Embedding libraries and cultural memory institutions in the Linked Data cloud

Thomas Baker

Dublin Core Metadata Initiative

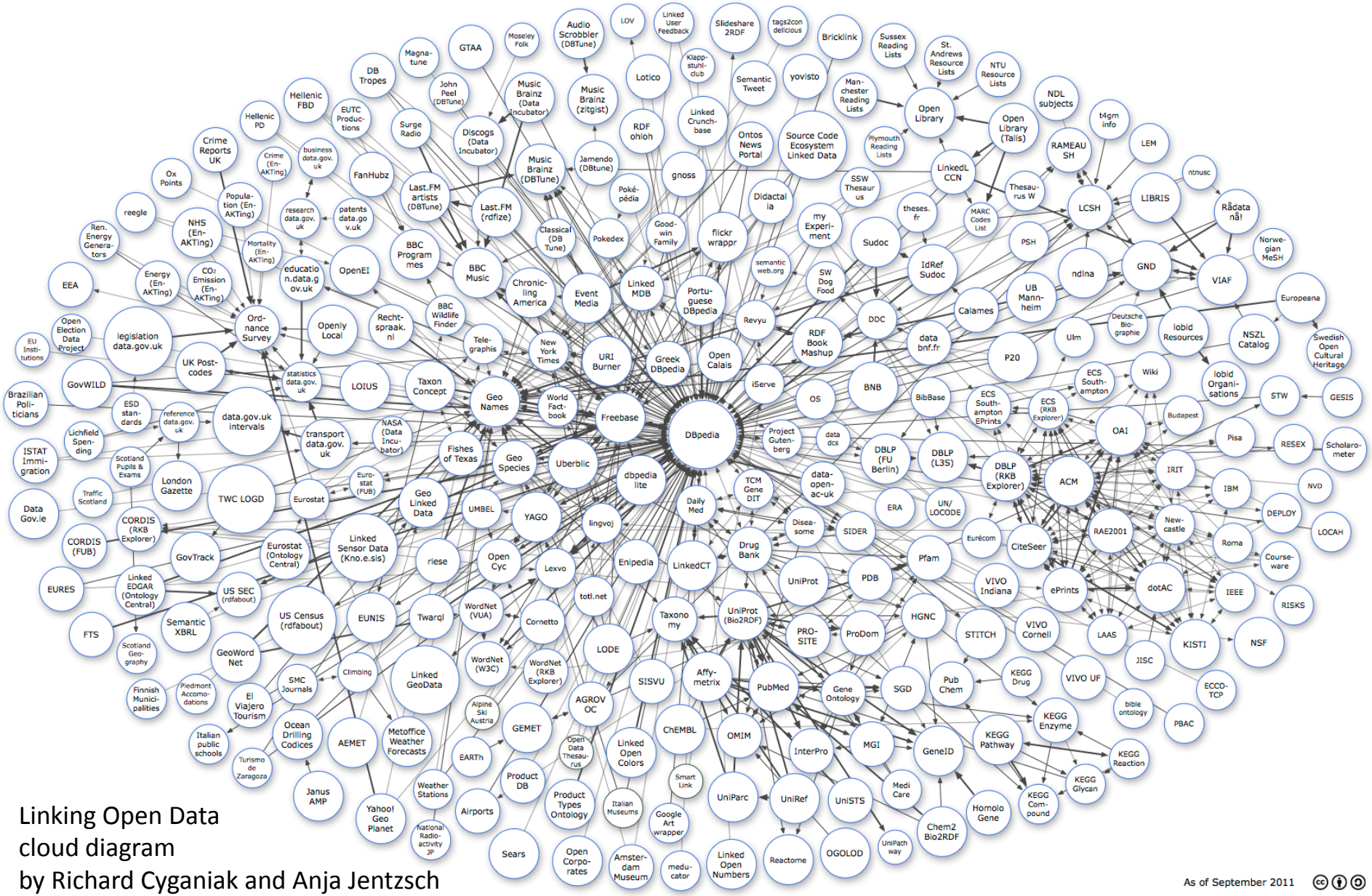
International Symposium on Knowledge Communities 2012

Research Center for Knowledge Communities

University of Tsukuba

15 December 2012

Data Web – the Cloud…



Linking Open Data cloud diagram
by Richard Cyganiak and Anja Jentzsch
http://lod-cloud.net/

As of September 2011

Description of a scientific article...

Triplifying the description…

Subject	Predicate	Object
agris:CN2009002389	has the Title	典型草原三种蝗虫种群死亡率的研究
agris:CN2009002389	has an Author	Lu Hui
agris:CN2009002389	has a Topic	Acrididae (grasshoppers)

A URI identifying the article

Resource A is identified by the URI http://agris.fao.org/resource/CN2009002389.
Let's abbreviate this as agris:CN2009002389.

Triplifying the description…

Subject	Predicate	Object
agris:CN2009002389	has the Title	典型草原三种蝗虫种群死亡率的研究
agris:CN2009002389	has an Author	agris-author:luhui
agris-author:luhui	has the Name	Lu Hui
agris:CN2009002389	has a Topic	Acrididae (grasshoppers)

A URI identifying the author

The author is identified by the URI http://agris.fao.org/author/luhui.
Let's abbreviate this as agris-author:luhui.

Triplifying the description…

Subject	Predicate	Object
agris:CN2009002389	has the Title	典型草原三种蝗虫种群死亡率的研究
agris:CN2009002389	has an Author	agris-author:luhui
agris-author:luhui	has the Name	Lu Hui
agris:CN2009002389	has a Topic	agrovoc:c_4416
agrovoc:c_4416	has preferred label	Acrididae (en)
agrovoc:c_4416	has preferred label	蝗科 (zh)

A URI identifying the subject

The description of http://aims.fao.org/aos/agrovoc/c_4416 in the AGROVOC Concept Scheme tells us how this concept is labeled in English and in Chinese.

