Linked Open Data in Agriculture
MACS-G20 Workshop in Berlin, September 27th–28th, 2017

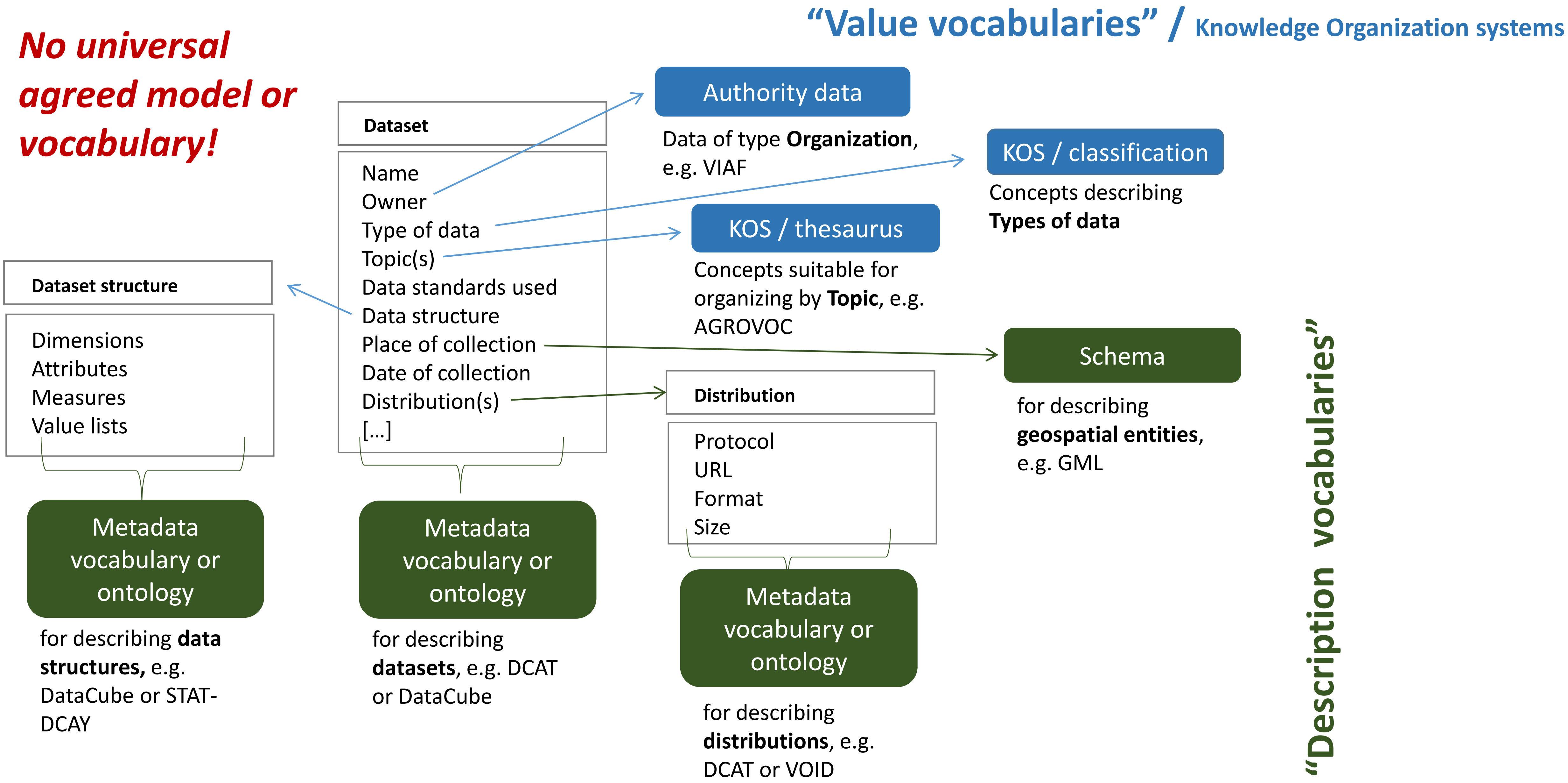
Semantic challenges in sharing dataset metadata and creating federated dataset catalogs

The example of the CIARD RING

Valeria Pesce (Global Forum on Agricultural Research and Innovation)

Semantics involved in describing datasets

No universal agreed model or vocabulary!



