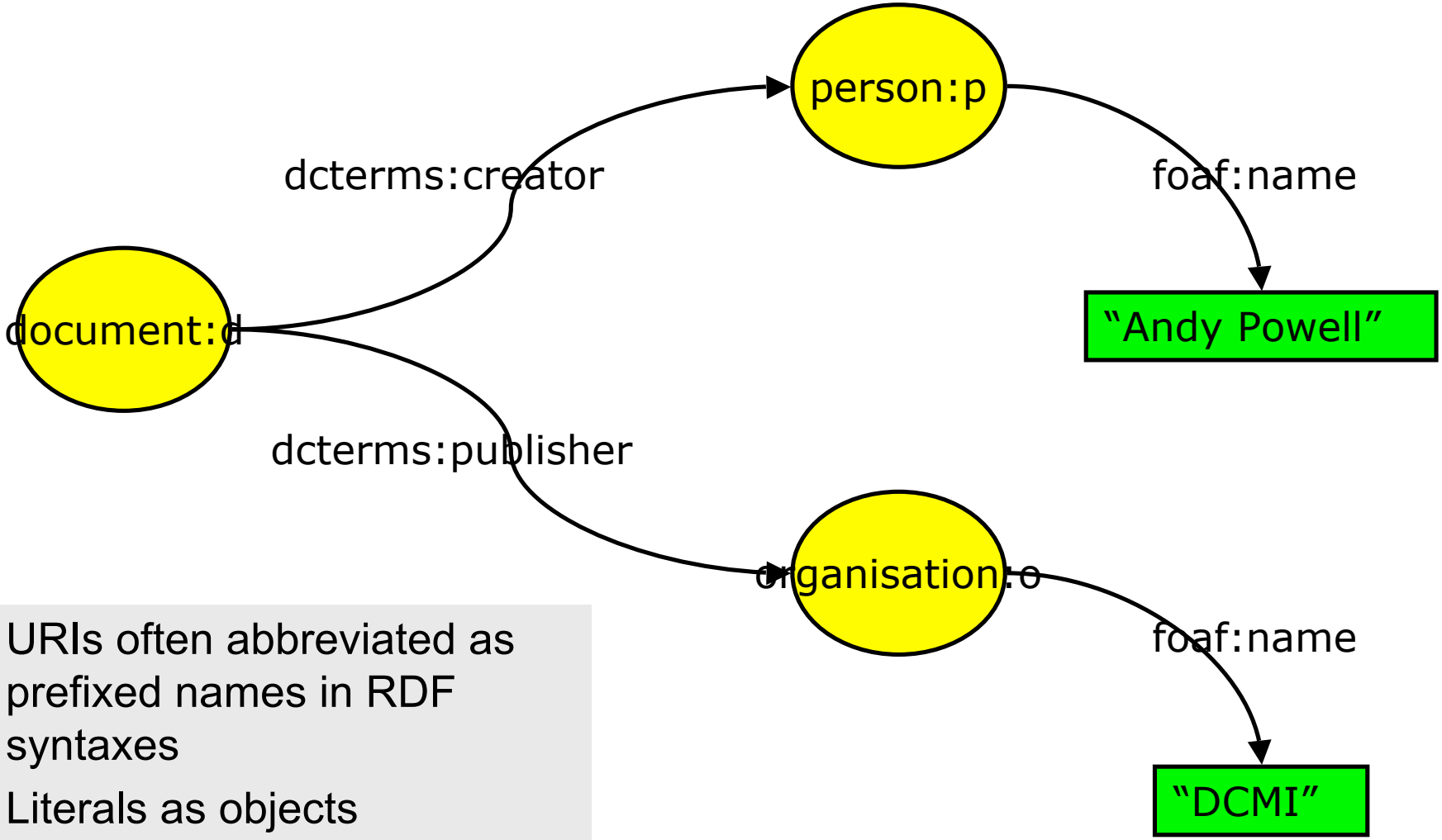
Resource Description Framework (RDF)



- URIs as subjects & objects
- URIs as predicates

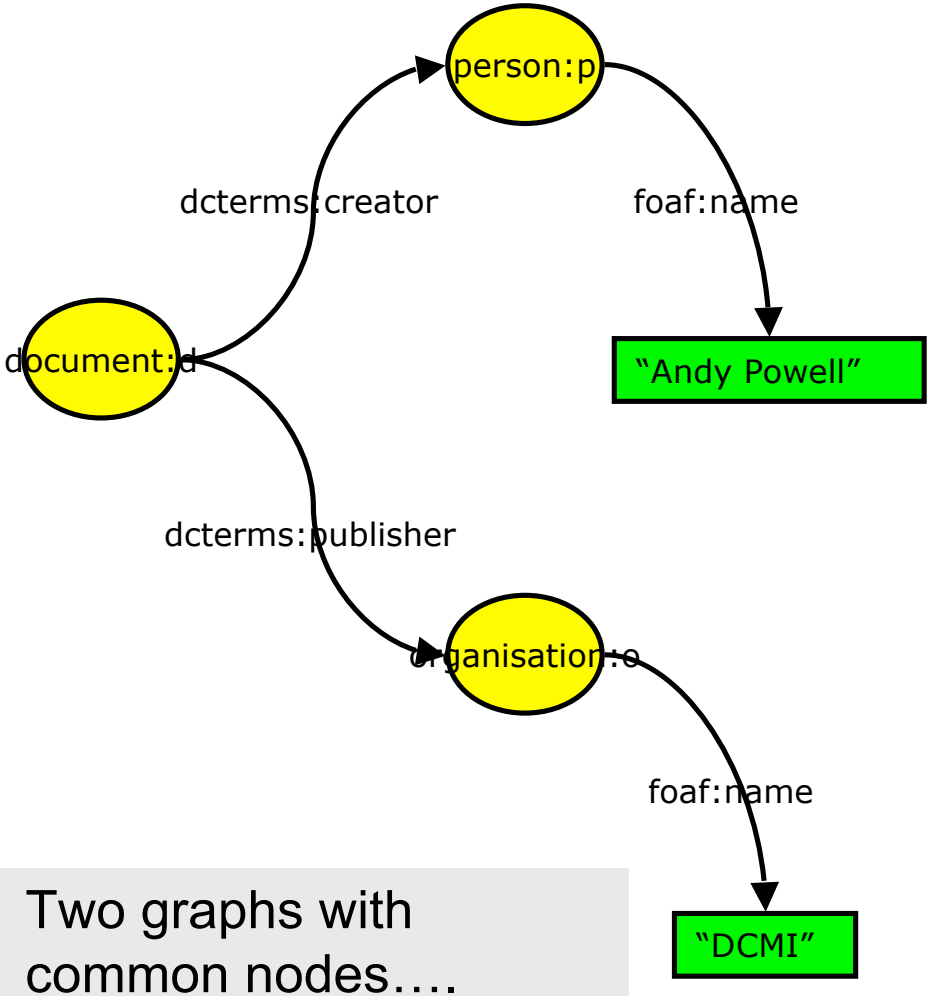
Subject	Predicate	Object
http://ex.org/doc/d	http://purl.org/dc/terms/creator	http://ex.org/per/p
http://ex.org/doc/d	http:// purl.org/dc/terms/publisher	http://ex.org/org/o

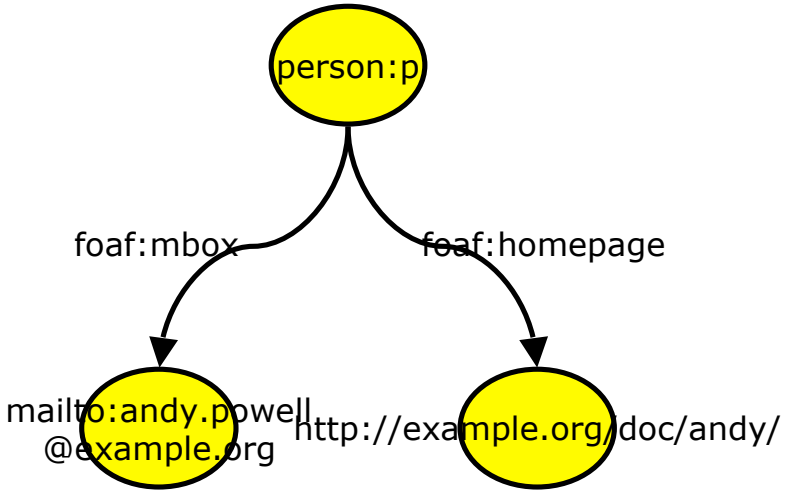
Resource Description Framework (RDF)



- URIs often abbreviated as prefixed names in RDF syntaxes
- Literals as objects

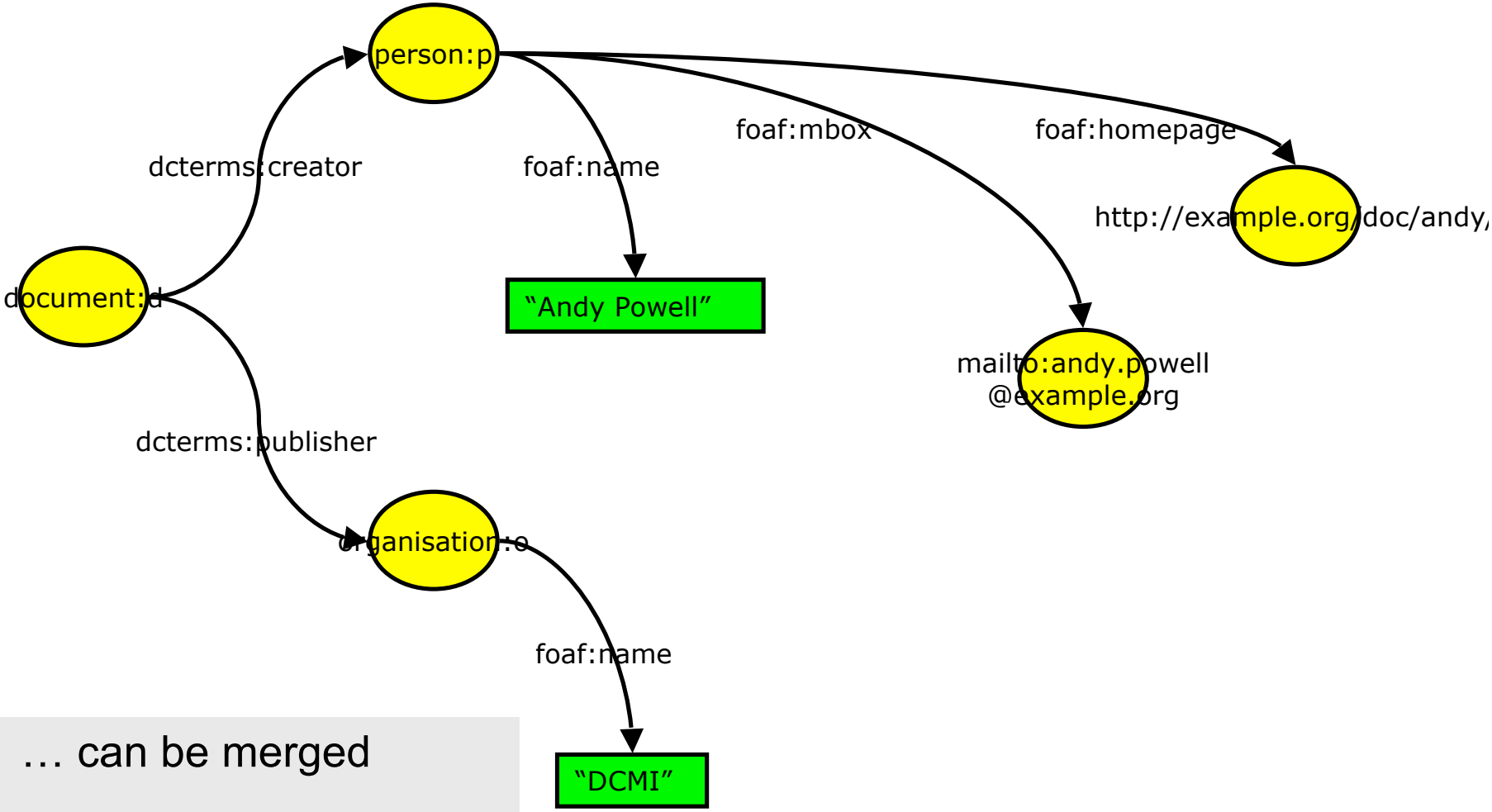
Resource Description Framework (RDF)





- Two graphs with common nodes….

Resource Description Framework (RDF)

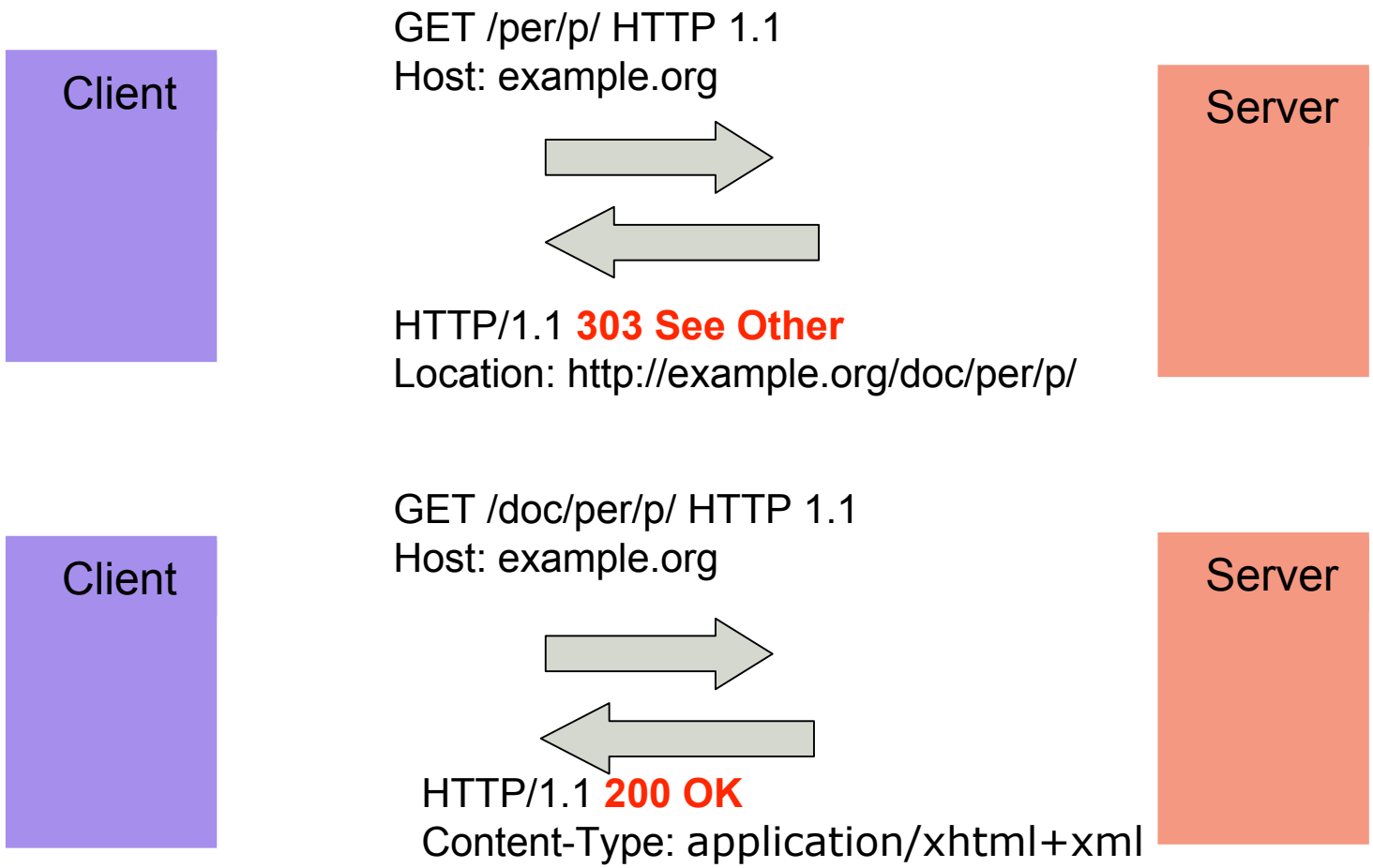


- … can be merged

Web Architecture: httpRange-14

- Widespread use of http URIs as names for things other than documents, while also serving representation
- Problem: does URI identify thing or document?
- Solution: the W3C TAG httpRange-14 resolution
  - If server responds to GET with 2xx, then resource is information resource (document)
  - If server responds to GET with 303, then resource could be anything
- Make a distinction between
  - information resources (Web documents), with URI, with representation
  - "non-information resources" ("real world things"), with URI, with no representation, but with description
- Sauermann, Cyganiak, *Cool URIs for the Semantic Web*
  - W3C Interest Group Note, 31 March 2008 http://www.w3.org/TR/2008/NOTE-cooluris-20080331/

W3C Web Architecture: httpRange-14



Linked Data

- Berners-Lee's Linked Data principles
  - Design Issues Note (2006) http://www.w3.org/DesignIssues/LinkedData.html
  1. Use URIs as names for things.
  2. Use http URIs so that people can look up those names.
  3. When someone looks up a URI, provide useful information.
  4. Include links to other URIs. So that they can discover more things

How to Publish Linked Data on the Web

- Bizer, Cyganiak, Heath: *How to Publish Linked Data on the Web* (2007)
  - http://sites.wiwiss.fu-berlin.de/suhl/bizer/pub/LinkedDataTutorial/
- Provides guidelines on
  - subset of RDF features (for merging & querying)
  - choosing URIs, linking and using URI aliases
  - "useful information to return"
    - description, backlinks, related descriptions, metadata
  - recipes for serving data

Open Archives Initiative Object Reuse & Exchange (OAI ORE)

With thanks to Herbert Van de Sompel (Los Alamos National Laboratory, USA) for several of the slides used in this section