Triplifying the description…

Subject	Predicate	Object
agris:CN2009002389	dct:title	典型草原三种蝗虫种群死亡率的研究
agris:CN2009002389	dct:creator	agris-author:luhui
agris-author:luhui	foaf:name	Lu Hui
agris:CN2009002389	dct:subject	agrovoc:c_4416
agrovoc:c_4416	has preferred label	Acrididae (en)
agrovoc:c_4416	has preferred label	蝗科 (zh)

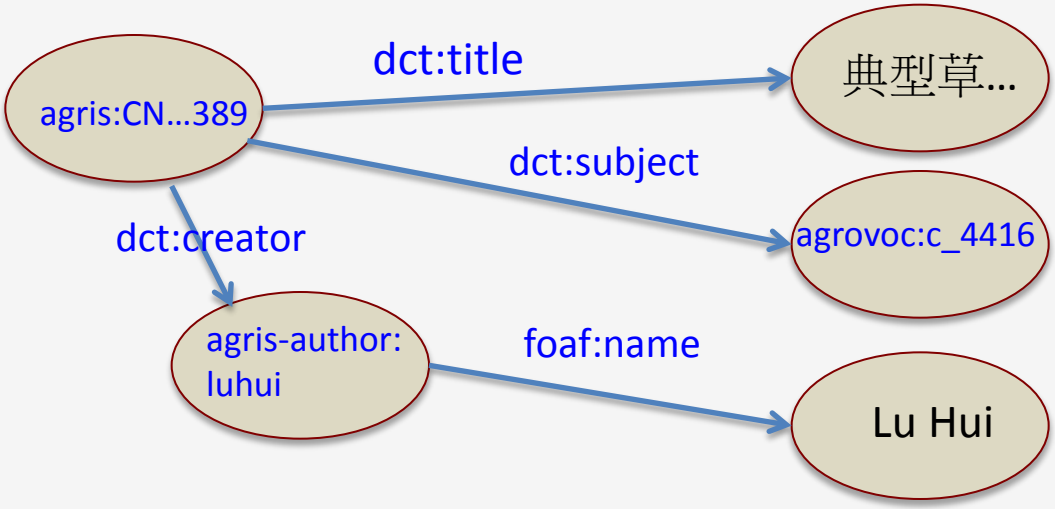
URIs identifying properties
(from Dublin Core and SKOS)

Triplifying the description…

Subject	Predicate	Object
agris:CN2009002389	dct:title	典型草原三种蝗虫种群死亡率的研究
agris:CN2009002389	dct:creator	agris-author:luhui
agris-author:luhui	foaf:name	Lu Hui
agris:CN2009002389	dct:subject	agrovoc:c_4416
agrovoc:c_4416	skos:prefLabel	Acrididae (en)
agrovoc:c_4416	skos:prefLabel	蝗科 (zh)

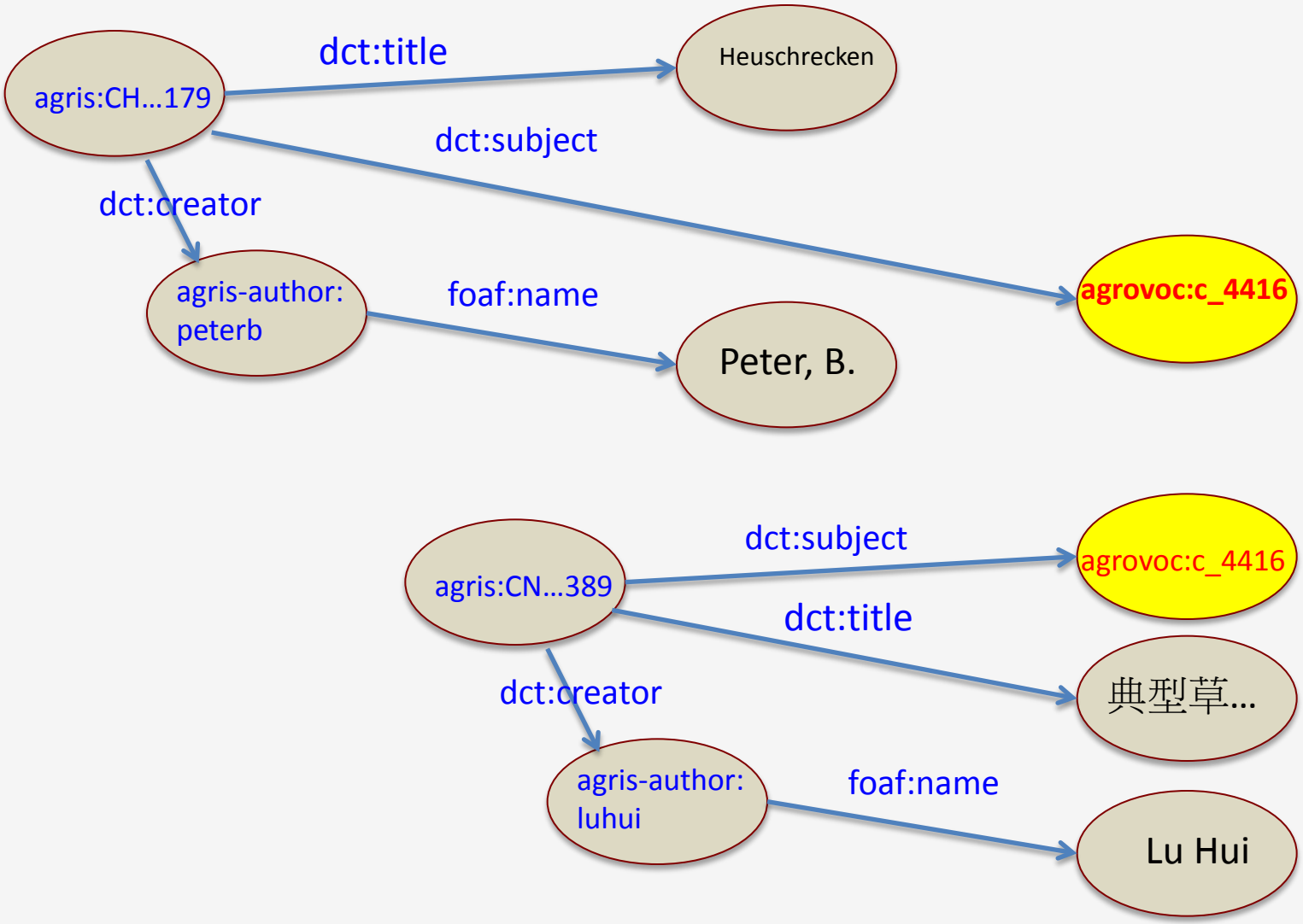
URIs identifying properties (from SKOS)

