

Usage data for metadata properties to support open registries and semantic wikis

Muriel Foulonneau

Public Research Centre Henri Tudor, Luxembourg
muriel.foulonneau@tudor.lu

Sébastien Martin, University Paris VIII, France

Jacques Ducloy, France

Thierry Daunois, Lorraine University, France

Slim Turki, Public Research Centre Henri Tudor,
Luxembourg

The need for vocabulary reuse

Many communities use common concepts

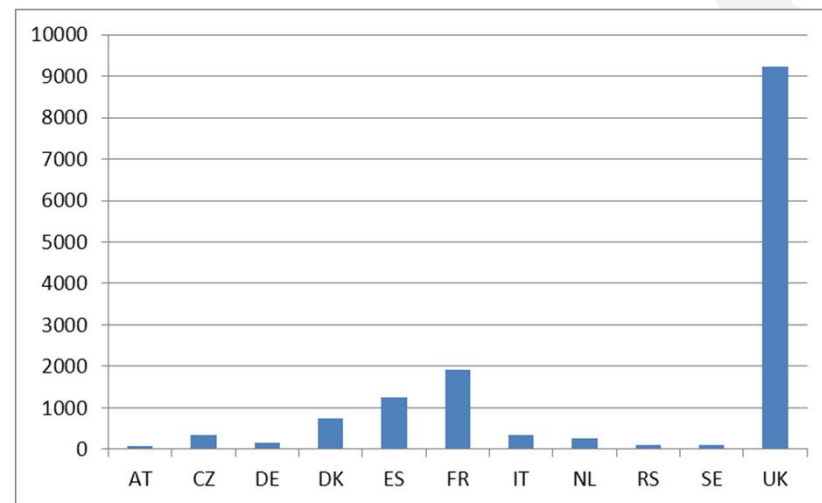
- The cost of new vocabularies (development, documentation, maintenance)
- Interoperability with services

A study on the Public Data EU catalogue

The *PublicData.eu* catalog currently collects data from 26 sources, has been the first to be published in Europe in 2011. It includes data beyond the European Union, e.g., Serbian datasets. Although not exhaustive, it represents a unique aggregation of European datasets.

17 027 datasets

An aggregation of catalogues



Many properties coming from each data source

236 distinct properties

Property	# occurrences
licence	140
License	15
license_url	5
License_details	34
License_ID	26
License_summary	55
License_title	26
License_uri	2
License_url	14
mandate	29

<http://publicdata.eu>

The lack of vocabulary reuse

Why do not people reuse other people's vocabulary?

Hypotheses:

- **I do not know about other people's vocabularies**
- **Not invented here**
- **I do not trust other people's vocabulary**

The Wicri network

Network of scientific wikis
SemanticMediaWiki



Some statistics: 13.155 pages (09/2013)

147 wikis

1 metadata wiki : vocabulary of the various wikis of the network
(i.e. kind of vocabulary registry)

Central maintenance with the help of contributors

The Wicri manager is responsible for the coherence of the
vocabularies used in the wikis

Wicri reuses vocabularies from

Semanticweb.org

e.g., « Has PC member »