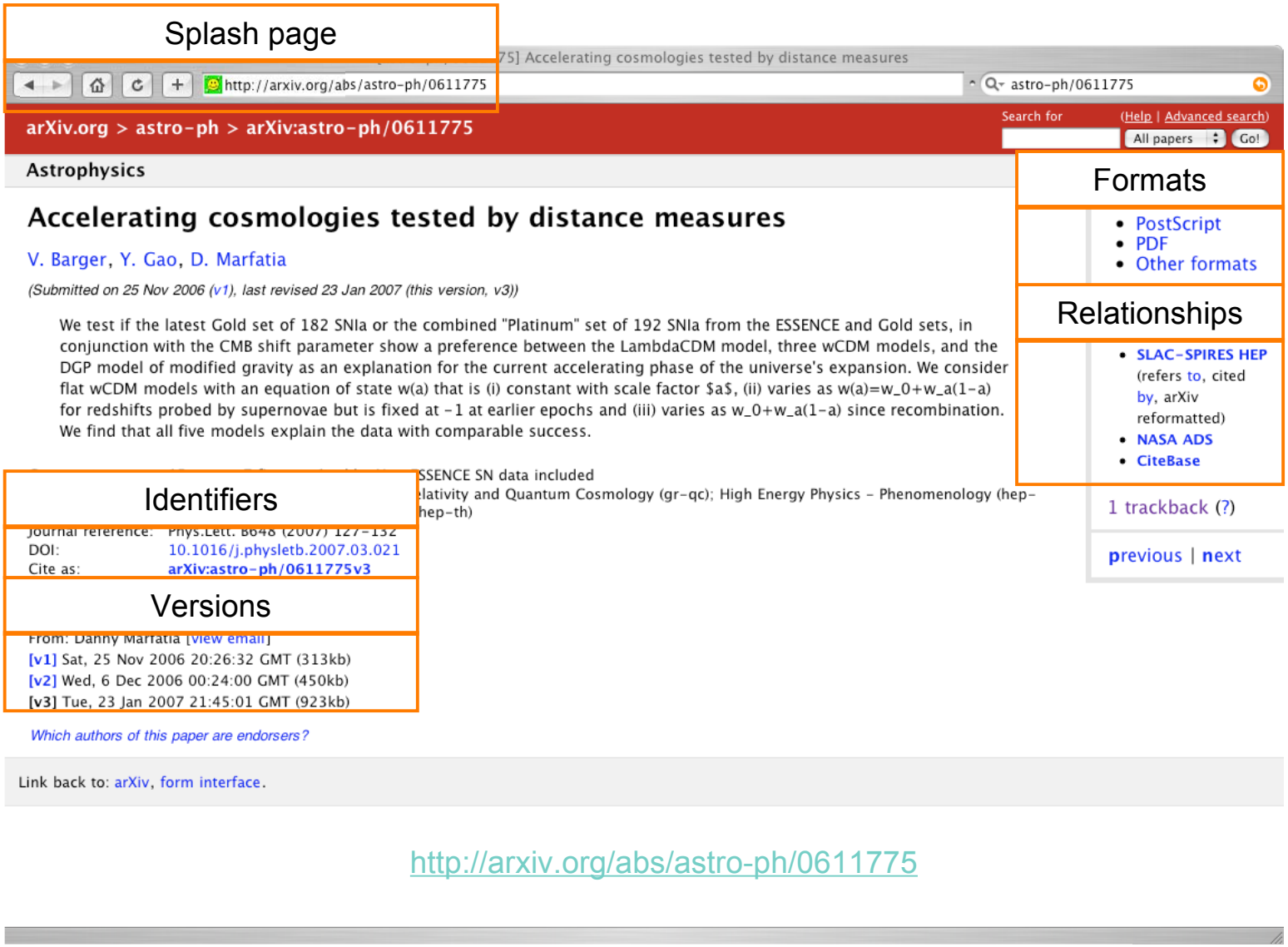
Open Archives Initiative Object Reuse & Exchange (ORE)

eduserv

- Project of Open Archives Initiative
  - http://www.openarchives.org/ore/
- Funded by
  - The Andrew W. Mellon Foundation
- Additional support from
  - The Coalition for Networked Information
  - Joint Information Systems Committee
  - Microsoft Corporation
  - The National Science Foundation
- Co-ordinated by Carl Lagoze & Herbert Van de Sompel

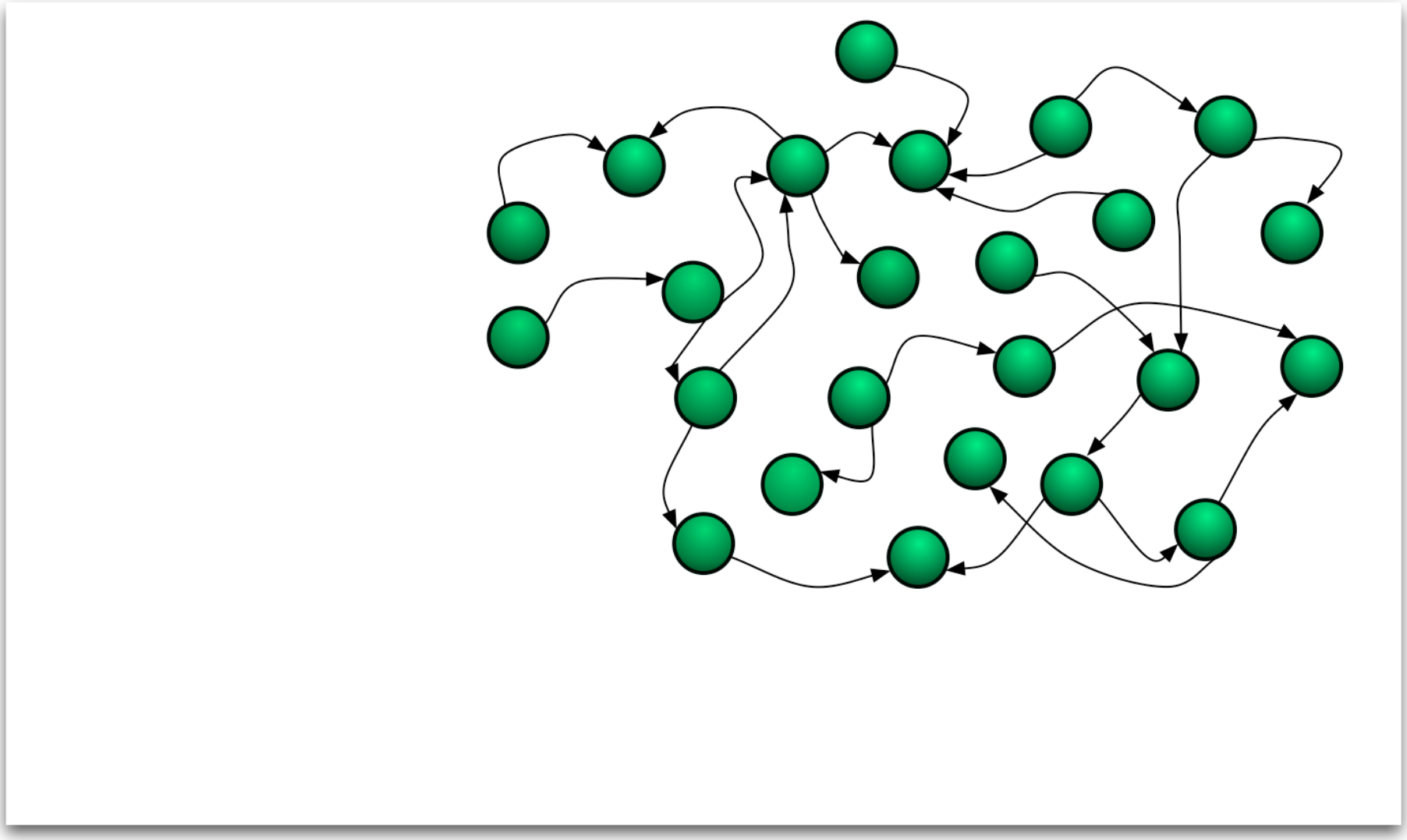
The problem ORE tries to solve

- Often useful to conceptualise (scholarly and other) resources as grouped into aggregations
- Typically, way in which resources made available on Web
  - does not make aggregate nature of resource explicit
  - presents component/member resources as linked in some way, but without making clear how related
  - introduces ambiguity between identification of set and identification of component/member



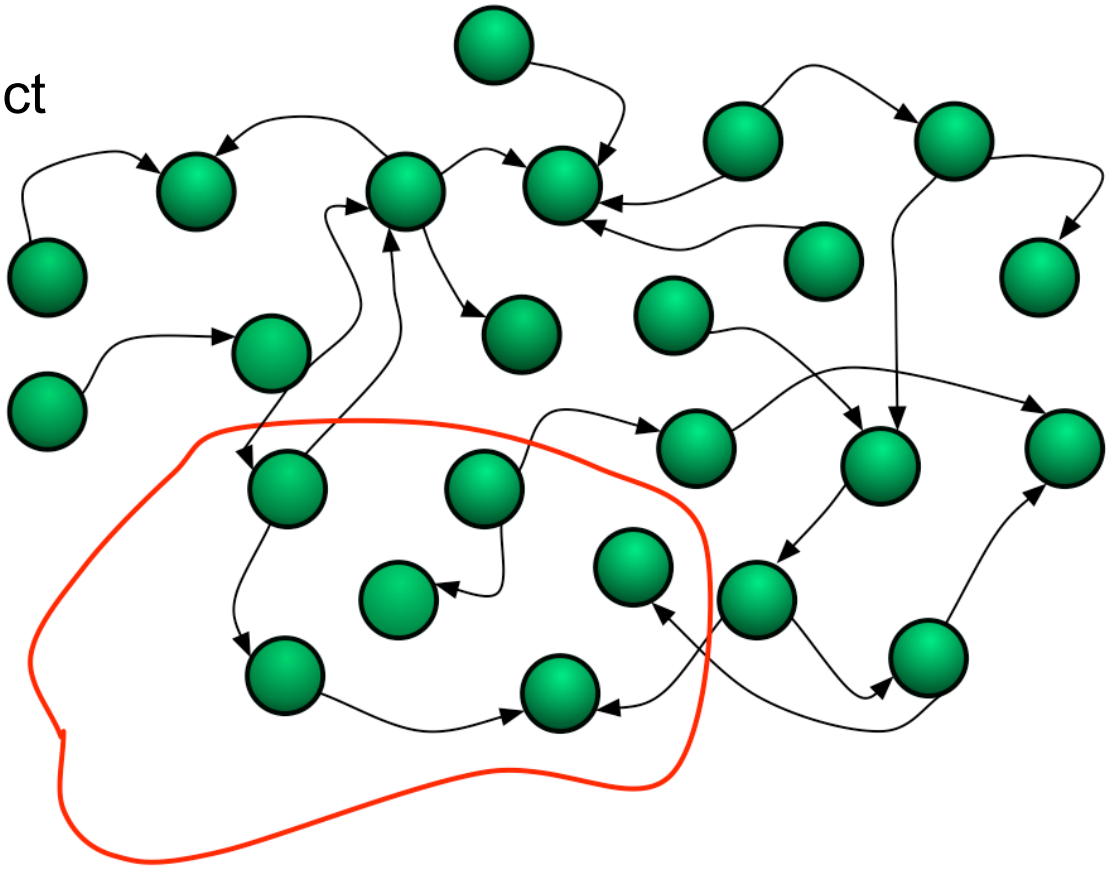
http://arxiv.org/abs/astro-ph/0611775

The Web



An Aggregation and the Web

- Resources of an Aggregation are distinct URI-identified Web resources

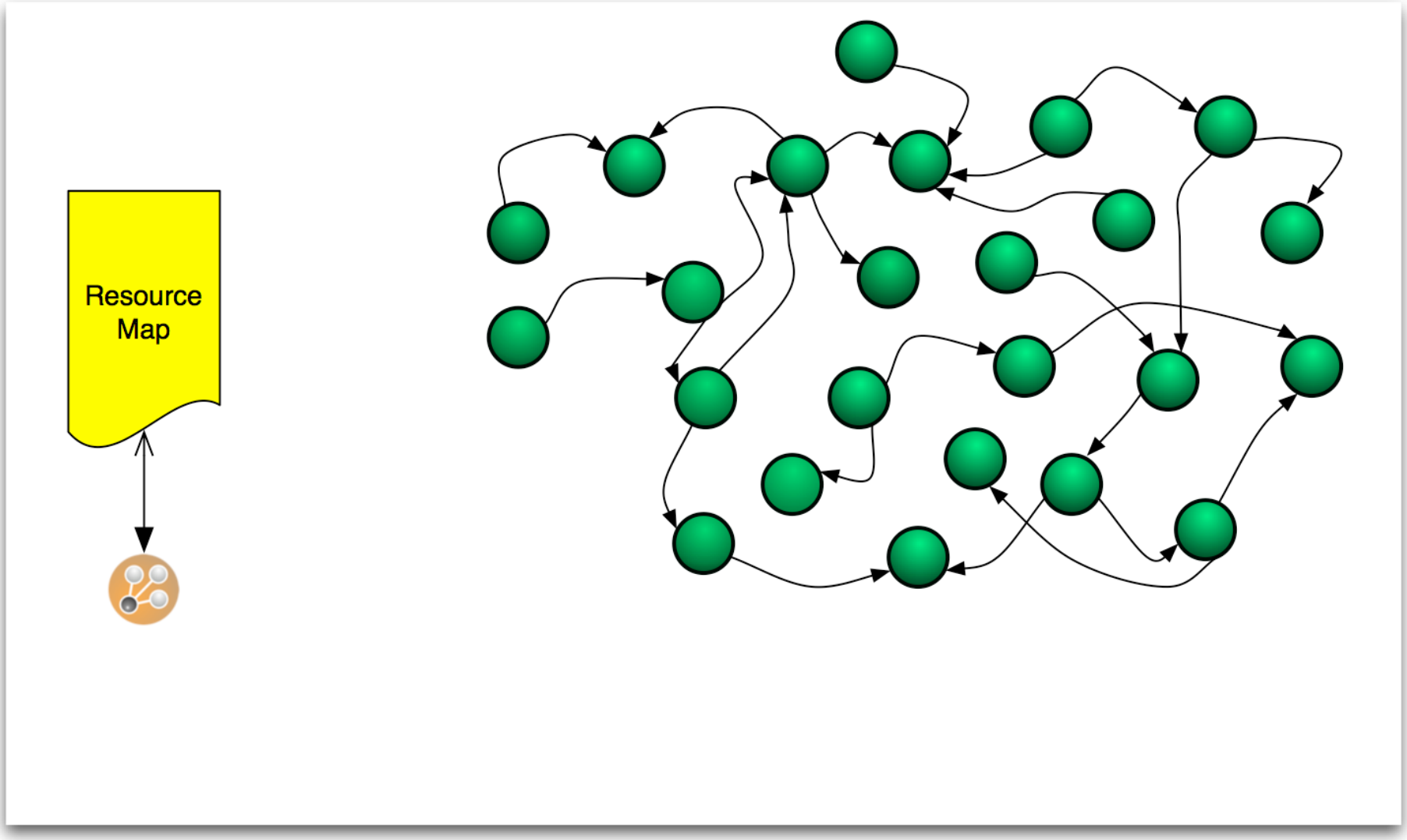


- Missing are:
  - The boundary that delineates the Aggregation in the Web
  - An identity (URI) for the Aggregation

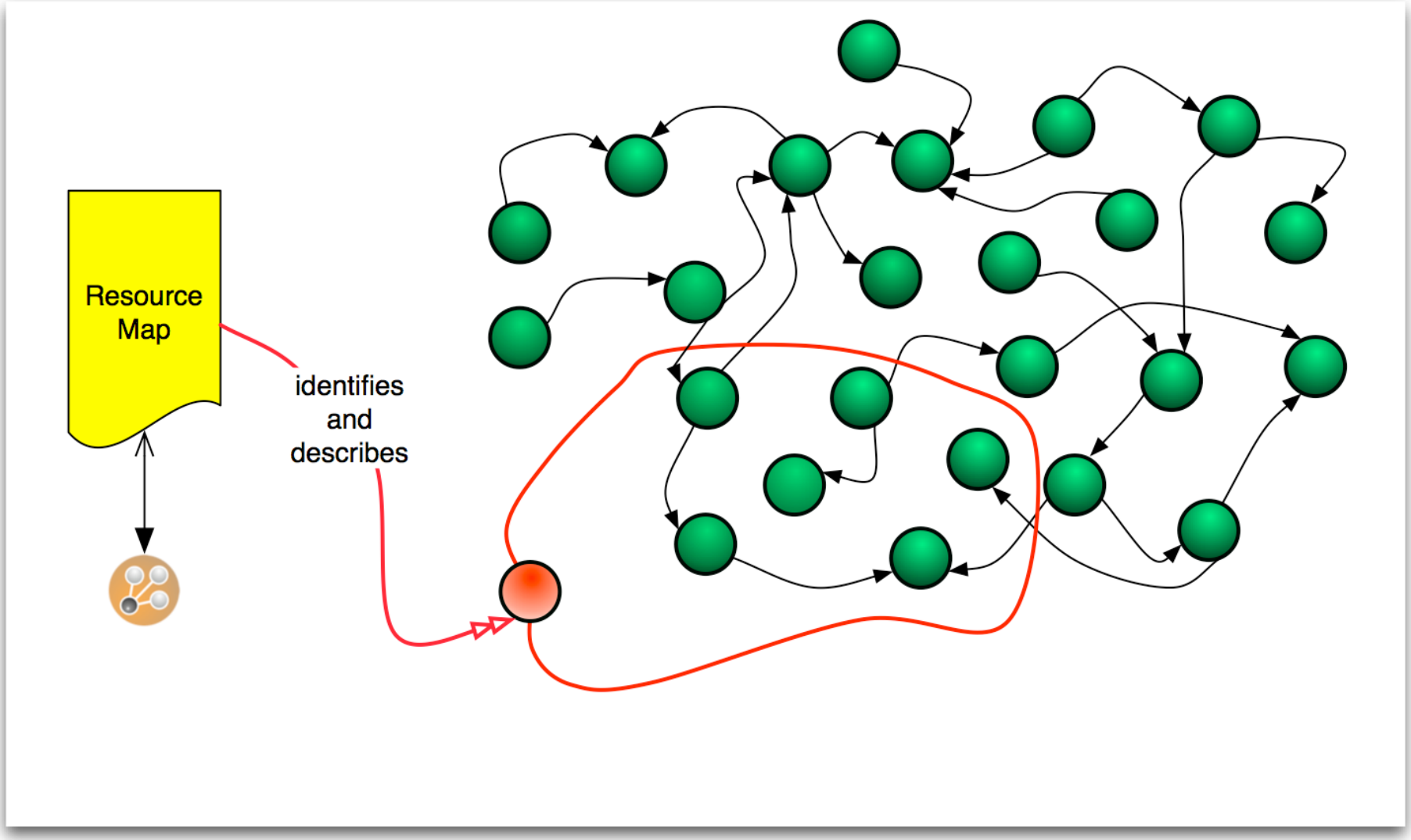
The ORE approach

- Make Aggregations resources, identified by URIs
- Create RDF-based descriptions of Aggregations (& their component resources)
  - relationships between Aggregation & component resources
  - attributes of Aggregation, component resources
  - relationships between Aggregation, component resources and other resources
- Make Resource Maps resources, identified by URIs
  - include attributes of Resource Map
  - relationships between Resource Map and other resources
- Make Resource Maps available using conventions of Web Arch/Cool URIs