Triples seen as a graph

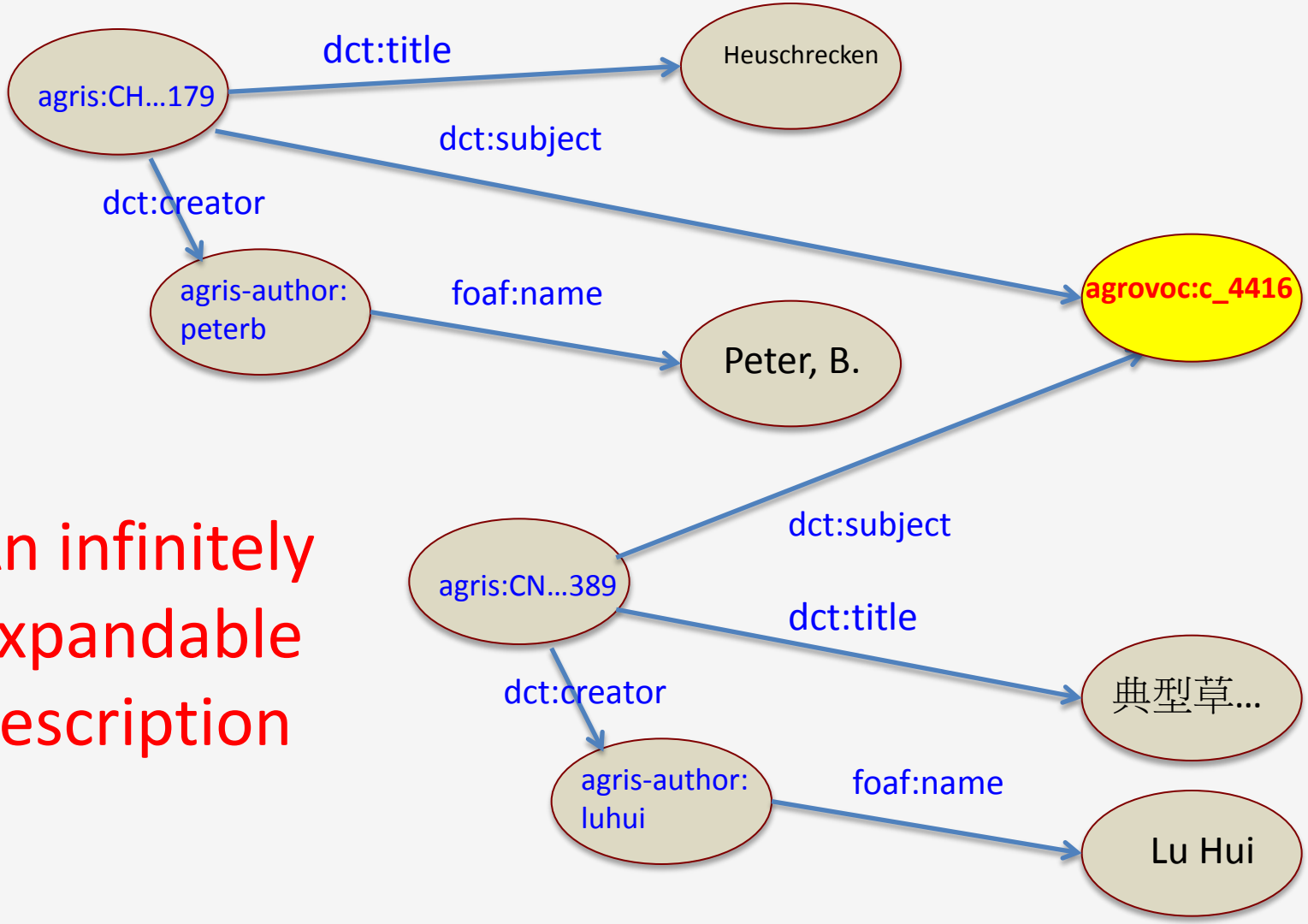


Subject	Predicate	Object
agris:CN2009002389	dct:title	典型草原三种蝗虫种群死亡率的研究
agris:CN2009002389	dct:creator	agris-author:luhui
agris-author:luhui	foaf:name	Lu Hui
agris:CN2009002389	dct:subject	agrovoc:c_4416

Computers detect URIs that match...