Eurovoc

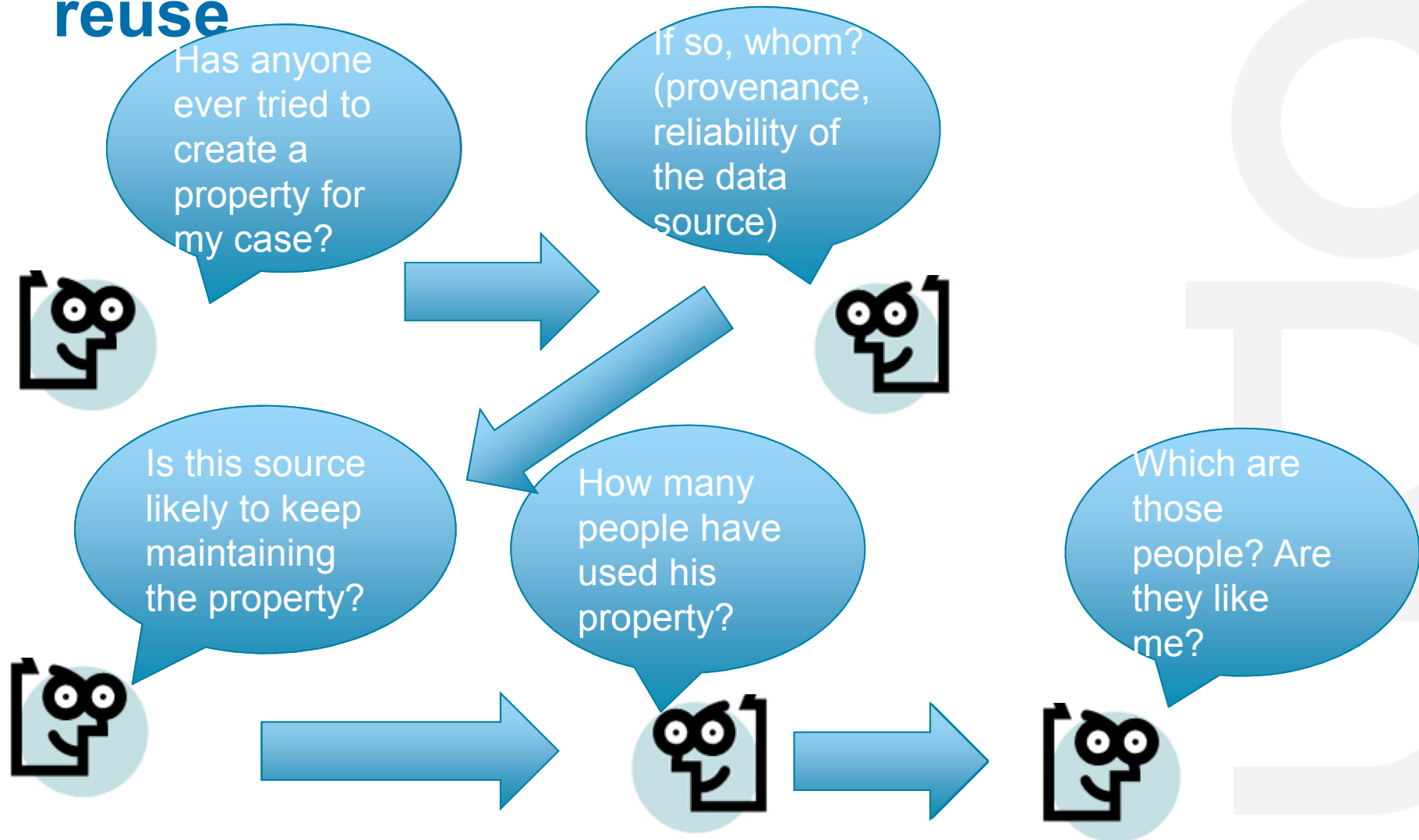
MeSH



Problems:

- when the same concept is used in different vocabularies
- concept ambiguity
- when reuse and when create a new concept
- coordinating generic and domain specific vocabularies

Decision making model for property reuse



Current vocabulary registries

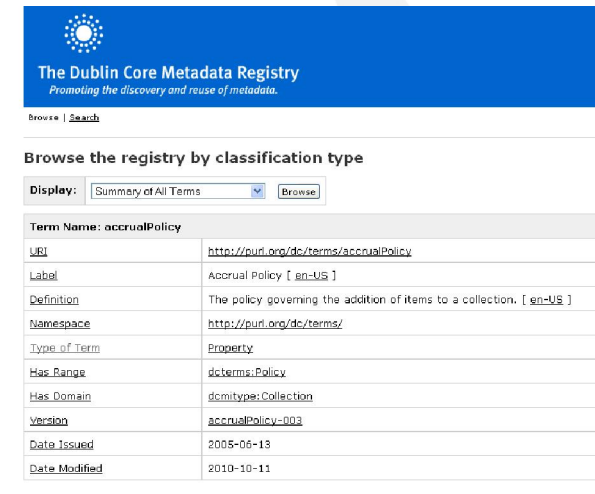
In order to assess an open source software project:

- production data (e.g., new releases)
- usage data (e.g., number of downloads)
- communication data (e.g., mailing list activity) can be analyzed

Current registries support descriptive metadata and some production data (releases)

=> They do not support usage data

=> **How to find usage data?**



The screenshot shows the Dublin Core Metadata Registry website. At the top is a blue header with the logo and text "The Dublin Core Metadata Registry Promoting the discovery and reuse of metadata." Below this is a search bar with "Browse" and "Search" buttons. The main content area is titled "Browse the registry by classification type" and includes a "Display:" dropdown menu set to "Summary of All Terms" and a "Browse" button. Below this is a table for the term "accrualPolicy".