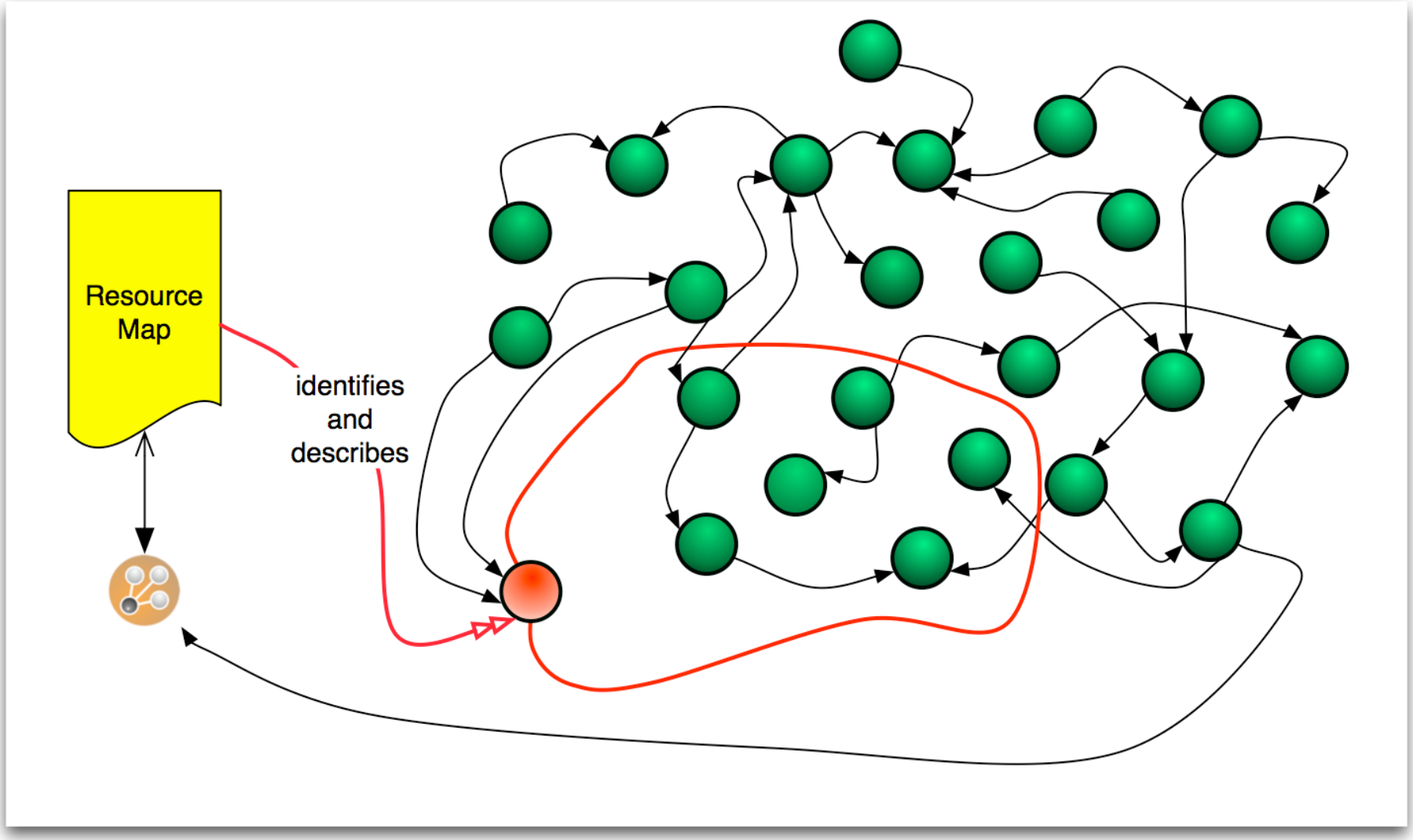
Publish a Resource Map to the Web



The Resource Map Identifies and Describes the Aggregation



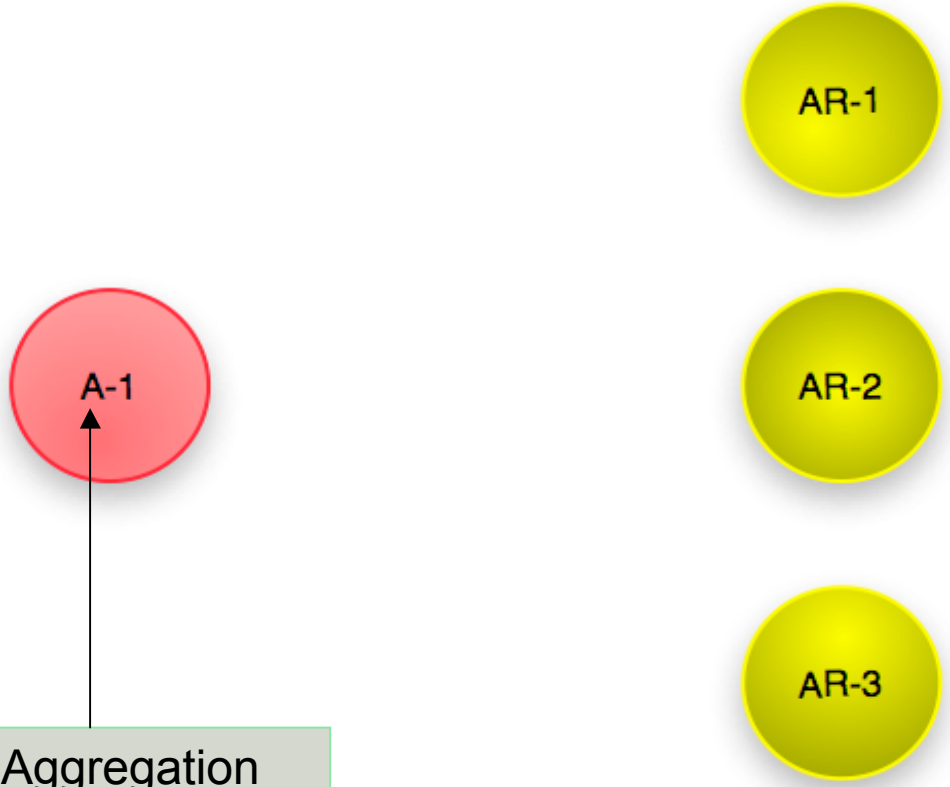
The Resource Map and Aggregation as (distinct) Web resources