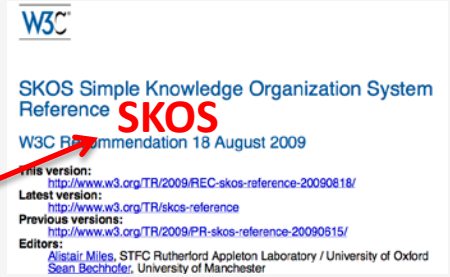
…thus “linking” the descriptions

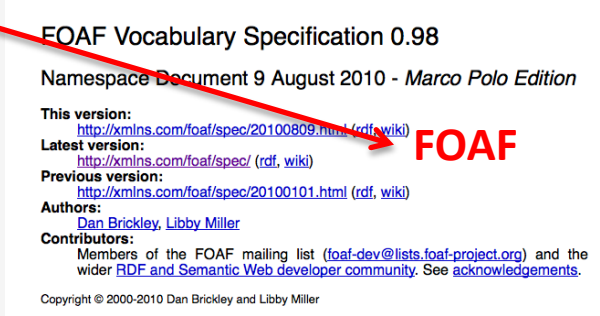


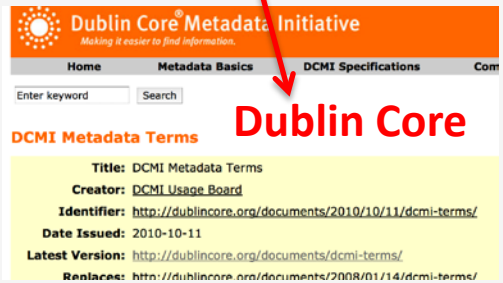
An infinitely expandable description