Term Name: accrualPolicy	
URI	http://purl.org/dc/terms/accrualPolicy
Label	Accrual Policy [en-US]
Definition	The policy governing the addition of items to a collection. [en-US]
Namespace	http://purl.org/dc/terms/
Type of Term	Property
Has Range	dcterms:Policy
Has Domain	dcmitype:Collection
Version	accrualPolicy-003
Date Issued	2005-06-13
Date Modified	2010-10-11

Ontology Repository

Ontologies

Displaying 219 of 219 ontologies [Filter](#)

[\[Select metrics\]](#)

DL Expressivity	Logical Axioms	Ontology	Download
ALCHOIN	108	people	RDF/XML Browse
EL++	696	pathway.cbo	RDF/XML Browse
ALCOIF	18775	fmaOwlDlComponent_1_4_0	RDF/XML Browse
SHIN(D)	3539	1.0	RDF/XML Browse

Known properties containing the term “license” in Sindice

[? license] [? enter a UR]

- <http://www.w3.org/1999/xhtml/vocab#license>
- <http://creativecommons.org/ns#license>
- <http://purl.org/dc/terms/license>
- <http://web.resource.org/cc/license>
- <http://backend.userland.com/creativecommonsRssModulelicense>
- <http://usefulinc.com/ns/doap#license>

HCARD [v]

Occurrences of the <foaf:mbox> property on Sindice

SEARCH

Group By Dataset: ☐

Sorted by: relevance ▼

Sindice search: (* <foaf:mbox> *) found 41,925 documents (in 0.79 seconds)

<http://trust.mindswap.org/trustFiles/out2.owl> (RDF)

+ 2010-04-14 – 568 triples in 75.2 kB

<http://trust.mindswap.org/trustFiles/out2.owl> ([Search](#)) Inspect: ([Cache](#)) ([Live](#))

<http://www.ecs.soton.ac.uk/~nmg/nmg-foaf.rdf> (RDF)

+ 2010-06-08 – 166 triples in 19.3 kB

<http://www.ecs.soton.ac.uk/~nmg/nmg-foaf.rdf> ([Search](#)) Inspect: ([Cache](#)) ([Live](#))

The Linked Open Vocabularies registry and JoinUp (at vocabulary level) provide statistics

The screenshot shows the Linked Open Vocabularies (LOV) registry interface. At the top, there's a header with the LOV logo and the text 'Linked Open Vocabularies (LOV)'. Below the header, there's a search bar with a 'Search' button. On the left, there's a 'Filter by Domain' section with a list of domains and their counts: City (5), Data & Systems (5), Library (19), Market (0), Media (10), Science (28), Upper & Meta (40), and Where & When (2). Below this is a 'Filter by Type' section with a list of types and their counts: rdfs:Class (22), rdfs:Property (92), voaf:Vocabulary (2), and Other (7). The main content area shows a search result for 'dcterms:creator'. A pop-up window titled 'Information about dcterms:creator' is displayed, showing metrics and element information.

Information about dcterms:creator

Metrics:

Metric	Value
Overall score with current parameters	0.763
Best ratio search words in labels	1.000
Nb occurrences in main labels	1
Nb occurrences in secondary labels	1
Nb occurrences in LOD datasets	7371553
Nb occurrences in LOV datasets	93
Nb of vocabularies which reference this element	57

Element Information:

Property	Value
URI	http://purl.org/dc/terms/creator
rdfs:type	owl:Thing > rdf:Property > owl:AnnotationProperty > owl:ObjectProperty
vocabulary	dcterms
is dc:hasPart of	All > Upper & Meta > Metadata
rdfs:label	Creator @en
dcterms:description	Examples of a Creator include a person, an organization, or a service. @en



DCAT application profile for data portals in Europe



Download

(5 ★★★★★)



Editor's choice

DCAT Application Profile for Data Portals in Europe - Final

Rating: 5/5 (based on 1 votes) | 342 reads

Description

The DCAT Application profile for data portals in Europe (DCAT-AP) is a specification based on the Data Catalogue vocabulary ([DCAT](#)) for describing public sector datasets in Europe. Its basic use case is to enable a cross-data portal search for data sets and

Conclusion

Current registries aim at property discovery

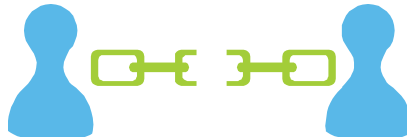
But contain mainly descriptive data, data on maintenance authority and production data

Miss key information regarding usage (and communication)

⇒ **Could we follow the decision making process for vocabulary reuse?**

⇒ i.e., join vocabulary registries and semantic search engines?

And provide a view of who uses what in which context?



Some references

Jacques Ducloy, Thierry Daunois, Muriel Foulonneau, Alice Hermann, Jean-Charles Lamirel, Stéphane Sire, Jean-Pierre Thomesse, & Christine Vanoirbeek. Metadata for WICRI, a Network of Semantic Wikis for Communities in Research and Innovation. Dublin Core conference 2010, Pittsburgh

Martin, S., Foulonneau, M., Turki, S. (2013). 1-5 stars: Metadata on the Openness Level of Open Data Sets in Europe. In Metadata and Semantics Research 2013. Communications in Computer and Information Science, Springer Verlag, Heidelberg