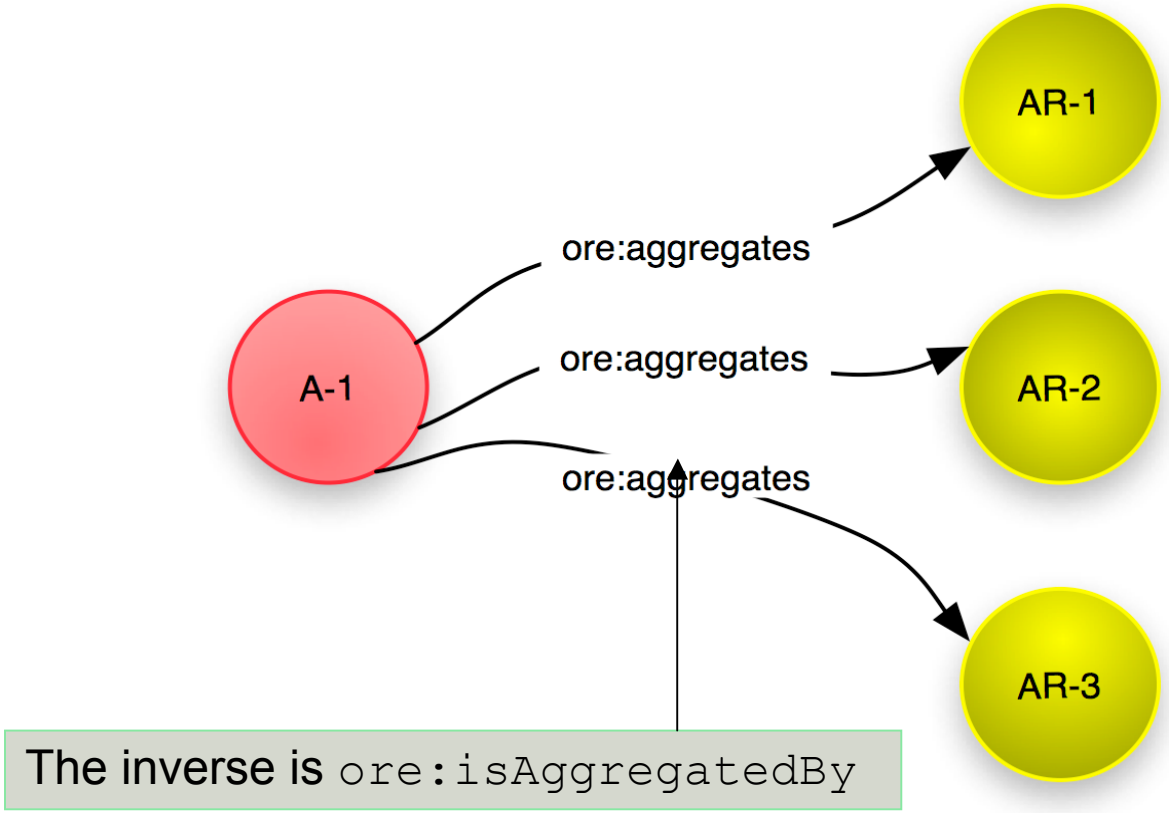
Constructing an ORE Resource Map

Some resources that belong together

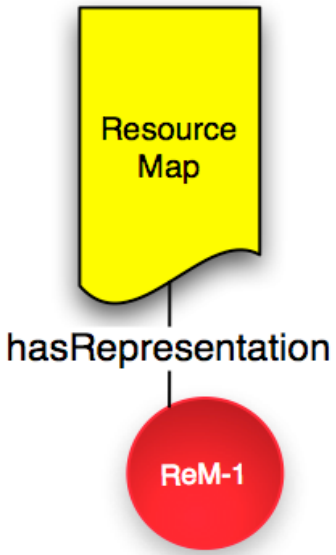
Introduce the Aggregation

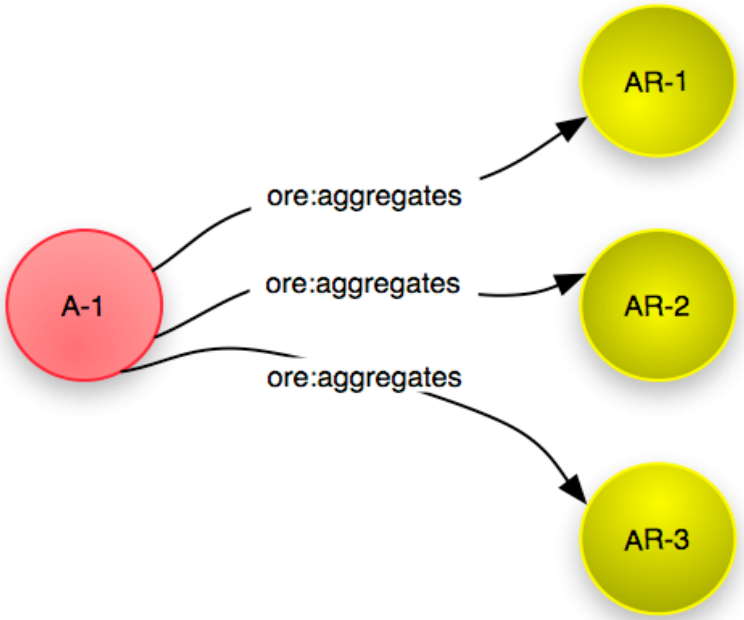


Express the `ore:aggregates` relationship

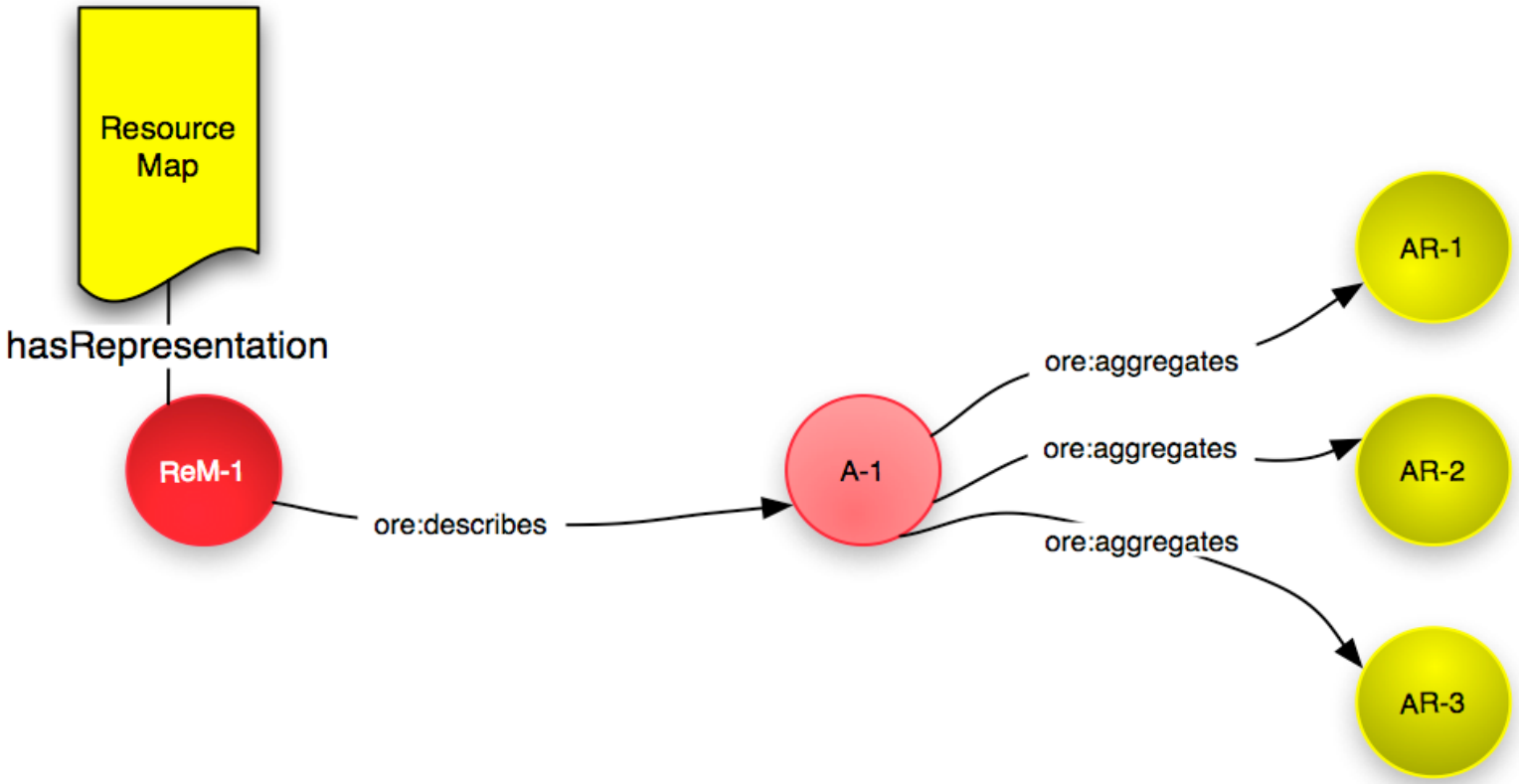


Introduce the Resource Map

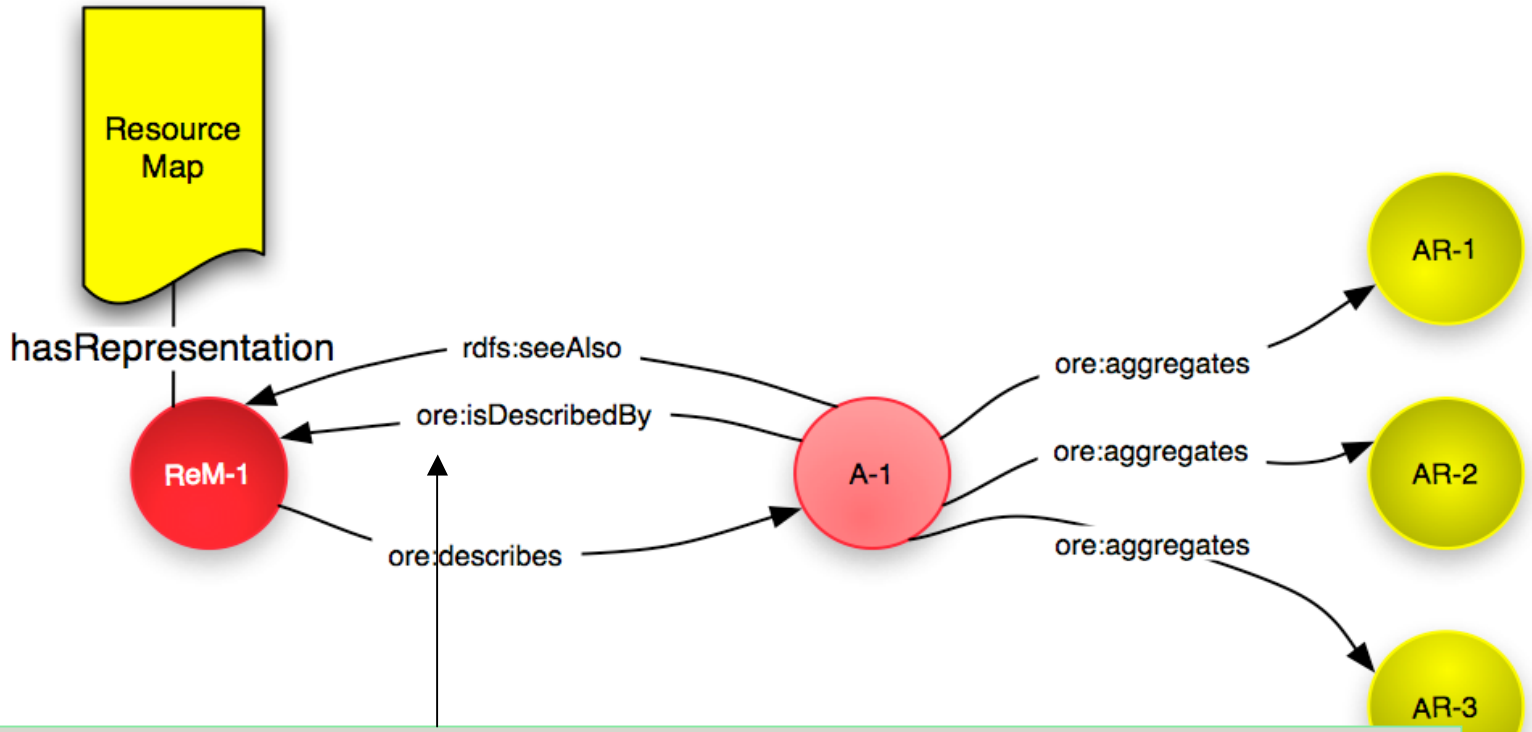




Express the `ore:describes` relationship

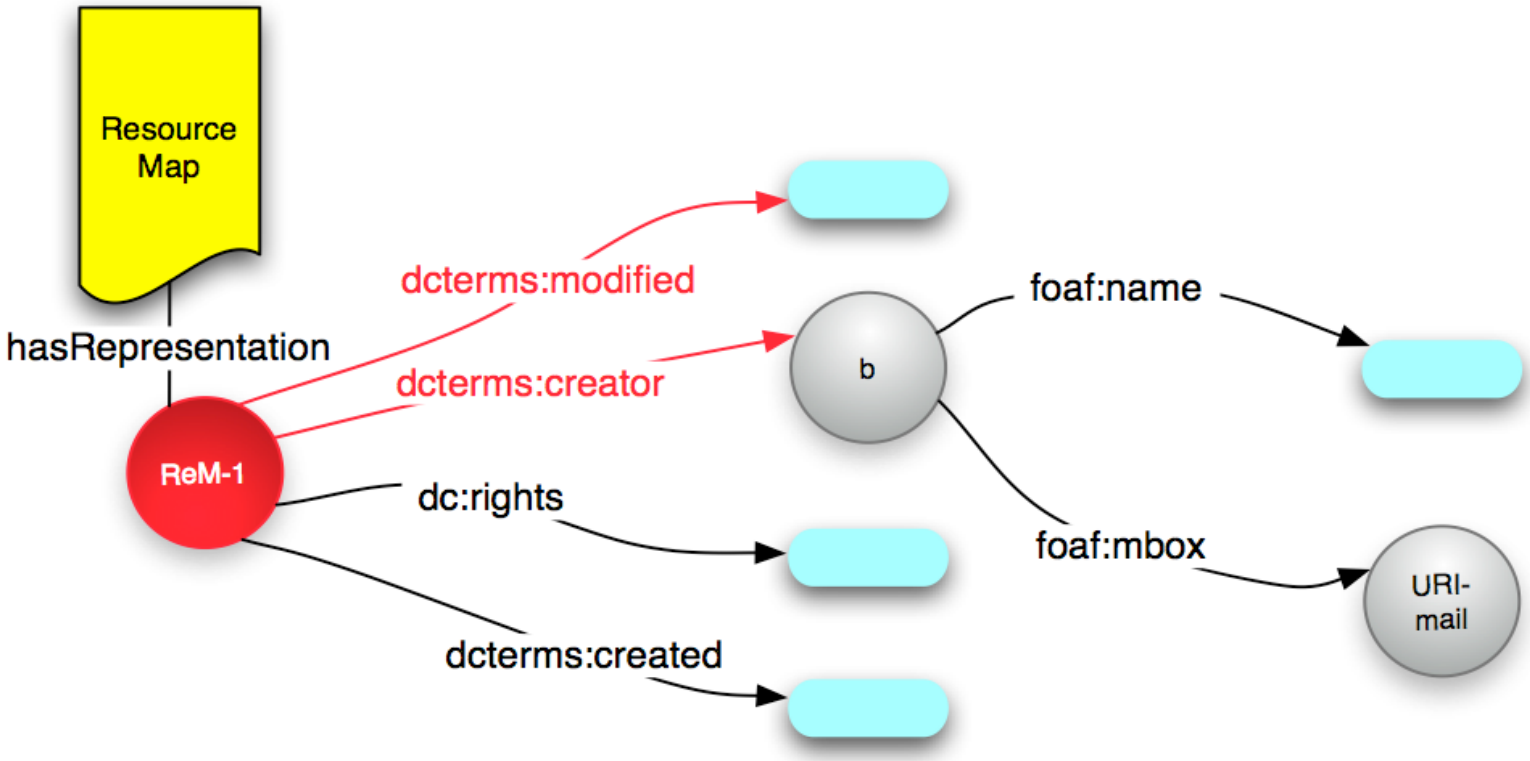


The `ore:isDescribedBy` relationship



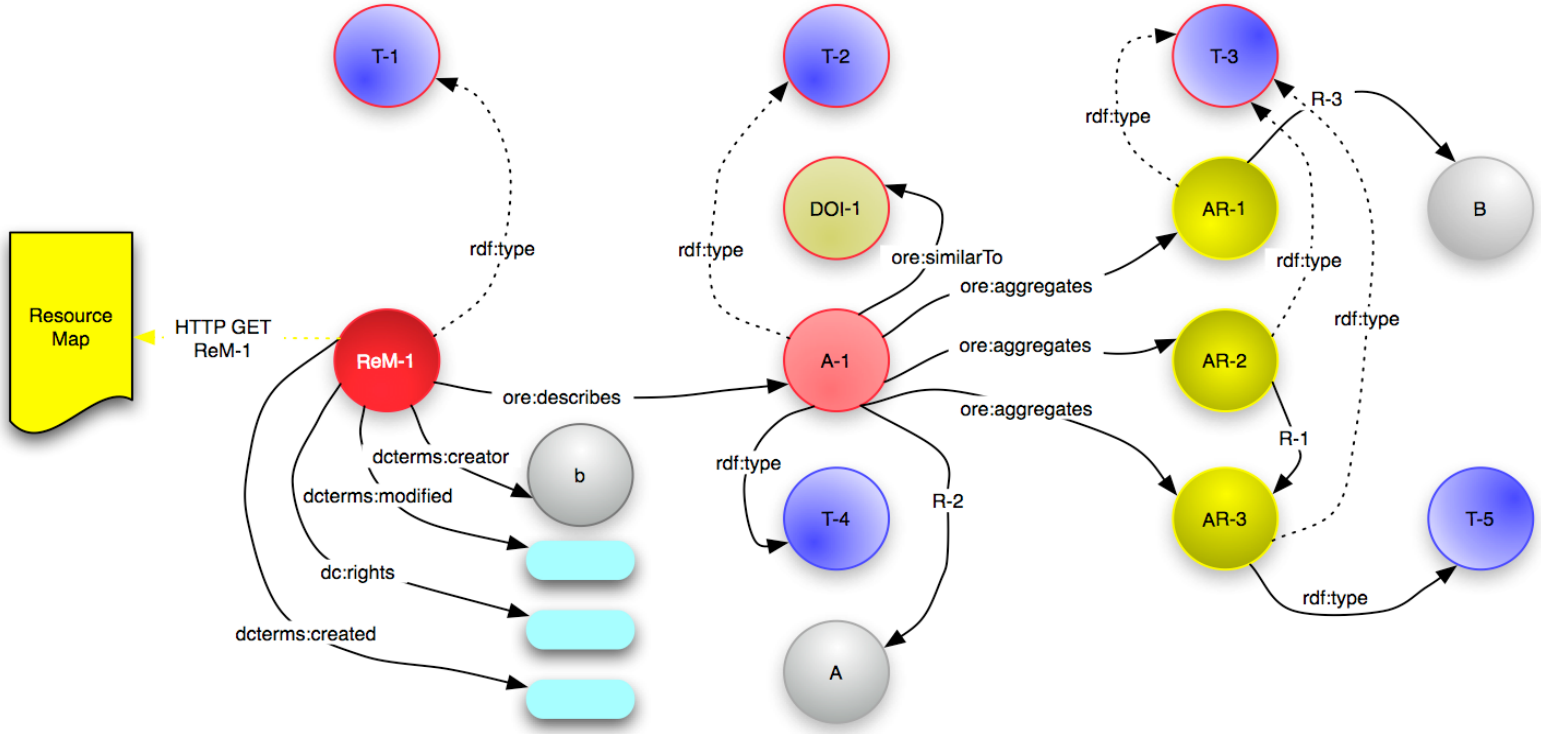
The inverse is `ore:isDescribedBy`; subproperty of `rdfs:seeAlso`

Express metadata about the Resource Map



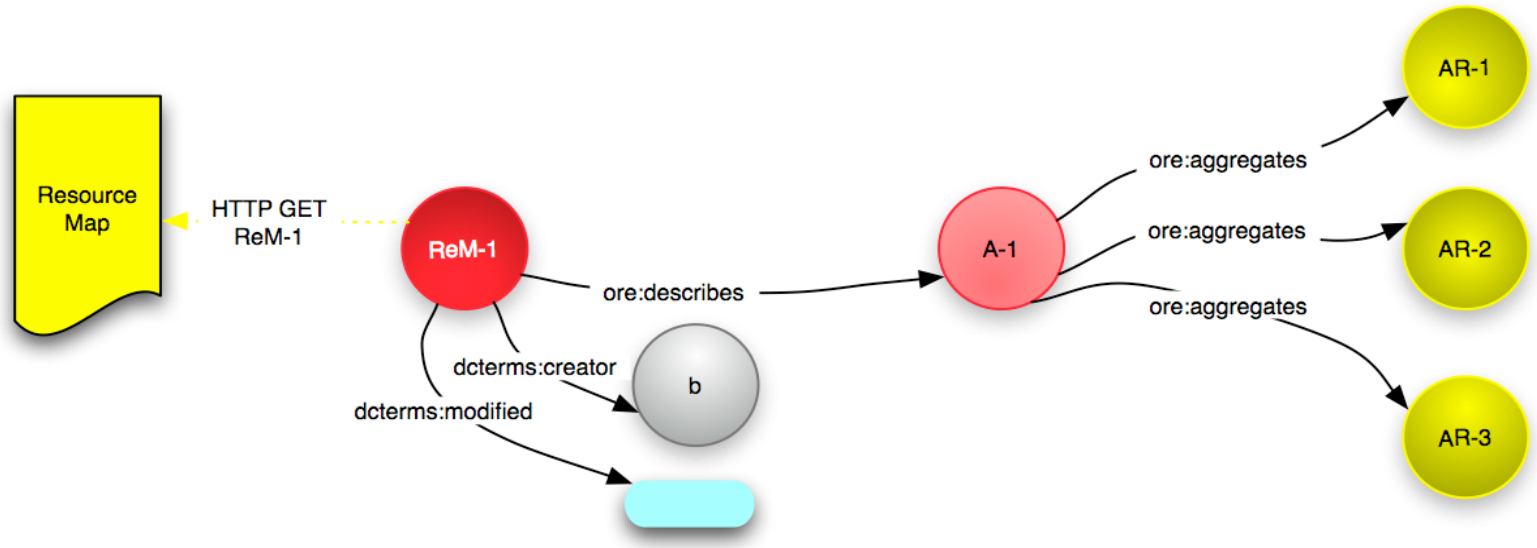
This corresponds to **metadata** from the Linked Data recommendations

A Resource Map can “say” a lot …



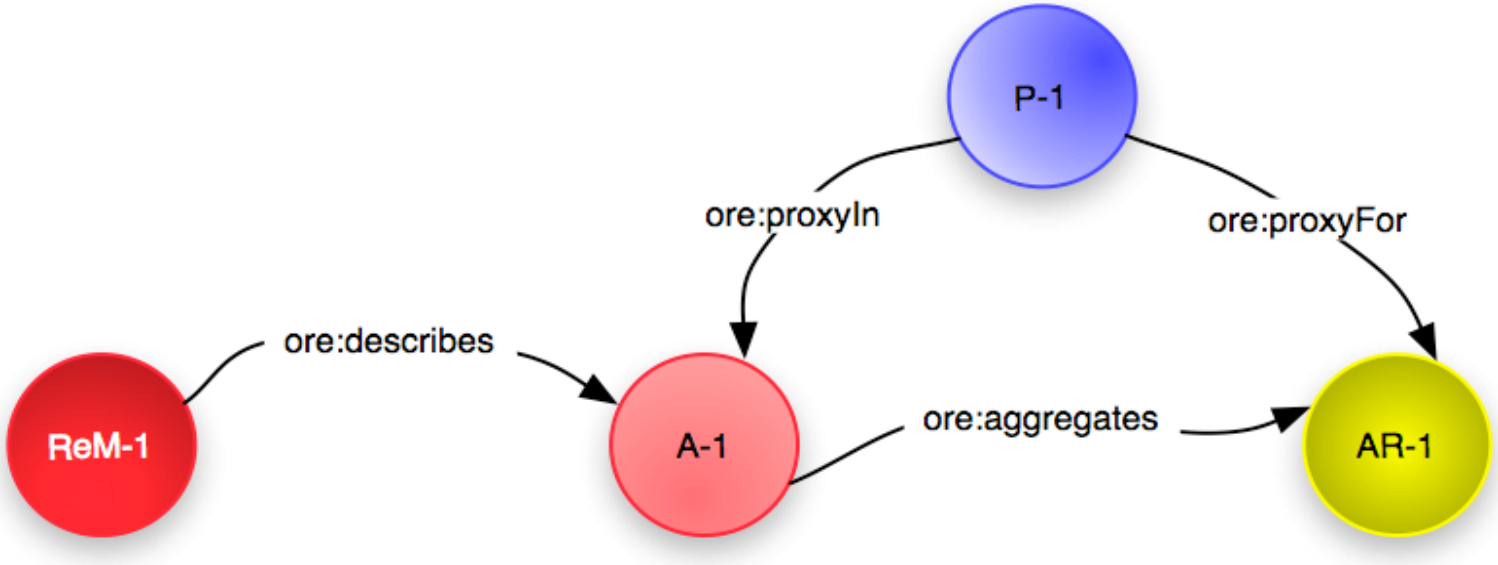
This corresponds to the **description**, **related descriptions**, **backlinks**, **metadata** from the Linked Data recommendations

But minimally it “says” this …



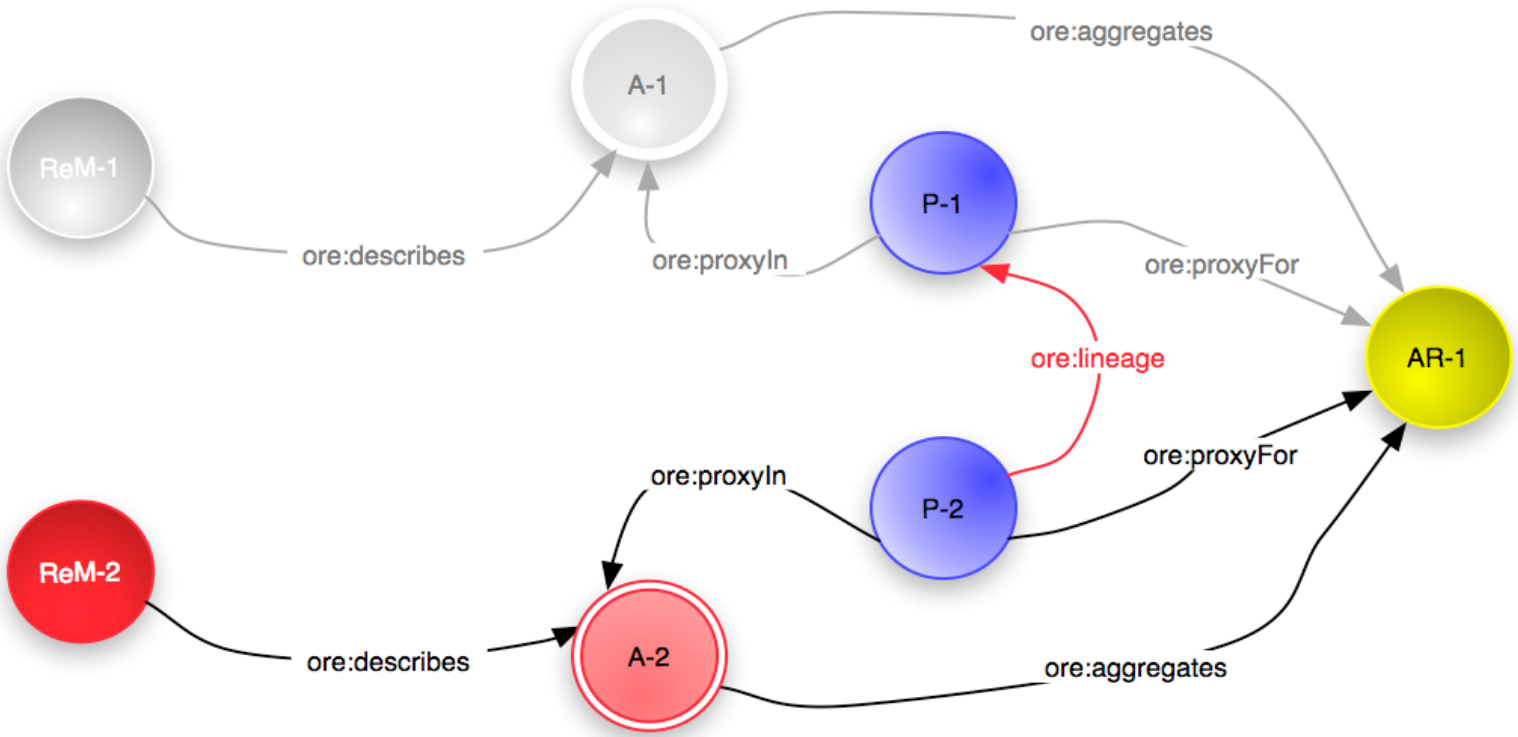
This corresponds to the **description** (minimal), and **metadata** from the Linked Data recommendations

Proxy: a Resource in the Context of an Aggregation



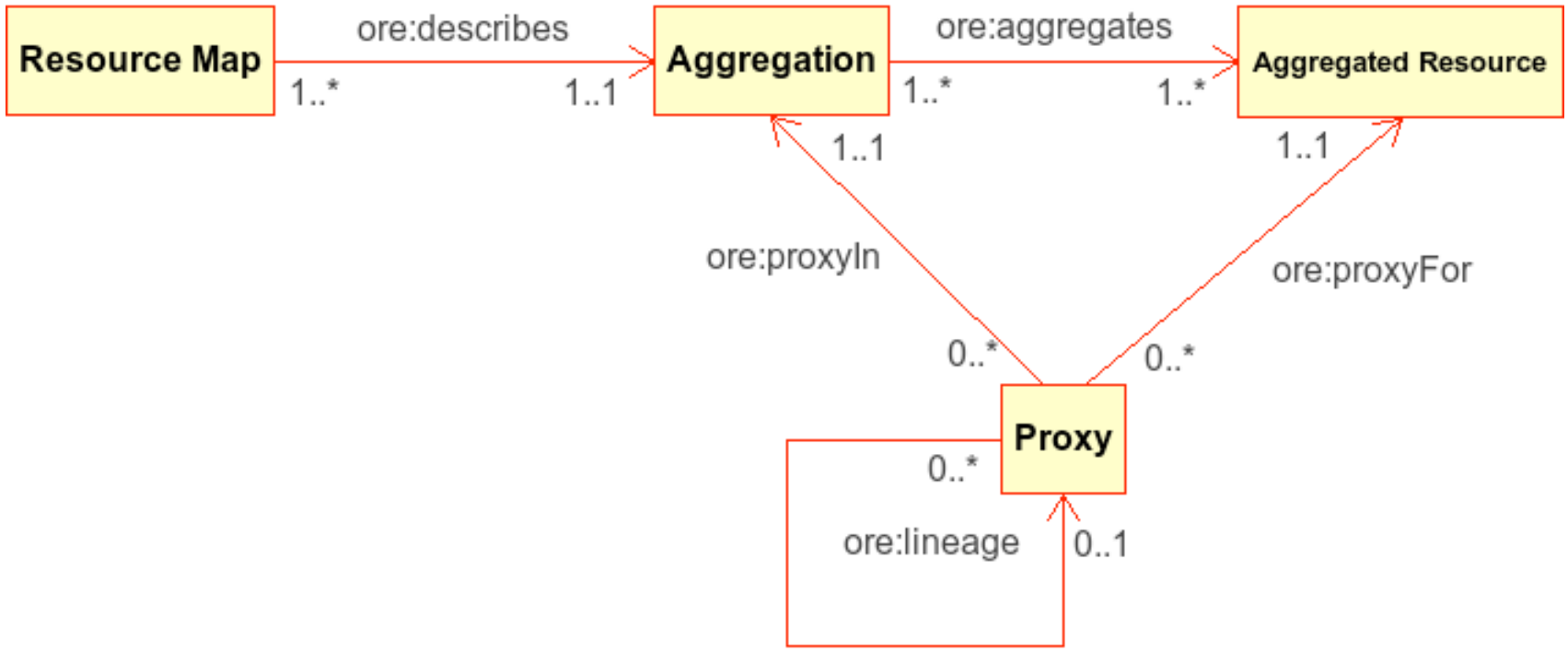
`ore:proxyFor` and `ore:proxyIn` to introduce a Proxy for an Aggregated Resource