Expanding to other Web resources

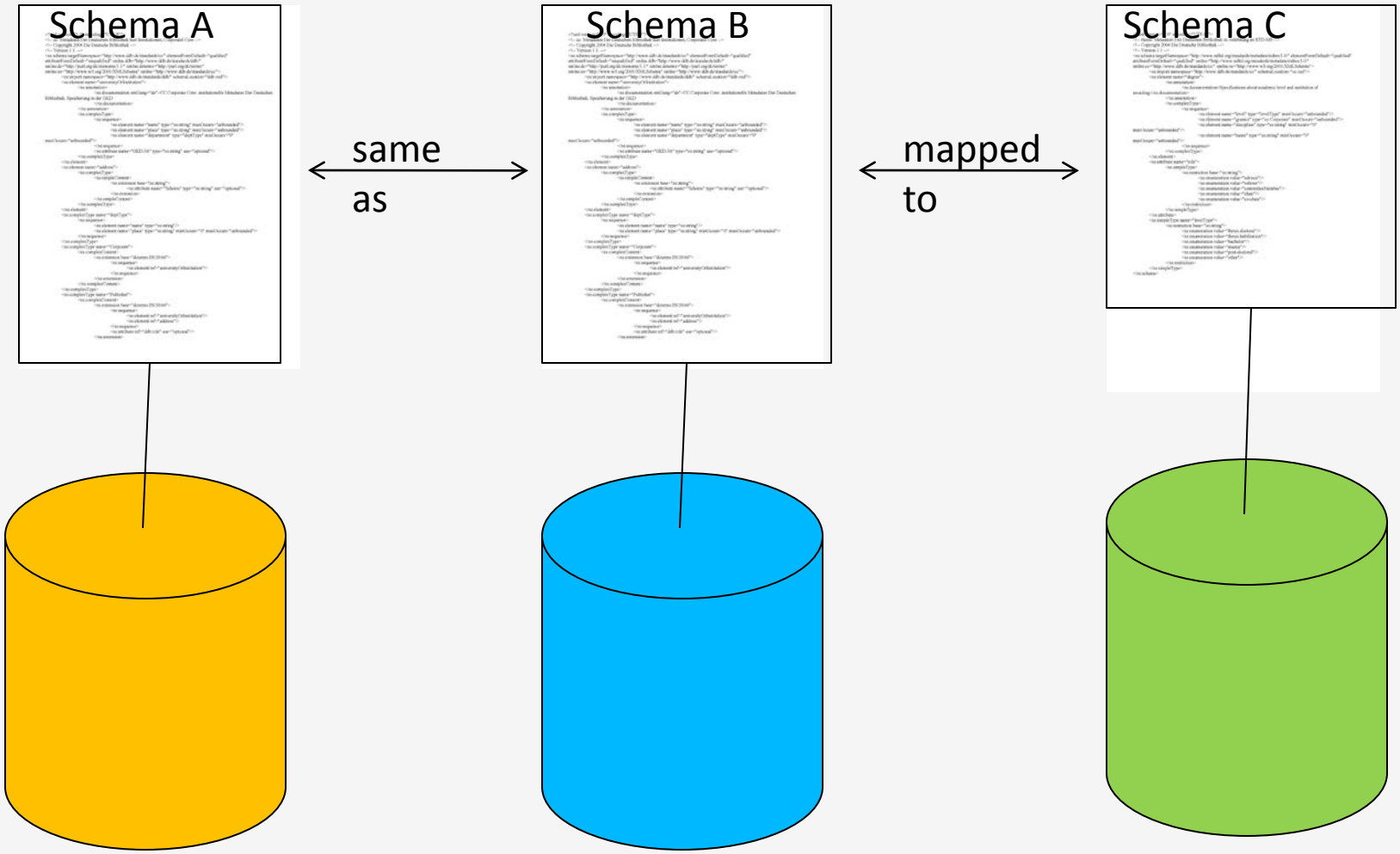
“Records” have become clusters of links



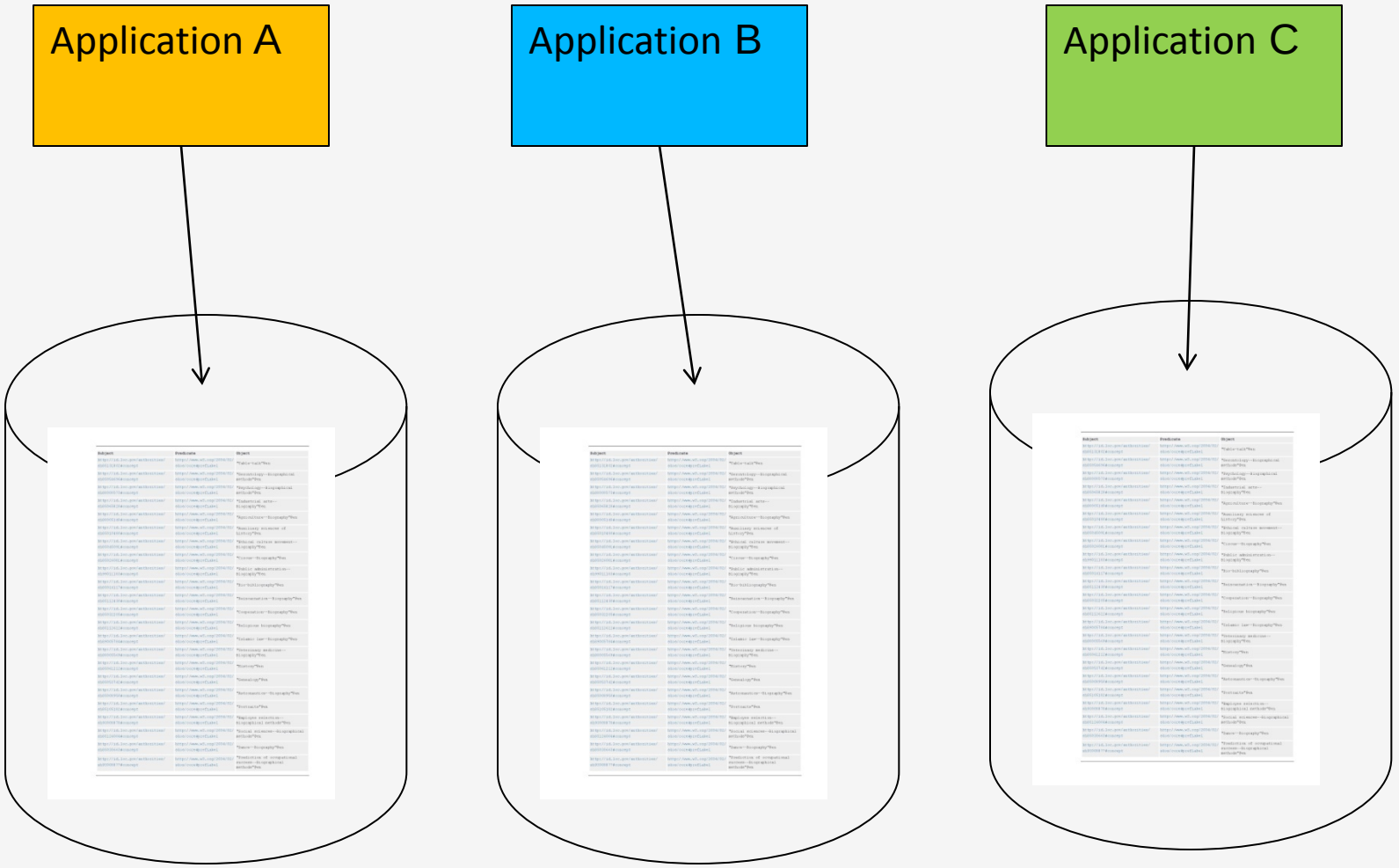




