Building a DCAT Merger

Pieter Heyvaert
Pieter Colpaert, Ruben Verborgh, Erik Mannens & Rik Van de Walle
Summary

What’s DCAT?

3 simple Linked Data problems

A patch: we’re launching DCAT Merger

Now What?
Distributing catalogs over the Web

- dcat:Catalog
  - dcat:Dataset
    - dcat:Distribution

Distributing catalogs over the Web
3 simple problems

ex1:Catalog a dcat:Catalog ;
    dcat:themeTaxonomy ex1:Tax ;
    dcat:dataset ex1:Dataset1 ;
    dct:creator [ foaf:name “Pieter Heyvaert” ].

ex1:Dataset1 a dcat:Dataset ;
    dcat:theme eurovoc:215481 .

ex2:Catalog a dcat:Catalog ;
    dcat:themeTaxonomy ex2:Tax ;
    dcat:dataset ex2:Dataset1 ;
    dct:creator [ schema:name “Pieter Heyvaert” ].

ex2:Dataset1 a dcat:Dataset ;
    dcat:theme dbpedia:Education .

1

Different ways of adding additional information
3 simple problems

SPARQL1.1

RDFa

Custom Web API

Different protocols to retrieve data
3 simple problems

What can we rely on in all castles?

“Local Querying” is not Web Querying
DCAT Merger (1)

Ghent
DCAT feed

Portorož
DCAT feed

London
DCAT feed

DCAT Merger
Node.js package

Merged & Enriched
DCAT feed
Enriching Feeds

Keywords
DCAT feed

Description
DCAT feed

NER

Subjects

Themes

Spatial Coverage

Custom
plugins
Advantages

different way of adding additional information → enriched feeds

different protocols → single feed

not possible to query all at the same time → single feed
Now What? (1)

Is there a better way to discover data catalogues?

Do we need to create a single point where all feeds URIs are collected? (distributed)

Is a hierarchy of DCAT Mergers a solution?
Now What? (2)

Can we consider republishing the (enhanced) feeds as a good thing?

Do we need to define protocol guidelines for the creation of data catalogues?