ore:lineage An Aggregated Resource *originated* in another Aggregation



ore:lineage is a relationship between Proxies

The ORE model in summary



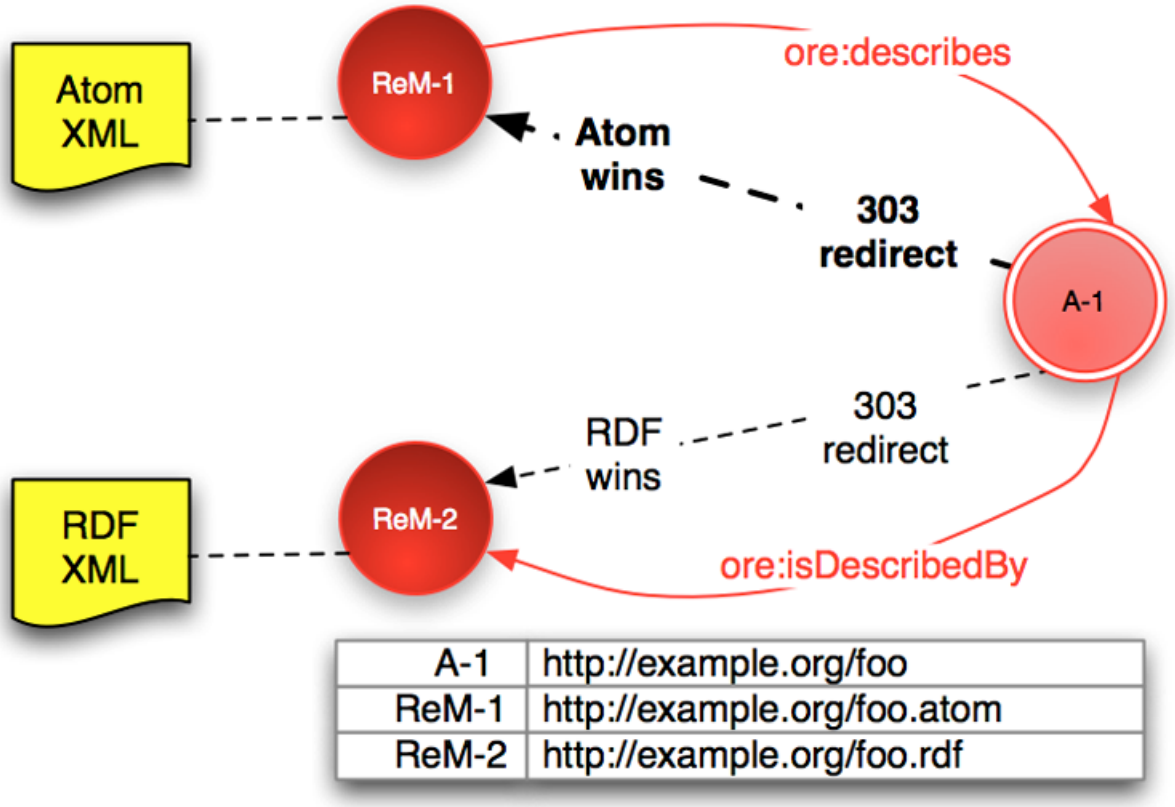
Representing an ORE Resource Map

- An ORE Resource Map is an RDF Graph, so can be represented using any RDF syntax
  - ORE provides guidelines for using
    - RDF/XML
    - RDFa
- Also guidelines for representing ORE Resource Map as Atom Entry Document
  - With GRDDL transformation available

Publishing an ORE Resource Map

- ORE Aggregation as “non-information resource”/”real-world thing”
- ORE Resource Map as “information resource”/”Web document”
- ORE recommends selected HTTP “recipes” from *Cool URIs for the Semantic Web*
  - “Hash URI”
  - Simple 303 redirect to URI of Resource Map
  - Content negotiation + 303 redirect to URI of Resource Map

Content negotiation with 303 redirect



A-1	http://example.org/foo
ReM-1	http://example.org/foo.atom
ReM-2	http://example.org/foo.rdf

Summary

- ORE addresses problem of describing Aggregations of resources
- Tries to do so in "resource-centric" way (c.f. OAI-PMH)
- Tries to cater for both Semantic Web contexts and Atom contexts
- Agnostic about
  - nature of aggregated resources
  - other relationships between aggregations, aggregated resources
- The least we can specify whilst being useful" (Jim Downing)
- Generic; potentially, broadly applicable
- Will typically be used alongside other vocabularies, models?

Applications in e-Science domain

- Applicable to any Aggregation!
- e.g.
  - Transfer/migration/archiving of “packages” of resources?
  - Collected outputs of activity as Aggregations?
  - Thesis + datasets as Aggregations (THEOREM project, Cambridge)
  - Papers + chemistry data (ORECHEM)
  - And more….

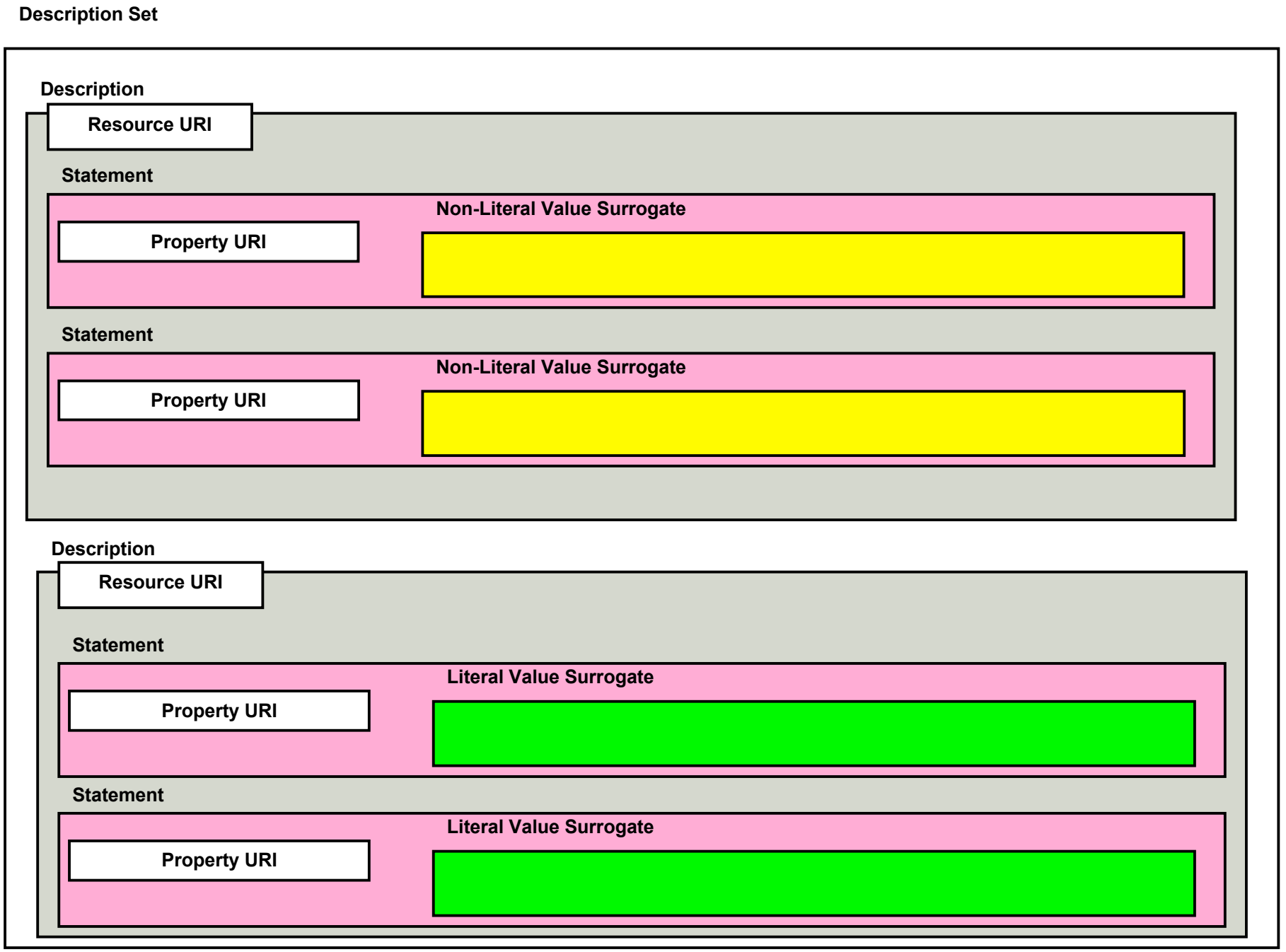
More Background: The DCMI Abstract Model and Dublin Core Application Profiles

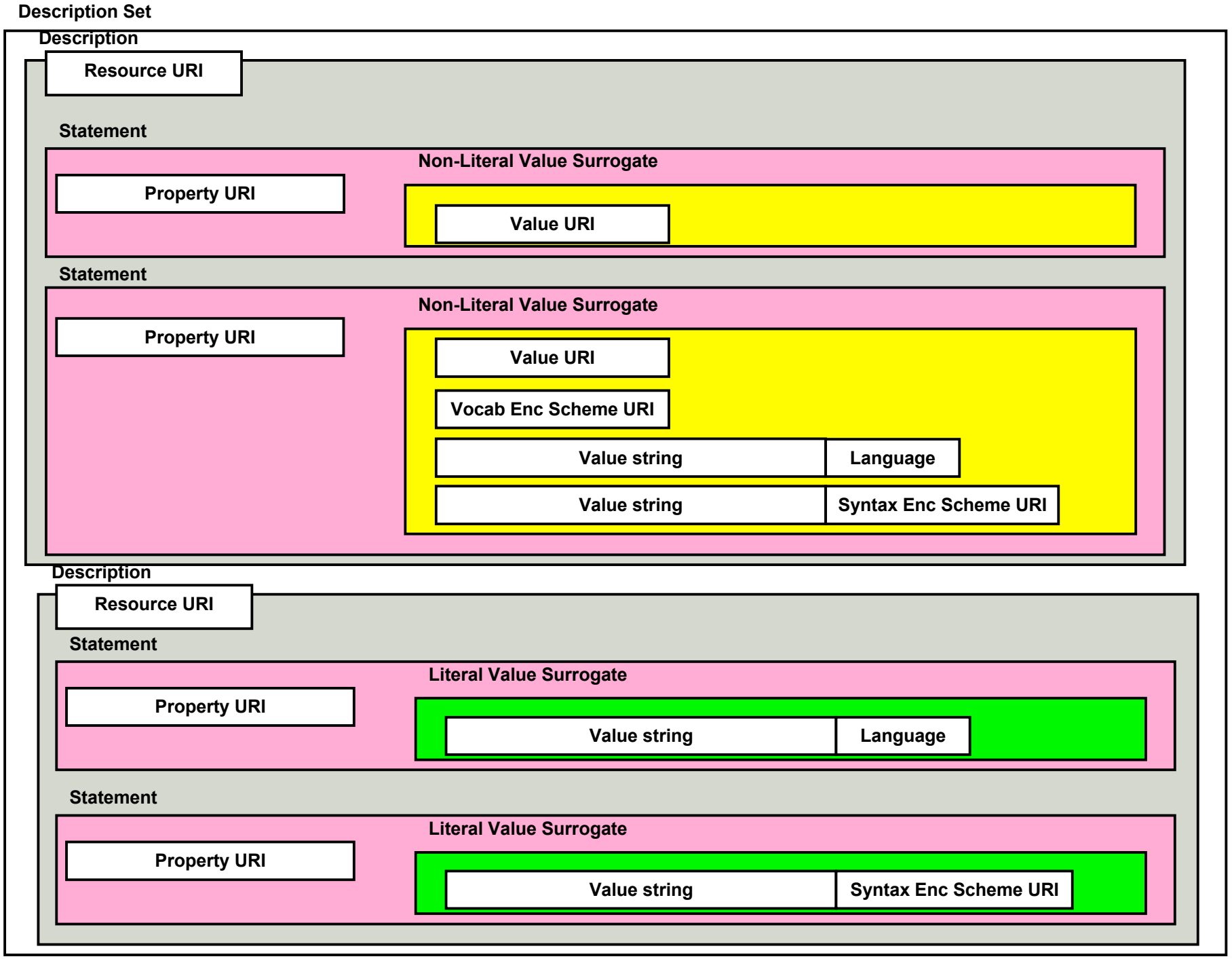
Why DCAM? Dublin Core in c2003

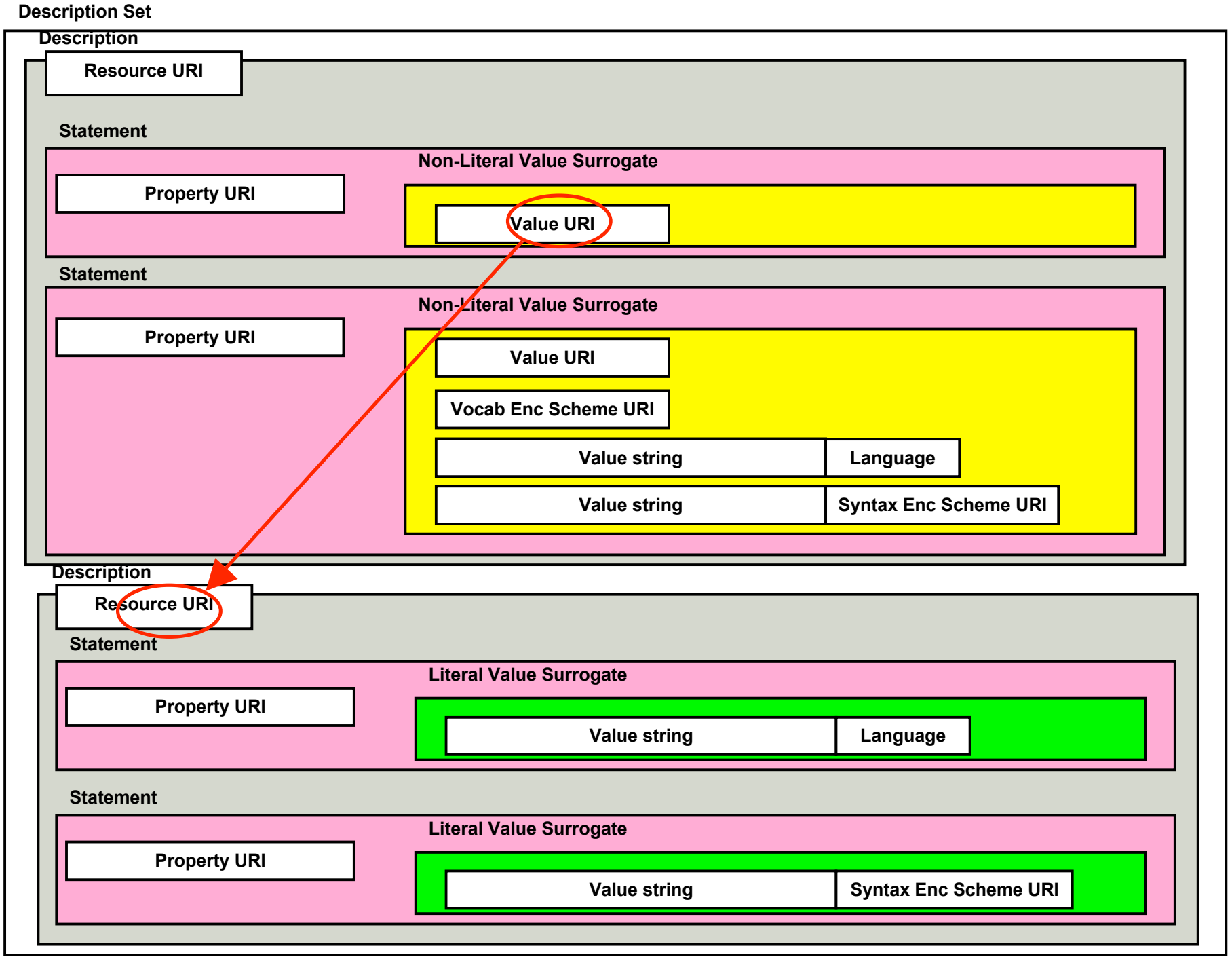
- Metadata vocabularies
  - … but what is a DC "element"?
  - … what sort of "terms" are used in DC metadata?
- Syntax independence & encoding guidelines
  - … but what is being "encoded"?
- "Dublin Core"? "Simple" and "Qualified" DC
  - … vocabularies?
  - … formats? (e.g. oai_dc)
  - … constraints on use of vocabularies? On which vocabularies?
- Grammatical Principles (including "1-to-1 rule")
  - … fairly informal
- Relationship between DC & Resource Description Framework
  - history of co-evolution