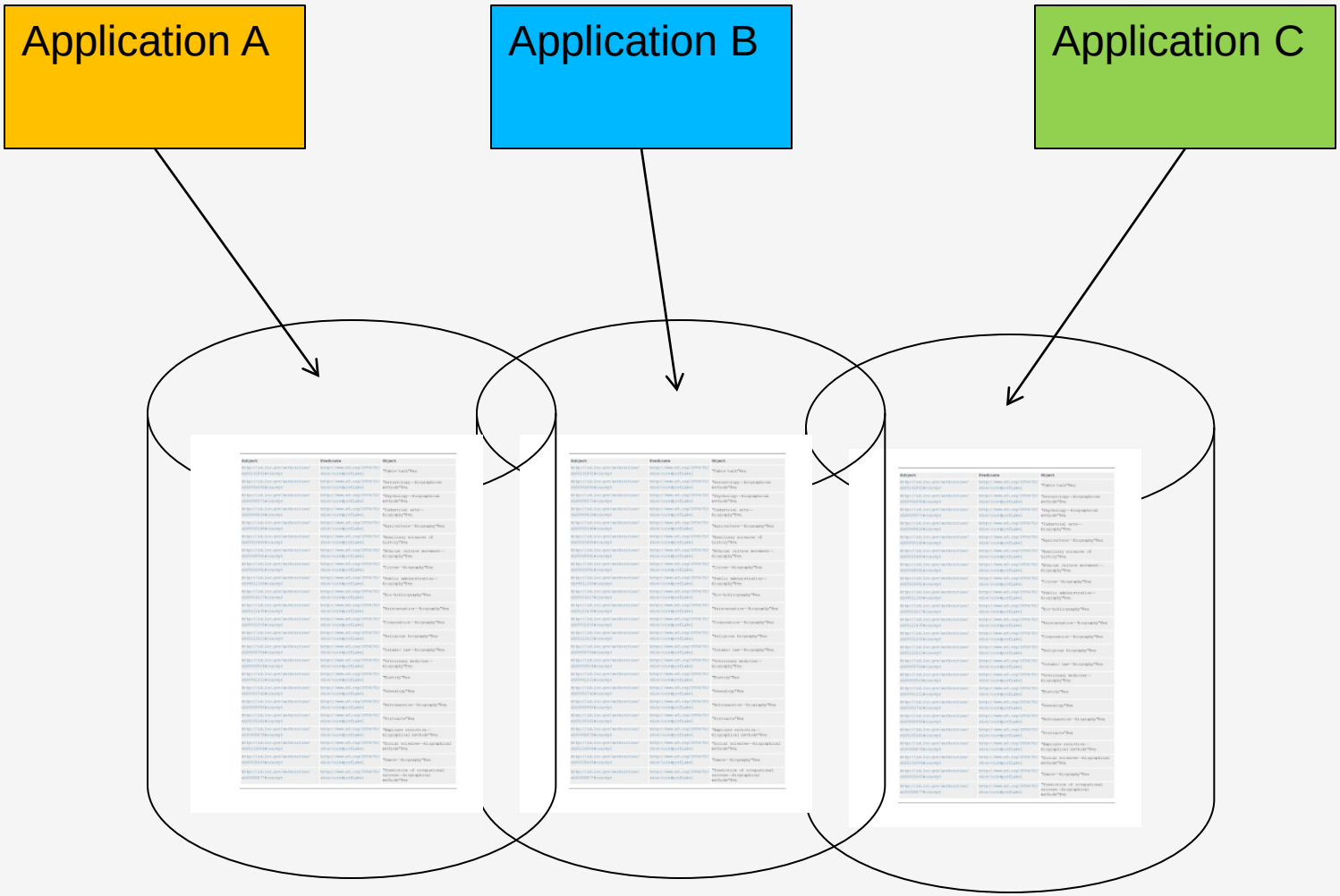
Traditional IT maps between Records



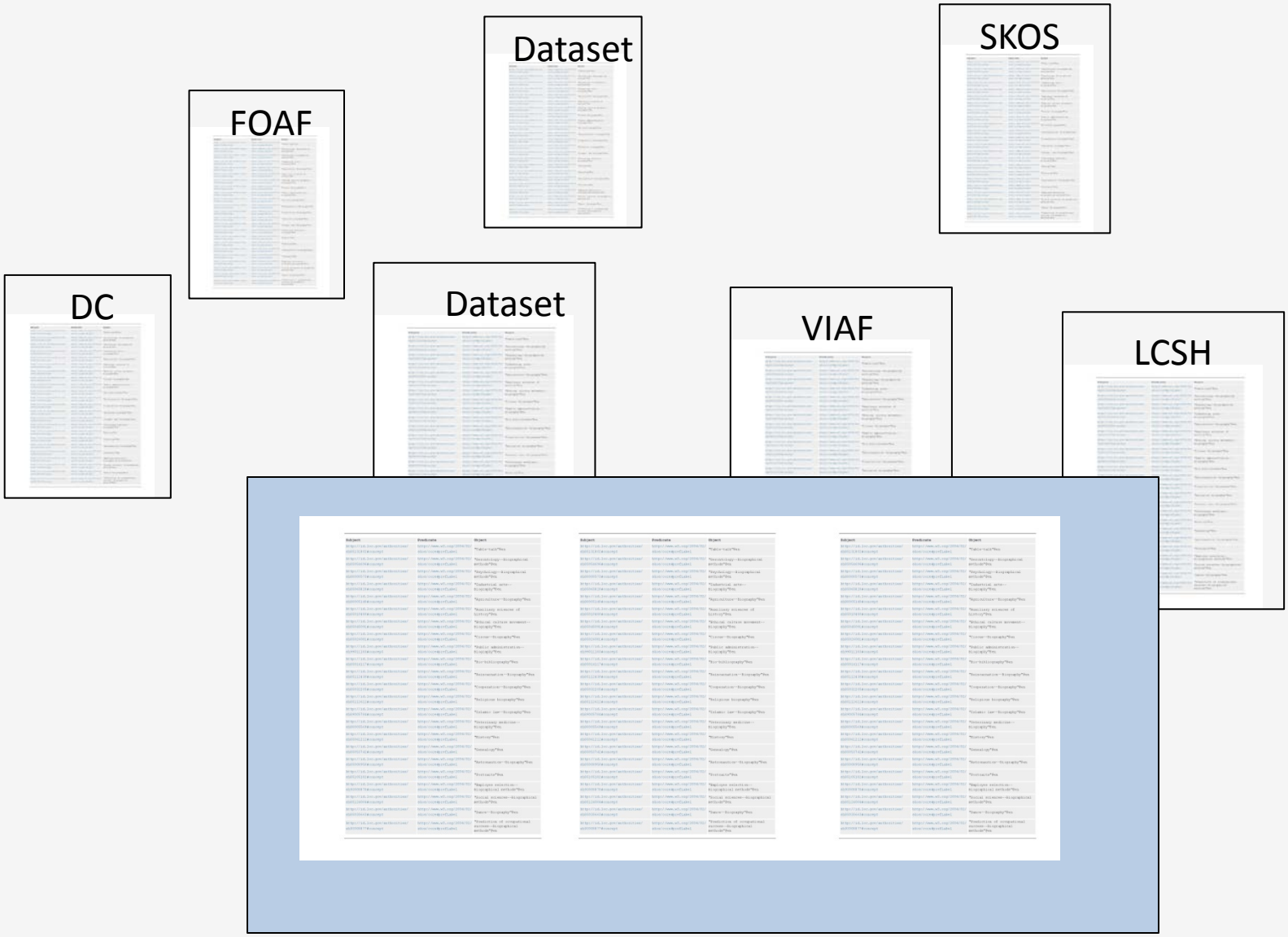
Linked Data applications create good triples



Linked Data “speaks for itself”...



…as long as source datasets and vocabularies are preserved



What Linked Data uses URIs for

Concept

Person

Object

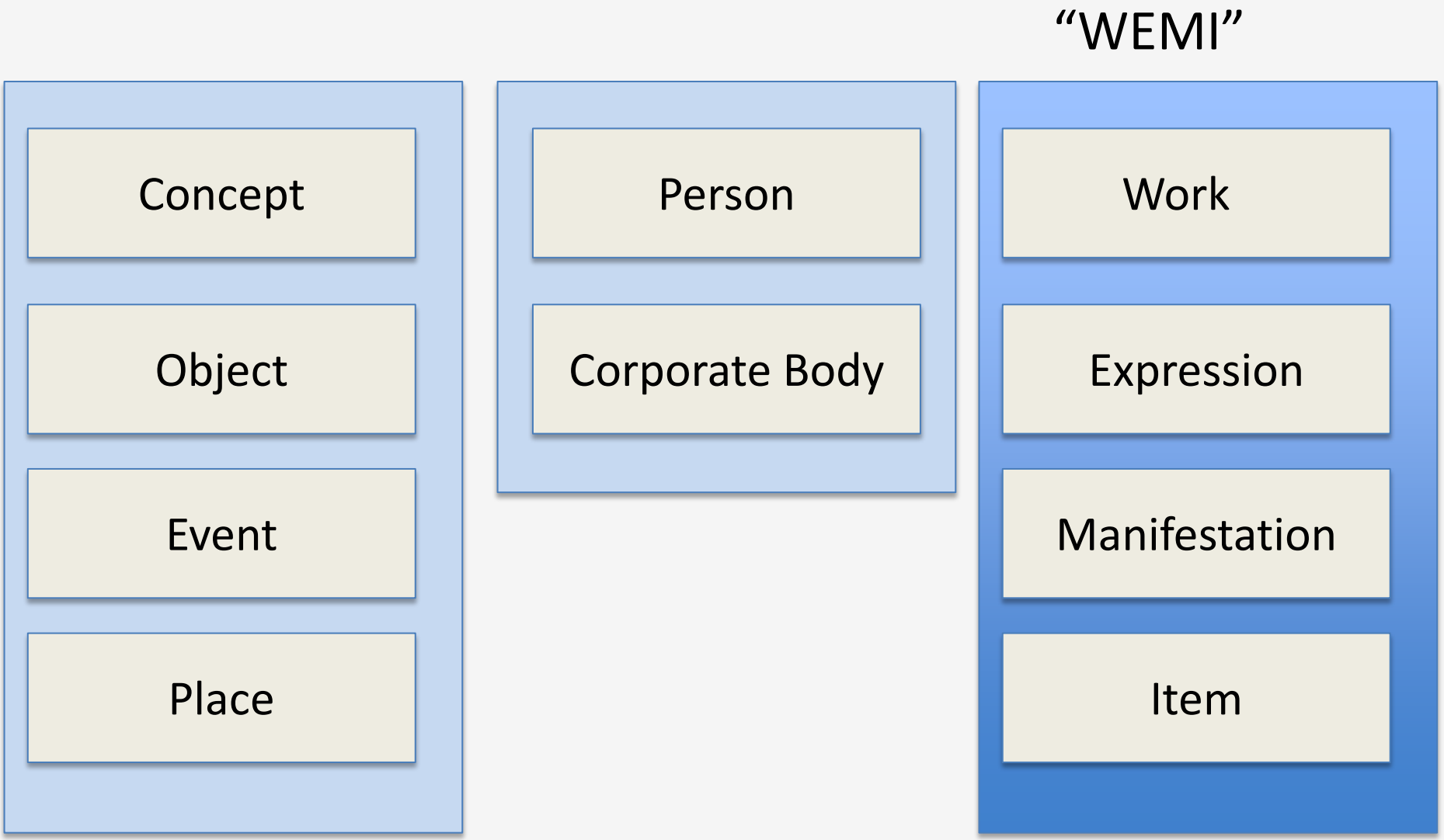
Corporate Body

Event

Place

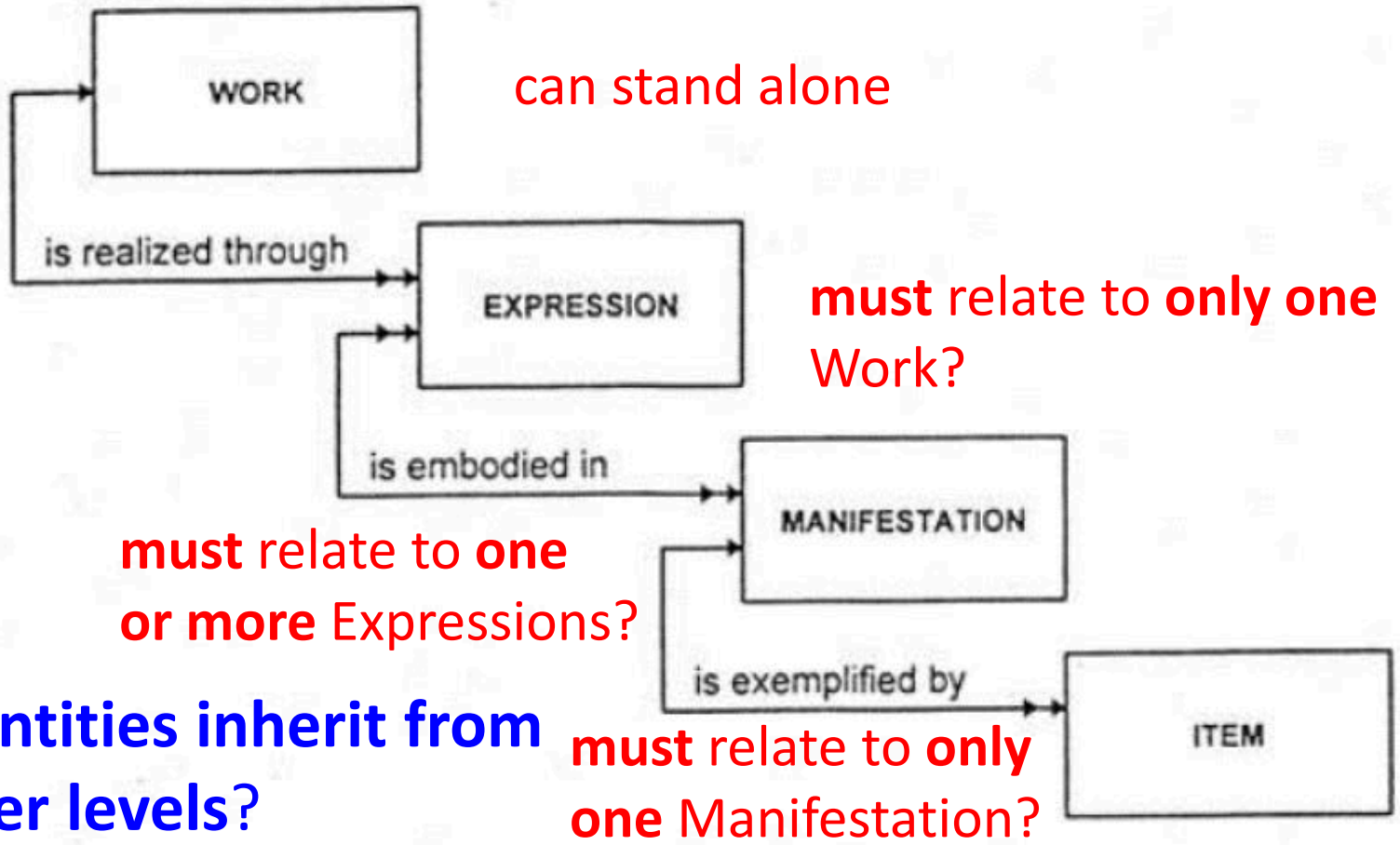
...and the relationships between them (properties)

FRBR theory


"WEMI"
Concept
Object
Event
Place
Person
Corporate Body
Work
Expression
Manifestation
Item

FRBR theory

"WEMI"



Do entities inherit from higher levels?

Replacing cataloging rules?

- Resource Description and Access (RDA)
  - Successor to AACR for library catalogs
  - Tied to FRBR ontology
  - Better support for relationships
  - Implies **Work** and **Expression** for every bibliographic item
  - Like FRBR, relatively *untested*

Replacing MARC format?

- Bibframe Initiative
  - Library of Congress replacement for MARC
  - http://www.loc.gov/marc/transition/pdf/marcld-report-11-21-2012.pdf
- Has **Work**, but model is simpler than WEMI
- First draft, November 2012:
  - More about data structures than “ontology”?
  - Like Application Profiles...?

LOD resource description

- Application Profiles (Dublin Core-style)
  - Use any syntax mappable to Linked Data
  - Use any combination of RDF vocabularies
  - Define “record”-like data packages
  - From very simple descriptions, to complex, depending on requirements...

LOD e-commerce description

- Schema.org
  - Structured information about People, Hotels, Recipes, Books, Movies…
  - Uses RDFa to embed descriptions in Web pages
  - Indexed by Google, Yahoo…
  - Oriented more to high-impact e-commerce than to the “long tail” of scientific and cultural data

LOD concept description