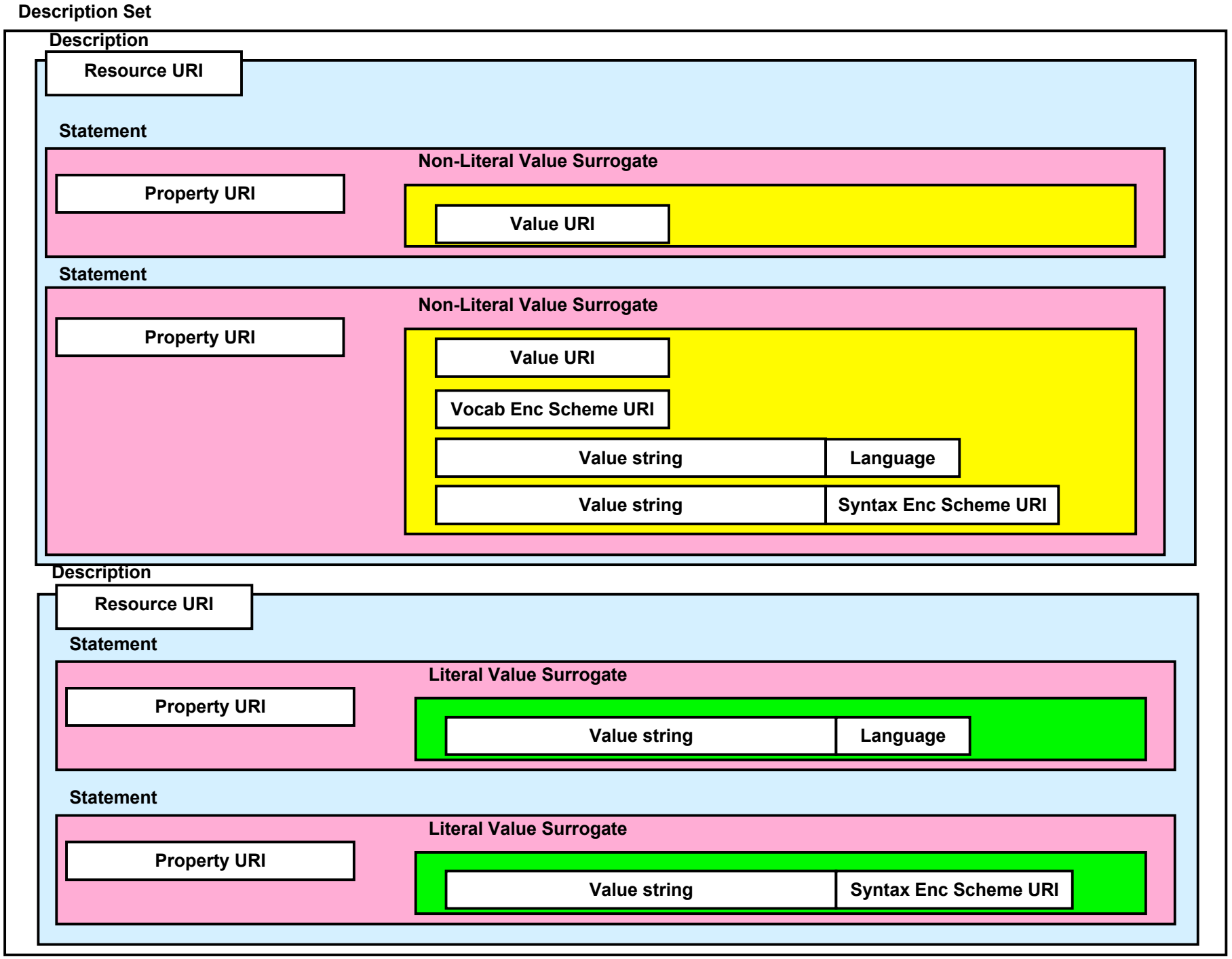
DCMI Abstract Model

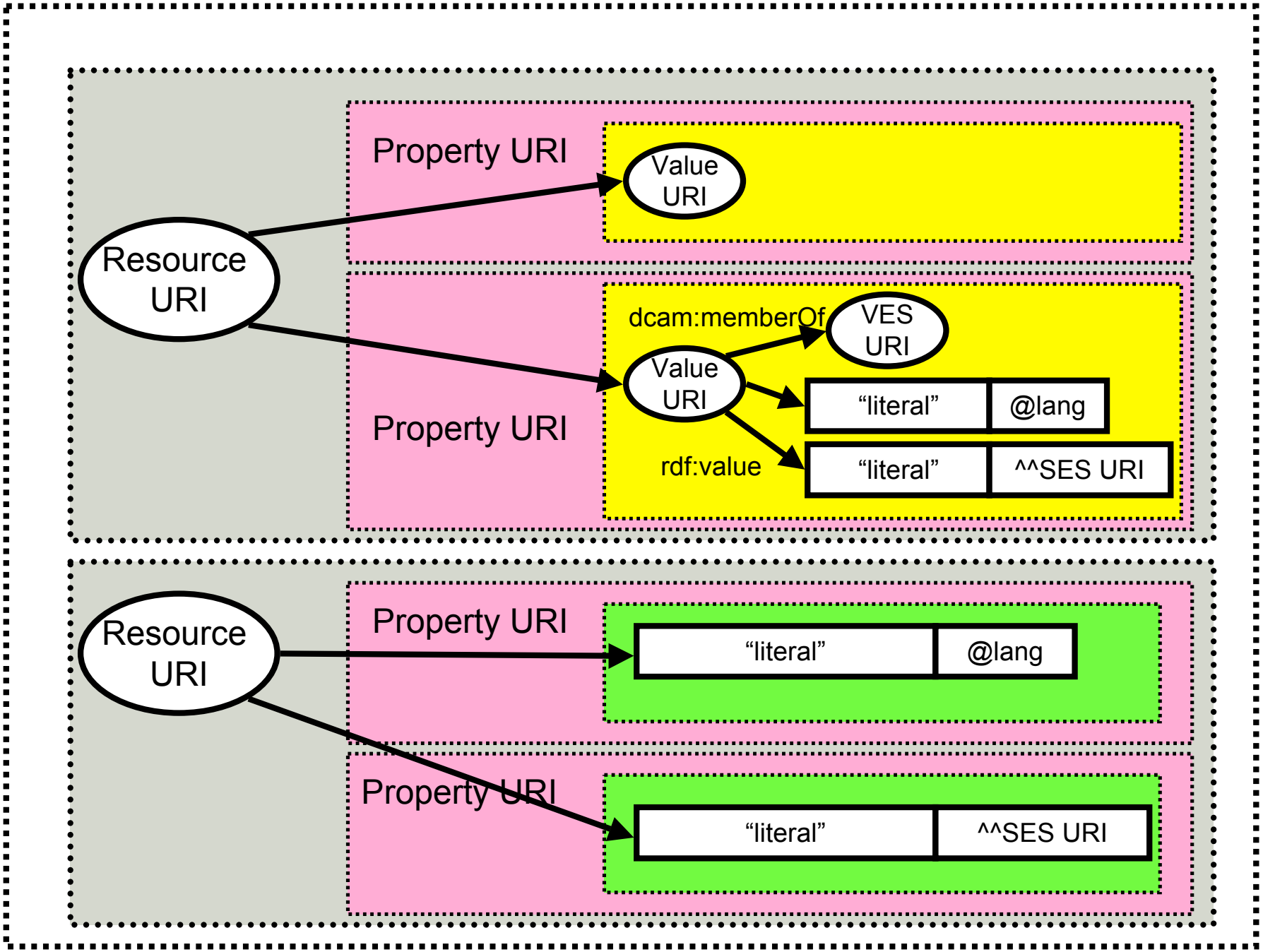
- Work by DCMI Architecture WG from mid-2003, initiated by Andy Powell (UKOLN, Eduserv)
- Second Version, DCMI Recommendation, 2007-06-04
  - http://dublincore.org/documents/2007/06/04/abstract-model/
- Based on RDF Semantics
  - Binary relations between resources
  - Use of URIs
  - RDF Schema inferencing
- DCAM uses the terminology
  - **Described resource**
  - **Property** = type of relationship
  - **Value** = other resource
- Introduces different "description model"
  - "Description set" as a community-specific "view" of an RDF Graph

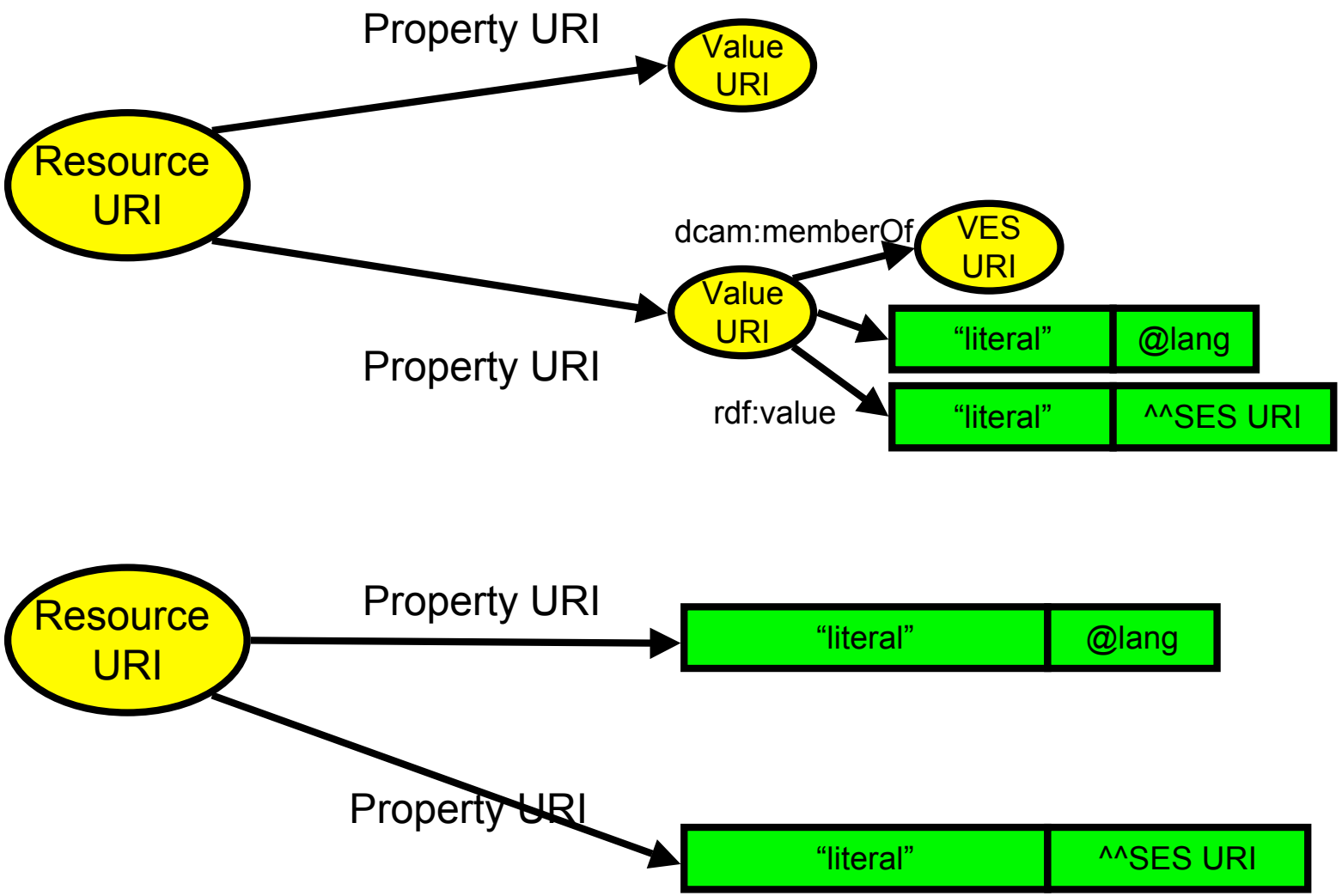

Description Set
Description
Resource URI
Statement
Non-Literal Value Surrogate
Property URI
Statement
Non-Literal Value Surrogate
Property URI
Description
Resource URI
Statement
Literal Value Surrogate
Property URI
Statement
Literal Value Surrogate
Property URI


Description Set
Description
Resource URI
Statement
Non-Literal Value Surrogate
Property URI
Value URI
Statement
Non-Literal Value Surrogate
Property URI
Value URI
Vocab Enc Scheme URI
Value string
Language
Value string
Syntax Enc Scheme URI
Description
Resource URI
Statement
Literal Value Surrogate
Property URI
Value string
Language
Statement
Literal Value Surrogate
Property URI
Value string
Syntax Enc Scheme URI

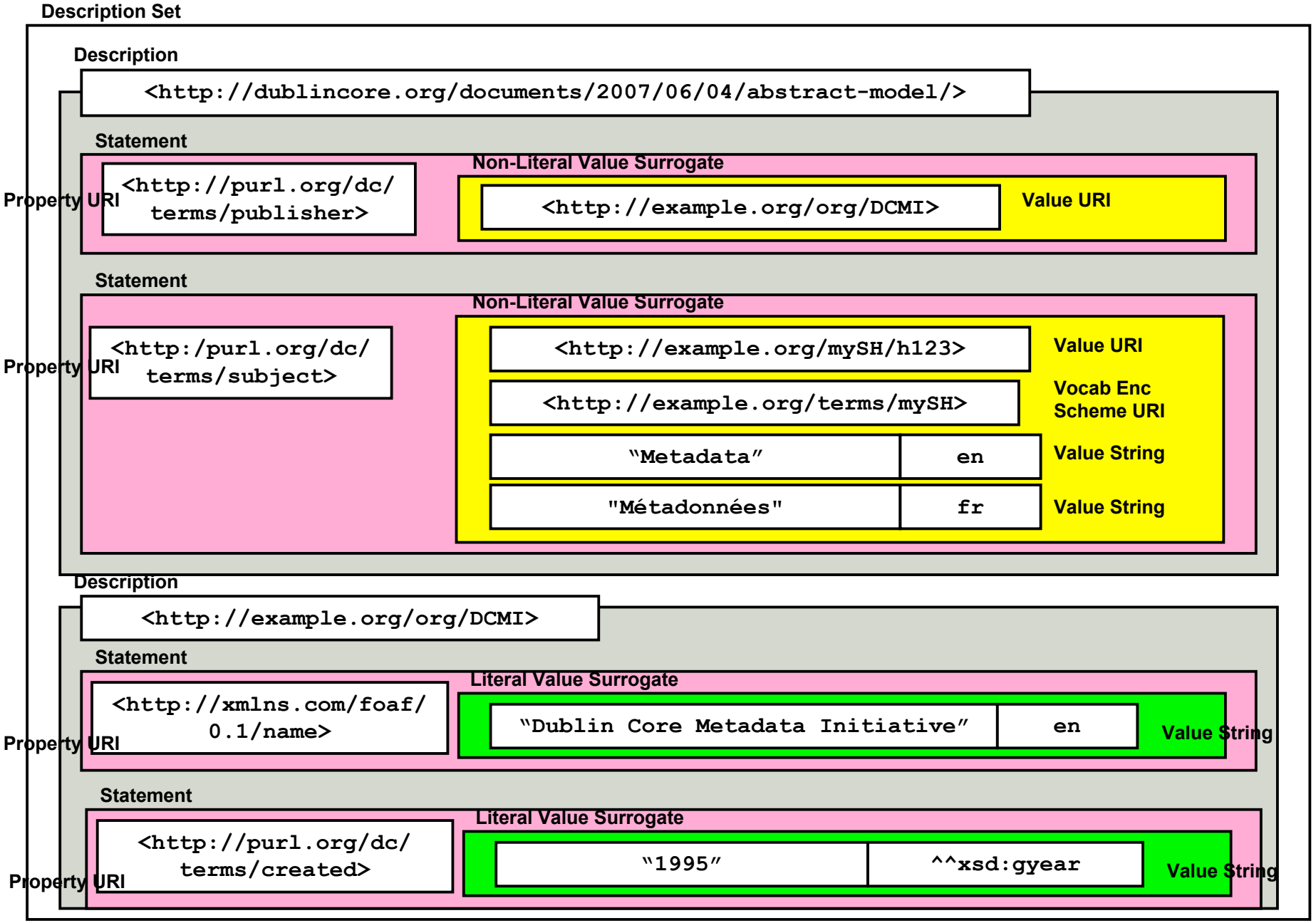

Description Set
Description
Resource URI
Statement
Property URI
Non-Literal Value Surrogate
Value URI
Statement
Property URI
Non-Literal Value Surrogate
Value URI
Vocab Enc Scheme URI
Value string
Language
Value string
Syntax Enc Scheme URI
Description
Resource URI
Statement
Property URI
Literal Value Surrogate
Value string
Language
Statement
Property URI
Literal Value Surrogate
Value string
Syntax Enc Scheme URI

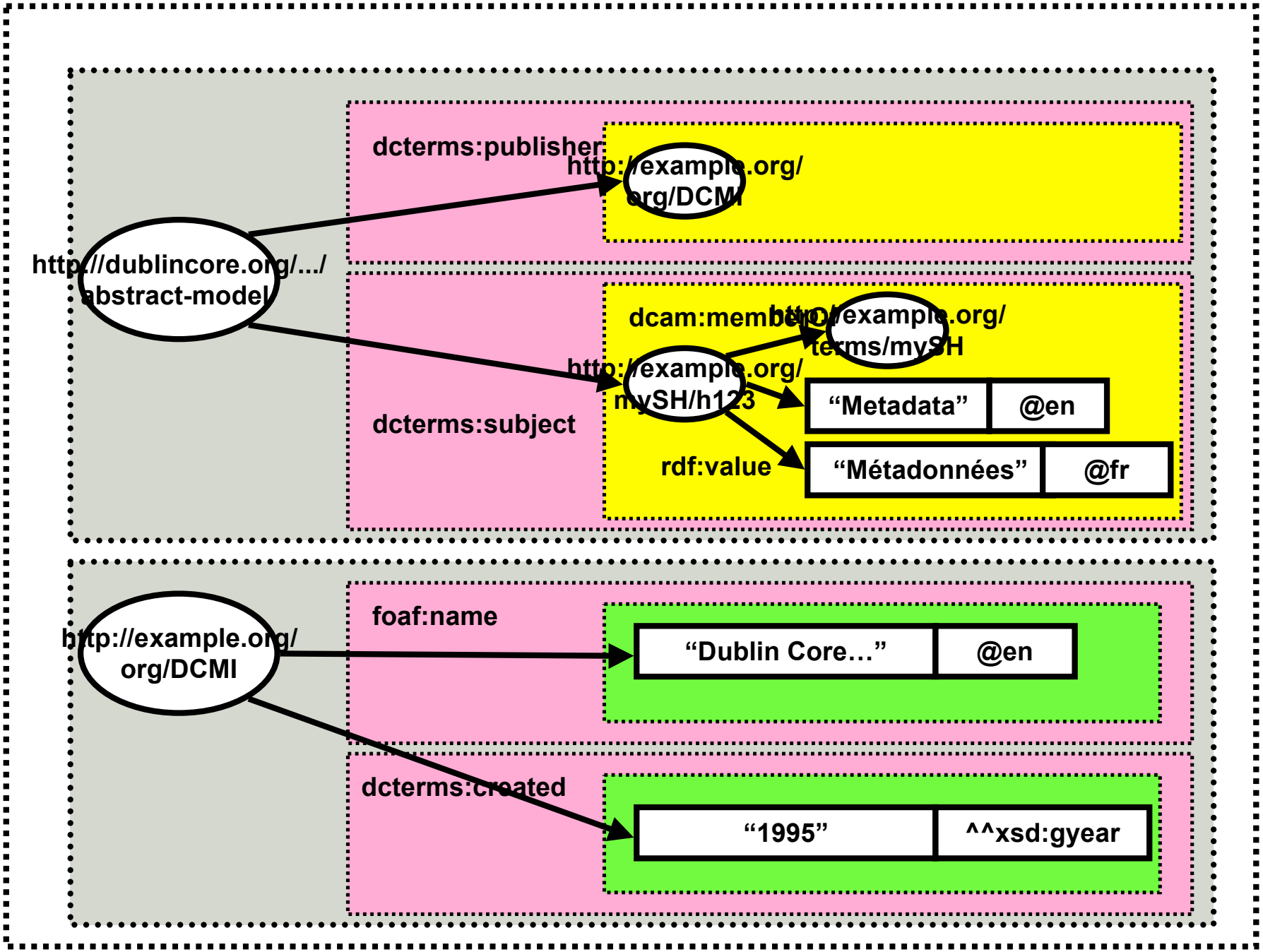

Description Set
Description
Resource URI
Statement
Non-Literal Value Surrogate
Property URI
Value URI
Statement
Non-Literal Value Surrogate
Property URI
Value URI
Vocab Enc Scheme URI
Value string
Language
Value string
Syntax Enc Scheme URI
Description
Resource URI
Statement
Literal Value Surrogate
Property URI
Value string
Language
Statement
Literal Value Surrogate
Property URI
Value string
Syntax Enc Scheme URI

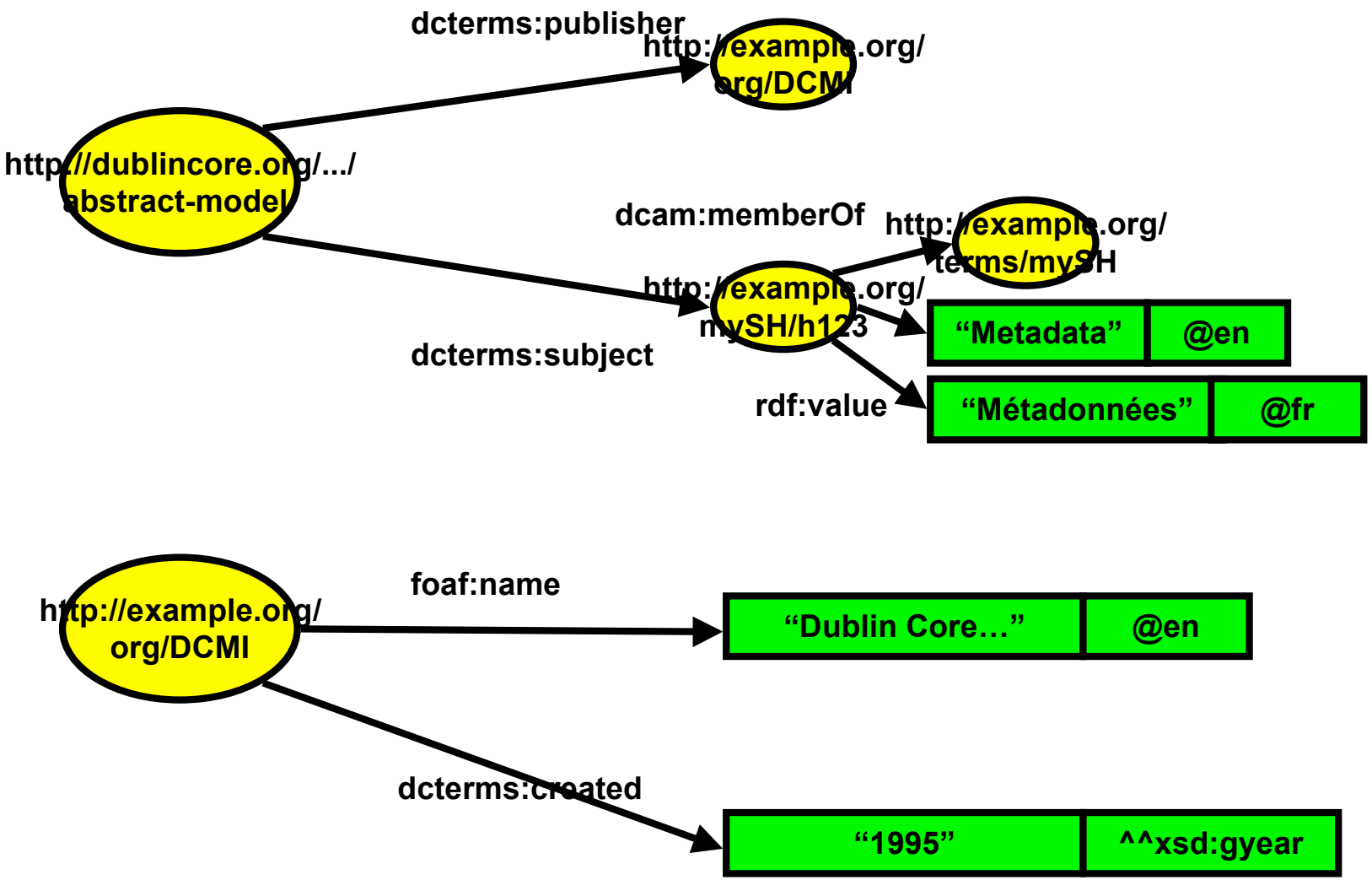

Property URI
Value URI
Resource URI
dcam:memberOf
VES URI
Value URI
"literal"
@lang
Property URI
rdf:value
"literal"
^^SES URI
Resource URI
Property URI
"literal"
@lang
Property URI
"literal"
^^SES URI


Property URI
Value URI
Resource URI
dcam:memberOf
VES URI
Value URI
"literal"
@lang
Property URI
rdf:value
"literal"
^^SES URI
Resource URI
Property URI
"literal"
@lang
Property URI
"literal"
^^SES URI

Example: Description of document, description of publisher




dcterms:publisher
http://example.org/org/DCMI
http://dublincore.org/.../abstract-model
dcam:memberOf
http://example.org/terms/mySH
http://example.org/mySH/h123
dcterms:subject
"Metadata"
@en
rdf:value
"Métadonnées"
@fr
http://example.org/org/DCMI
foaf:name
"Dublin Core…"
@en
dcterms:created
"1995"
^^xsd:gyear


dcterms:publisher
http://example.org/org/DCMI
http://dublincore.org/.../abstract-model
dcam:memberOf
http://example.org/terms/mySH
http://example.org/mySH/h123
"Metadata"
@en
dcterms:subject
rdf:value
"Métadonnées"
@fr
http://example.org/org/DCMI
foaf:name
"Dublin Core…"
@en
dcterms:created
"1995"
^^xsd:gyear

Dublin Core Application Profile

- Specification of how to construct & deploy description sets (descriptions, statements)…
  - … to serve some purpose
  - … meet some set of requirements
- At core, a profile of a "description set"
  - a set of constraints on the description set
  - based on E-R model of problem space
- Enables
  - structural validation
  - predictability for processing, querying etc -> tools

DCMI Description Set Profile (DSP)

- A way of describing **structural constraints** on a description set
  - the resources that may be described by descriptions in the description set
  - the properties that may be referenced in statements
  - the ways a value surrogate may be given
- Description templates, statement templates
- Model & XML Syntax for DSP
  - Working draft by Mikael Nilsson (Royal Institute of Technology, Sweden)
  - http://dublincore.org/documents/2008/03/31/dc-dsp/

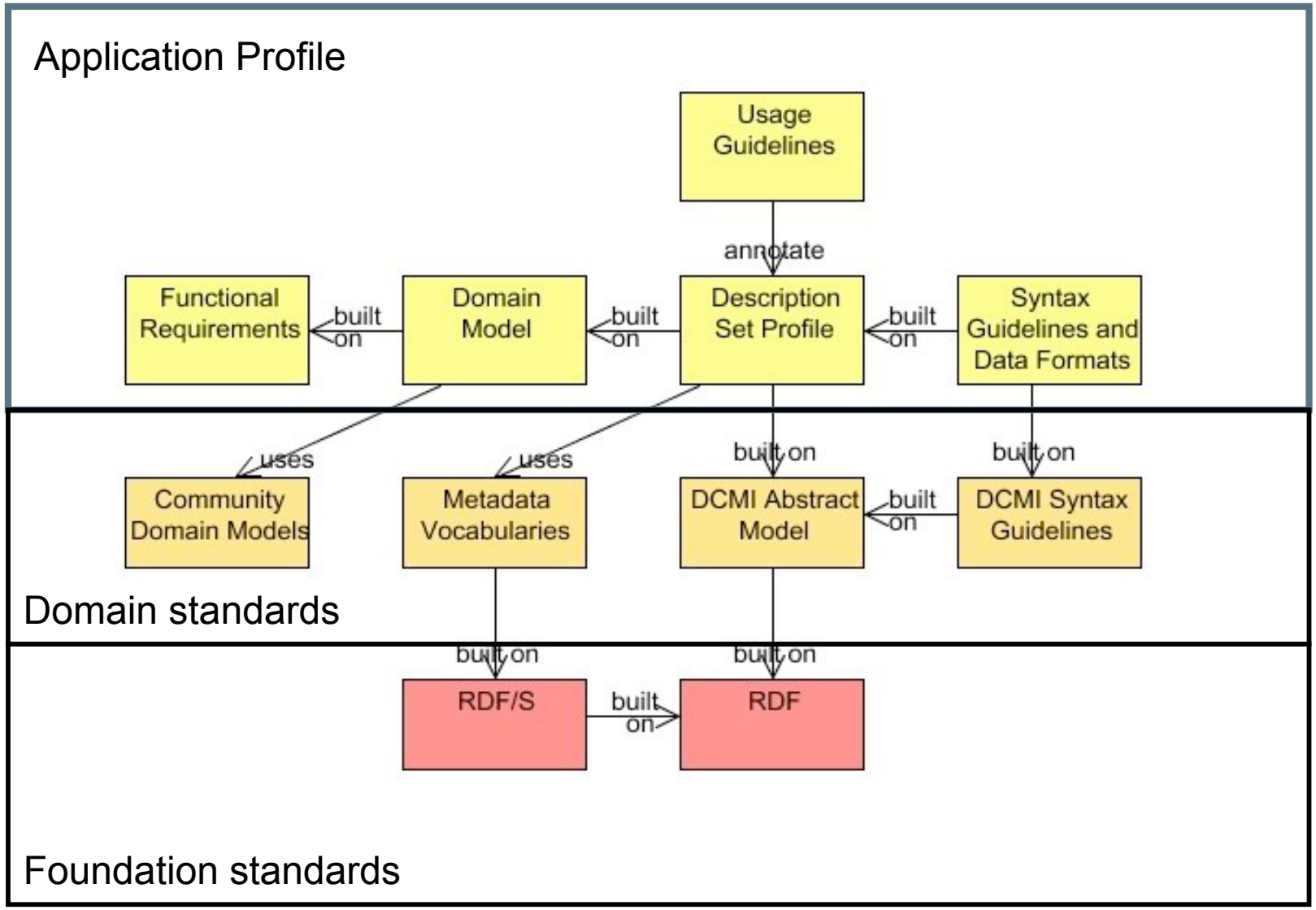
Description & Statement Templates

```
<DescriptionTemplate
   minOccurs="1"
   maxOccurs="1"
   standalone="yes">
  <StatementTemplate
    minOccurs="1"
    maxOccurs="1“
    type="literal">
   <Property>http://purl.org/dc/terms/title</Property>
  </StatementTemplate>
   …
</DescriptionTemplate>
```

Dublin Core Application Profile

- A DC Application Profile is “packet of documentation” which consists of:
  - Functional requirements (desirable)
  - Domain model (mandatory)
  - Description Set Profile (DSP) (mandatory)
  - Usage guidelines (optional)
  - Encoding syntax guidelines (optional)

The “Singapore Framework”



Scholarly Works (ePrints) DC Application Profile (SWAP)

Scholarly Works DC Application Profile

- SWAP/eprints Application Profile
  - http://www.ukoln.ac.uk/repositories/digirep/index/Eprints_Application_Profile
- Development funded by JISC, Summer 2006
- Co-ordinated by Julie Allinson (UKOLN) & Andy Powell (Eduserv Foundation)
- Specific focus on "eprint":
  - a ''scientific or scholarly research text‟ (Budapest Open Access Initiative)
  - e.g. peer-reviewed journal article, preprint, working paper, thesis, book chapter, report, etc.

The problem SWAP tries to solve

eduserv

- Emphasis on open access to research outputs
- eprints & descriptions of eprints typically made available on Web using “repository systems”
  - usually supporting OAI-PMH
- Inconsistency/ambiguity over “what is being described”
  - eprint-as-“instance” /copy
  - eprint-as-content made available in multiple “instances”
  - eprint as set of related resources
- Problems compounded by (over-)emphasis on use of “Simple DC”/oai_dc
- Difficult/impossible to
  - reliably access full-text
  - determine relationships between “instances” e.g. revisions, translations etc

The SWAP approach

- Apply the Functional Requirements for Bibliographic Records (FRBR) model to the case of eprints
- Make Works, Expressions, Manifestations & Items resources, identified by URIs
- Create DCAM-based descriptions of Works, Expressions, Manifestations & Items
  - relationships between W, E, M, I
  - attributes of W, E, M, I
  - relationships between W, E, M, I & Agents
- Mostly neutral on use of protocols
  - designed with OAI-PMH in mind
  - but does take a resource-centric approach

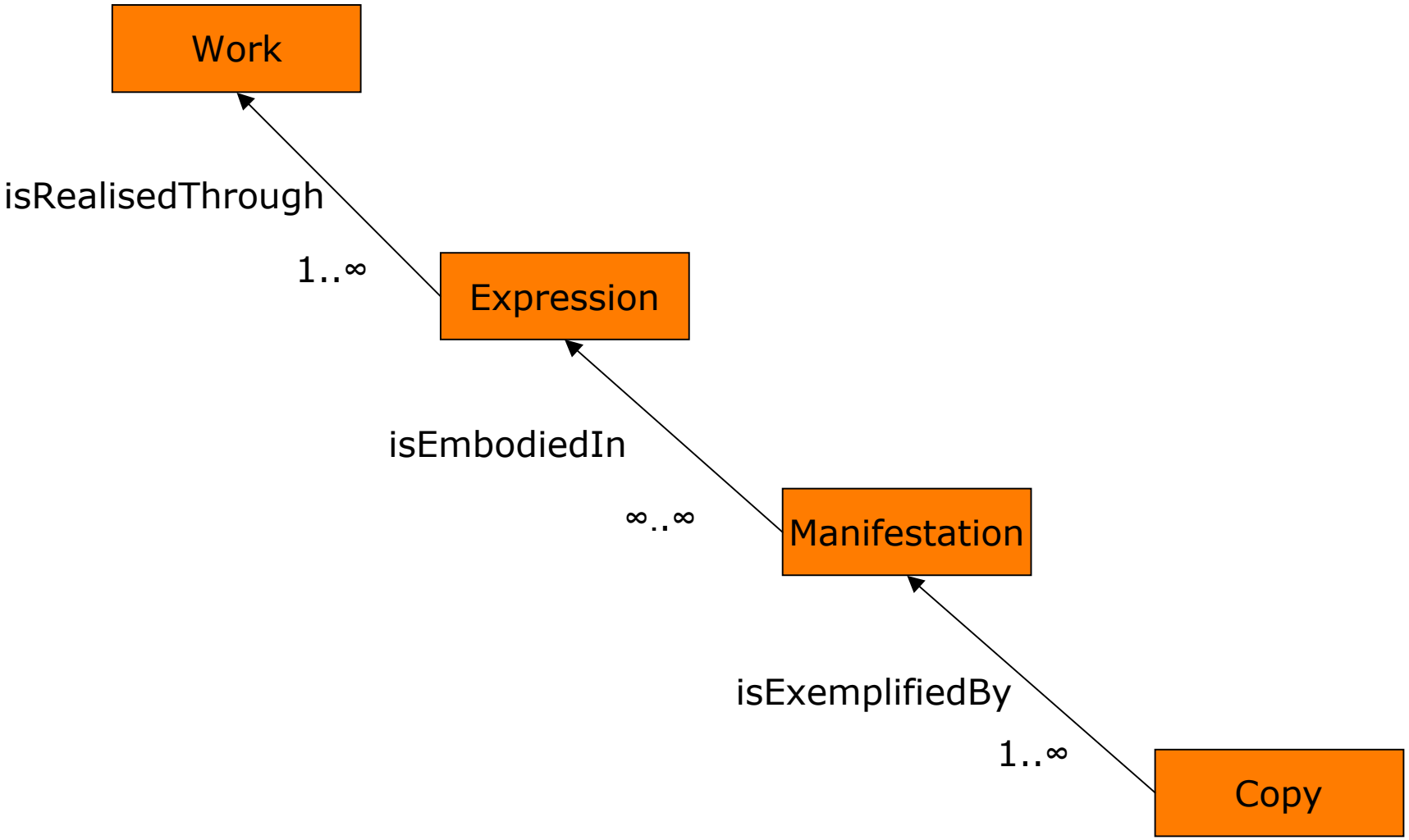
Components delivered

- Functional requirements specification
- “Domain model”
  - Based on subset of FRBR
- The “Scholarly Works Description Set Profile"
  - plus human-readable commentary, usage guidelines
- New vocabularies of metadata terms
  - With URIs like http://purl.org/eprint/terms/xyz
- Eprints DC-XML XML format
  - Based on work in progress within DCMI at time

Functional Requirements for Bibliographic Records (FRBR)

- Report of IFLA Study Group, 1998
- Entity-Relational model for the "world" that ***bibliographic records*** describe
- FRBR models the world using 4 key entities (Group 1 Entities):
  - a **work** is a distinct intellectual or artistic creation. A work is an abstract entity
  - an **expression** is the intellectual or artistic realization of a work
  - a **manifestation** is the physical embodiment of an expression of a work
  - an **item** is a single exemplar of a manifestation. The entity defined as item is a concrete entity
- Primary relationships
  - Work -- is realized through --> Expression
  - Expression -- is embodied in --> Manifestation
  - Manifestation -- is exemplified by --> Item

FRBR Group 1 Entities



Functional Requirements for Bibliographic Records (FRBR)

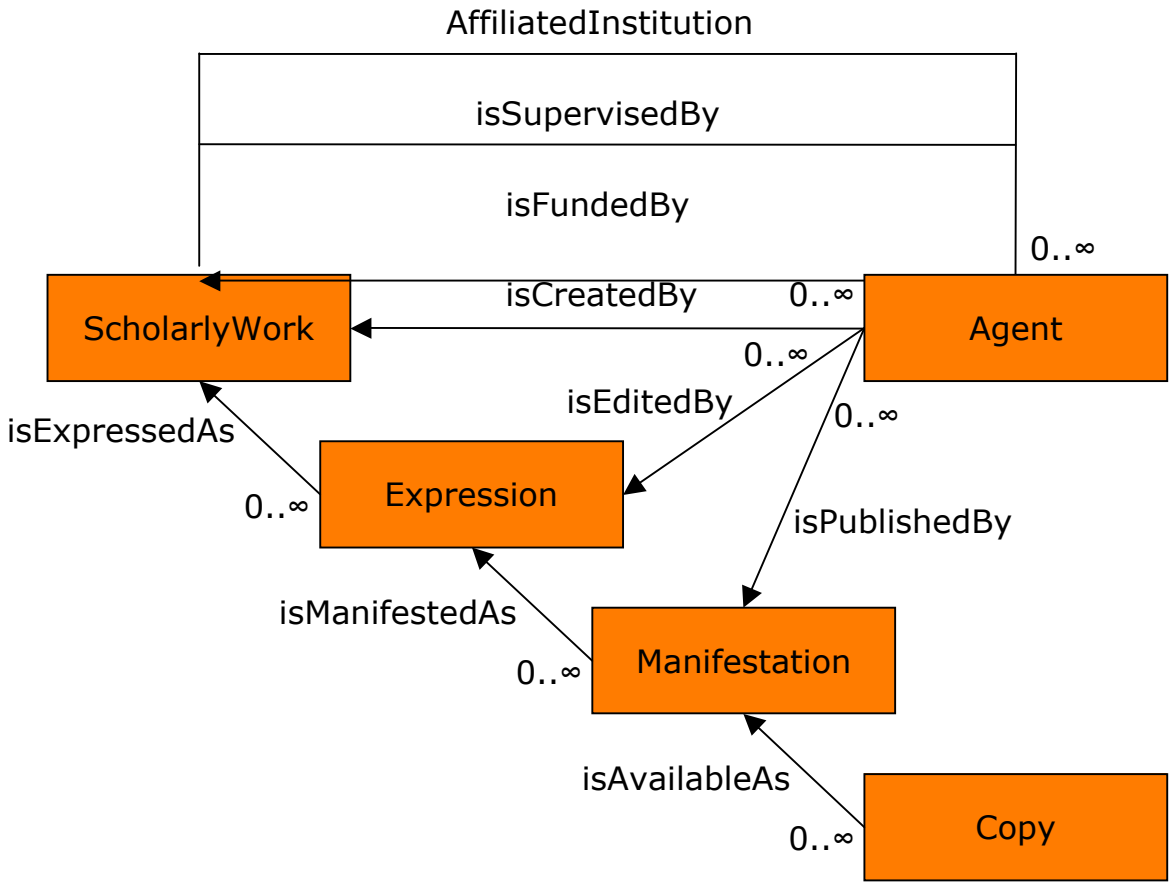
- Work-Work Relationships
  - Successor, Supplement, Adaptation etc
  - Whole-Part
- Expression-Expression Relationships
  - Abridgement, Revision, Translation etc
  - Whole-Part
- Manifestation-Manifestation Relationships
  - Reproduction, Alternate
  - Whole-Part
- Item-Item Relationships
  - Reconfiguration, Reproduction
  - Whole-Part