- W3C Web Ontology Language (OWL) models "reality" to support inference
  - If *Dog* is a sub-class *Mammal*, and "Oscar" is a *Dog*, then "Oscar" is a *Mammal*
- Define a "cartoon universe" (*Dogs*, *Cats*, *Mammals*)
- Can be difficult and expensive to design

LOD concept description

- SKOS Concept Schemes
  - W3C **Simple** Knowledge Organization System
  - SKOS pragmatically ***avoids*** supporting inference
  - If *Car* is broader than *Wheel*, and *Machine* is broader than *Car*, *Machine* is **not** "broader than" *Wheel*
- OWL and SKOS provide URIs for Linked Data
  - SKOS is easier, often "good enough"

Making RDF publication easier

- Drupal 7 content management
  - Easy workflows help non-experts publish on Web
  - Uses RDFa embedded in Web pages (like schema.org)
  - Can output other RDF serialization formats (RDF/XML, Turtle, JSONLD…)
  - Support for SPARQL endpoints

Simplicity scales better

- **Many** can agree on **small** standards (SKOS, DC, RDF)
- **A few** can agree on **large complex** standards (FRBR, RDA)
- Rare for **many** to agree on **large complex** standards

Recipe for Interoperability

1. What elements of your data **really** need to interoperate?
2. Map those parts to a core model.
3. Define the core model using **URIs** and **RDF**.
4. Publish the core data elements as Linked Data.

Learning Linked Data

- RDF is a language for data
  - DC, SKOS, LCSH... – the “words” of its dictionaries
  - RDF Triples based on a “sentence grammar”
- Fluency in RDF
  - Language designed for humans for processing by machines
  - Software tools to read, analyze, compose data...
  - “RDF as a second language”

tom@tombaker.org