Functional Requirements for Bibliographic Records (FRBR)

- Group 2 Entities: Person, Corporate body
  - Responsibility relationships
    - Work is-Created-By Person/CB
    - Expression is-Realised-By Person/CB
    - Manifestation is-Produced-By Person/CB
    - Item is-Owned-By Person/CB
- Group 3 Entities: Concept, Object, Event and Place
  - Subject relationships
    - Work has-as-Subject Work/Expression/Manifestation/Item
    - Work has-as-Subject Person/CB
    - Work has-as-Subject Concept/Object/Event/Place

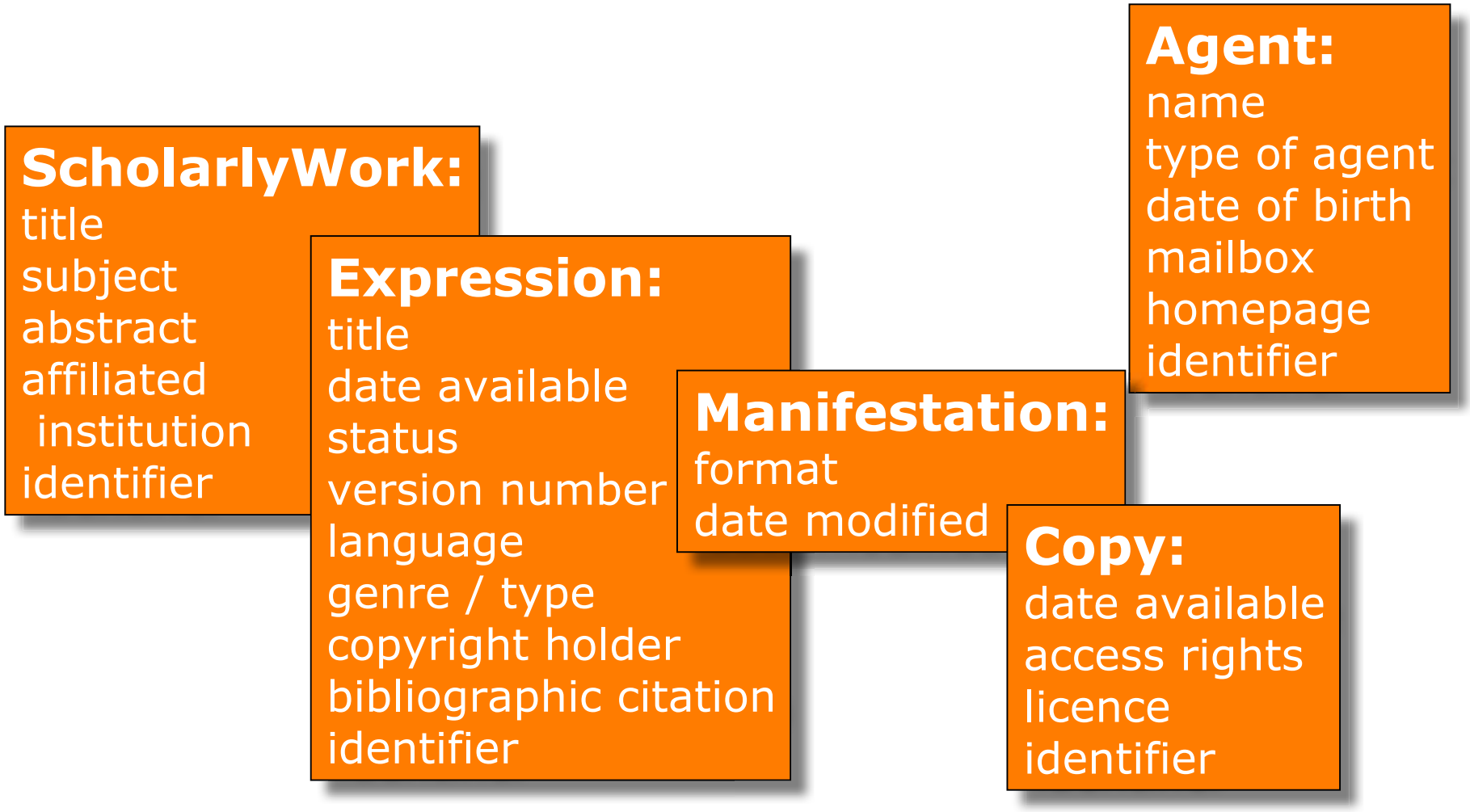
The SWAP Domain Model: Entities & Relationships

- SWAP adopts/extends a subset of the FRBR model

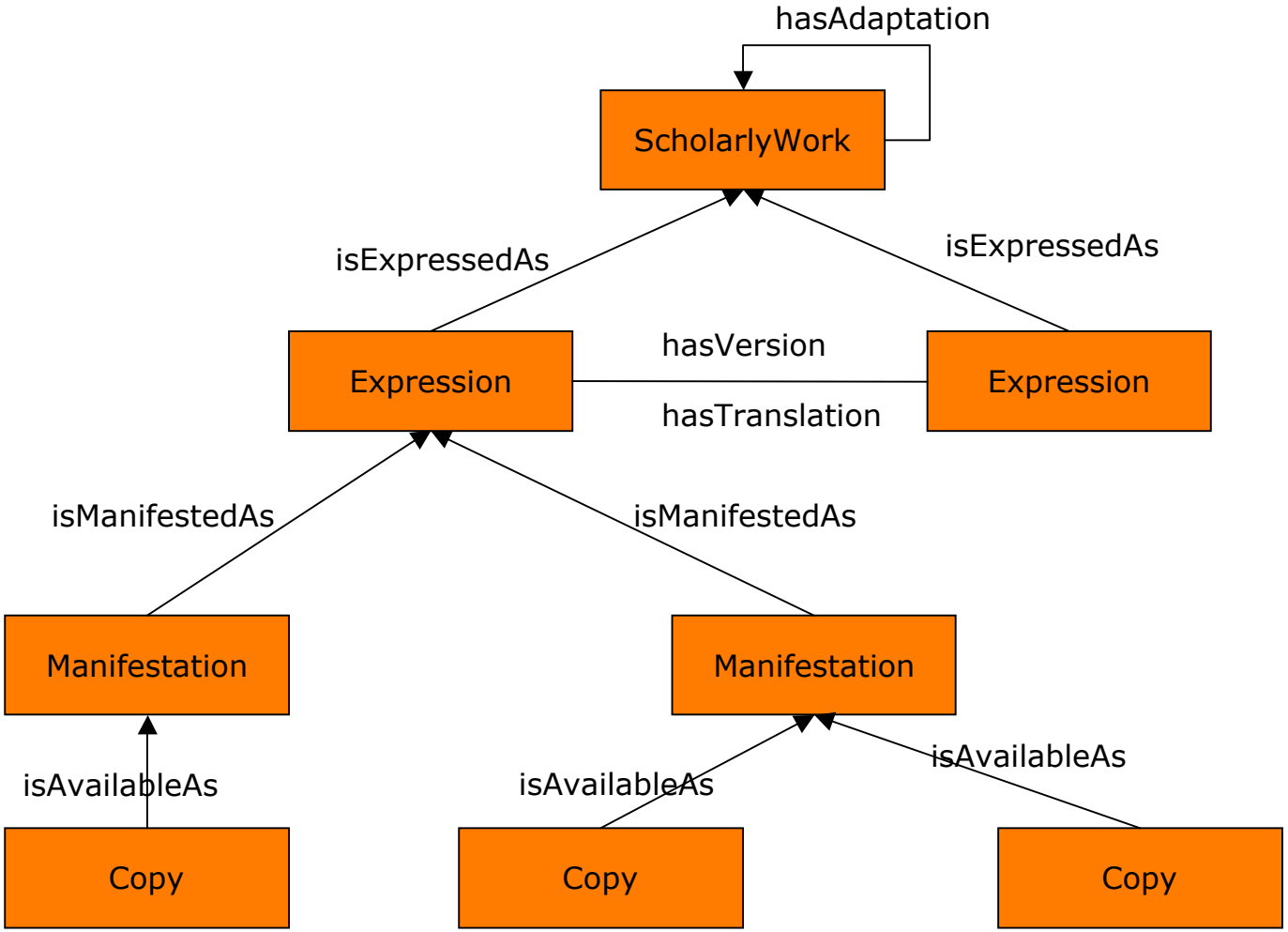


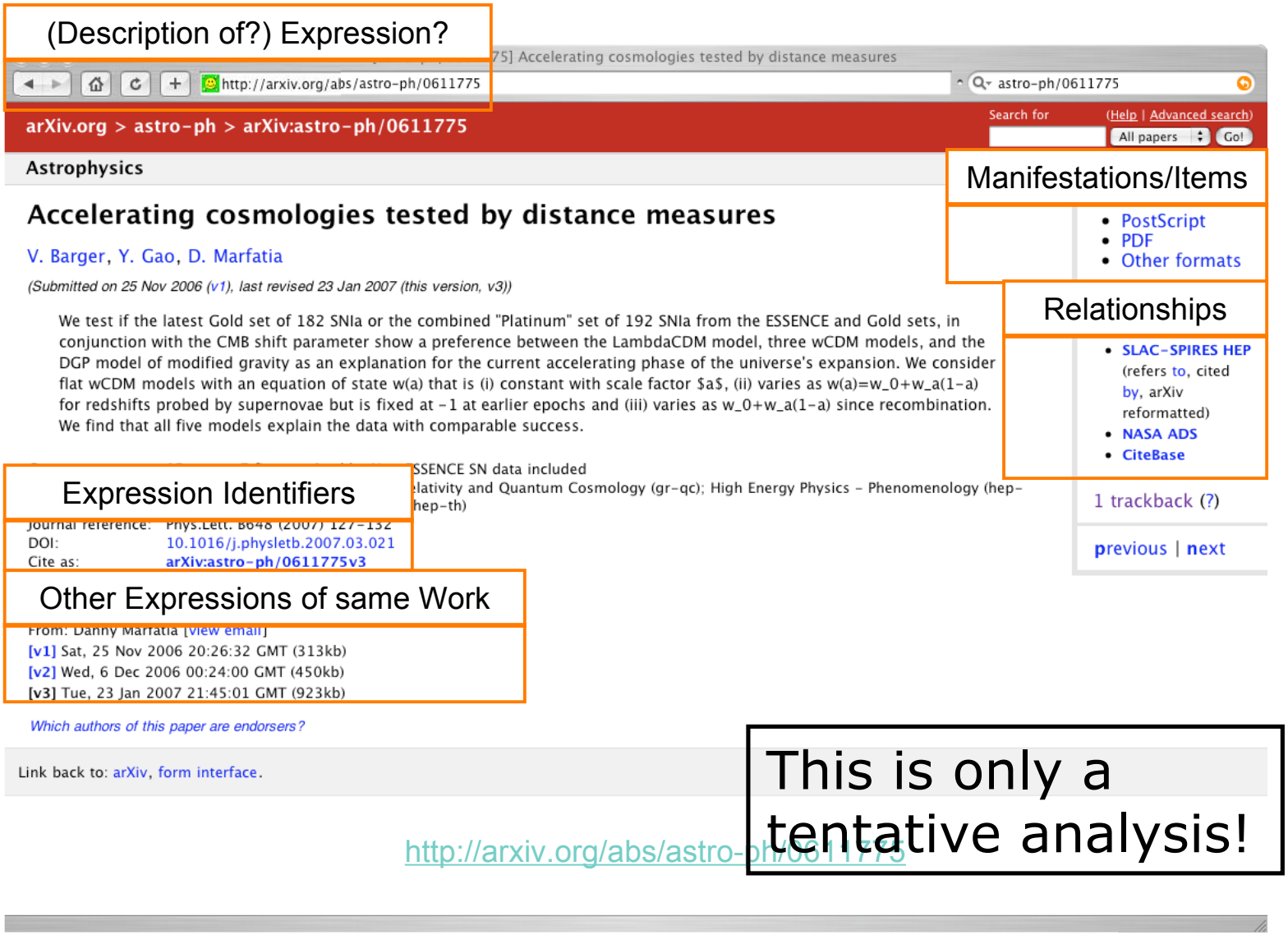
The SWAP Domain Model: Attributes

eduserv



The SWAP Domain Model




(Description of?) Expression?
http://arxiv.org/abs/astro-ph/0611775
astro-ph/0611775
arXiv.org > astro-ph > arXiv:astro-ph/0611775
Search for
(Help | Advanced search)
All papers
Go!
Astrophysics
Accelerating cosmologies tested by distance measures
V. Barger, Y. Gao, D. Marfatia
(Submitted on 25 Nov 2006 (v1), last revised 23 Jan 2007 (this version, v3))
We test if the latest Gold set of 182 SNIa or the combined "Platinum" set of 192 SNIa from the ESSENCE and Gold sets, in conjunction with the CMB shift parameter show a preference between the LambdaCDM model, three wCDM models, and the DGP model of modified gravity as an explanation for the current accelerating phase of the universe's expansion. We consider flat wCDM models with an equation of state w(a) that is (i) constant with scale factor a, (ii) varies as w(a)=w_0+w_a(1-a) for redshifts probed by supernovae but is fixed at -1 at earlier epochs and (iii) varies as w_0+w_a(1-a) since recombination. We find that all five models explain the data with comparable success.
SSENCE SN data included
elativity and Quantum Cosmology (gr-qc); High Energy Physics - Phenomenology (hep-
hep-th)
Manifestations/Items
PostScript
PDF
Other formats
Relationships
SLAC-SPIRES HEP (refers to, cited by, arXiv reformatted)
NASA ADS
CiteBase
1 trackback (?)
previous | next
Expression Identifiers
Journal reference: Phys.Lett. B648 (2007) 127-132
DOI: 10.1016/j.physletb.2007.03.021
Cite as: arXiv:astro-ph/0611775v3
Other Expressions of same Work
From: Danny Marfatia [view email]
[v1] Sat, 25 Nov 2006 20:26:32 GMT (313kb)
[v2] Wed, 6 Dec 2006 00:24:00 GMT (450kb)
[v3] Tue, 23 Jan 2007 21:45:01 GMT (923kb)
Which authors of this paper are endorsers?
Link back to: arXiv, form interface.
This is only a tentative analysis!
http://arxiv.org/abs/astro-ph/0611775

The SWAP Description Set Profile

- Specifies “description templates” for descriptions of instances of the five entity types
  - Scholarly Work, Expression, Manifestation, Copy (Item), Agent
- Specifies “statement templates” for statements within those descriptions
  - e.g. a description of a Scholarly Work must contain a statement using the dc:title property with a literal value
  - e.g. when a description of a Scholarly Work contains a statement using the eprint:isExpressedAs property, a description of the value must follow the constraints of the “Expression” description template
- Small set of required data, much is optional

Thoughts on the Approach

- Clarity on "what is being described"
- Easier to rationalise 'traditional' and 'modern' citations
  - traditional citations tend to refer to 'expressions'
  - hypertext links tend to refer to 'copies'/'items'
- Clarity on relationships between resources
  - facilitates managing "versions"
- Relatively complex underlying model may be manifest in relatively simple cataloguer and/or end-user interfaces
- FRBR applicable to other "intellectual creations"
  - Potential to merge with other data
  - FRBR as key part of RDA initiative to update library cataloguing standards

Thoughts on the Approach

eduserv

- **However**, limited implementation of SWAP to date (in JISC context)
  - Inherent complexity?
  - Fitting in to workflow?
  - Compatibility with existing internal repository software models?
- See e.g.
  - Jenny Delasalle (Warwick University), "SWAP and e-prints structures don't match" http://blogs.warwick.ac.uk/wrap/entry/swap_and_e-prints/
  - eprints.org model of "eprint" + "document" v SWAP Work, Expression, Manifestation & Item

Applications in e-Science domain

- Applicable to scientific eprints!
- SWAP was scoped to focus on eprints
- FRBR model more broadly applicable
- Current JISC-funded work on DCAPs for
  - Geo-Spatial resources
  - Still Images
  - Time-Based Media
  - Learning Materials (scoping study)
- Preliminary investigations on datasets
  - Relationship of FRBR & aggregation models (ORE?) to description of datasets

Concluding thoughts

Some concluding thoughts

	ORE	SWAP
Nature	Both seek to address problems of ambiguity of identification, explicitness of relationship types in (Semantic) Web-friendly ways	
Scope	General in scope (though roots in scholarly communication domain)	Scoped to specific resource type (scholarly works) (though FRBR broader)
Domain model	Aggregations (ore:aggregates)	Subset of FRBR model (several different relationship types); SWAP doesn't address aggregation
Description model	RDF (with constraints)	DCAM, layered on RDF
Vocabularies	Requires core set of properties, allows for others	Specifies set of properties to be used
Deployment	Explicitly aligns with Web Arch/httpRange-14 etc	Less explicit, doesn't include the "describes" relationship, but resource-centric

Acknowledgements

The sections on Web Architecture and OAI ORE make use of material from a presentation by Herbert Van de Sompel, "An Introduction to the ORE interoperability framework", presented at the 4th Search/Find Workshop, Ghent, Belgium, 22 August 2008

The section on SWAP adapts material from a presentation by Julie Allinson, "A Dublin Core Application Profile for Scholarly Works", presented to the JISC CETIS Metadata & Digital Repositories SIG, Manchester, UK, 16 April 2007

Title slide photo "Iron ore flakes" by Flickr user lars hammar
See http://www.flickr.com/photos/7603557@N08/1195127627/
Made available under CC Attribution-NonCommercial-Share-Alike 2.0 license

eduserv

ORE & SWAP: Composition & Complexity

14 October 2008

Metadata Infrastructures, MPG eScience Seminar, Berlin

Pete Johnston, Eduserv Foundation
pete.johnston@eduserv.org.uk
http://www.eduserv.org.uk/foundation