“Value vocabularies” / Knowledge Organization systems

Dataset

Name
Owner
Type of data
Topic(s)
Data standards used
Data structure
Place of collection
Date of collection
Distribution(s)
[…]

Metadata vocabulary or ontology

for describing **datasets**, e.g. DCAT or DataCube

Dataset structure

Dimensions
Attributes
Measures
Value lists

Metadata vocabulary or ontology

for describing **data structures,** e.g. DataCube or STAT-DCAY

Authority data

Data of type **Organization**, e.g. VIAF

KOS / classification

Concepts describing **Types of data**

KOS / thesaurus

Concepts suitable for organizing by **Topic**, e.g. AGROVOC

Schema

for describing **geospatial entities**, e.g. GML

Distribution

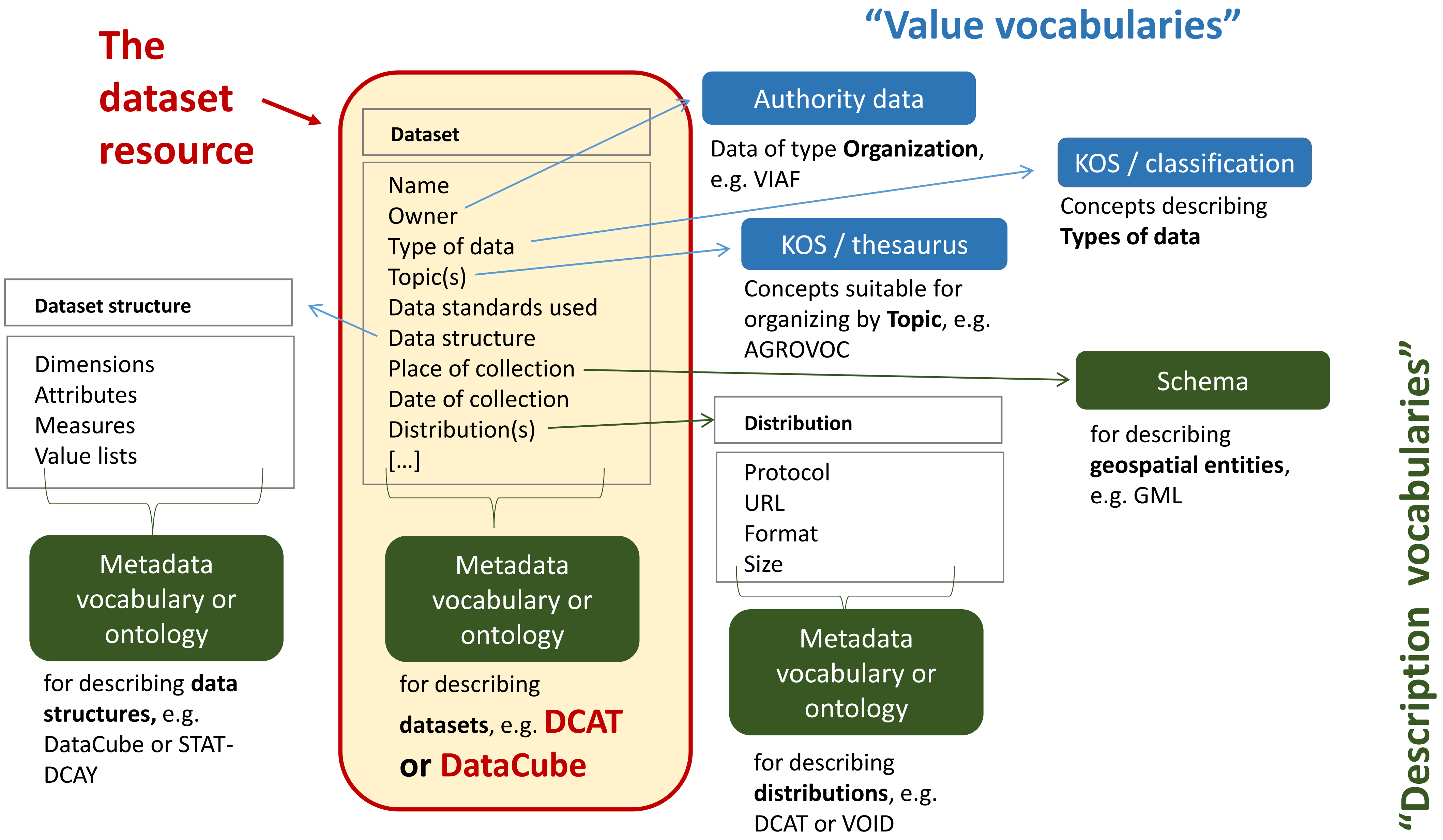
Protocol
URL
Format
Size

Metadata vocabulary or ontology

for describing **distributions**, e.g. DCAT or VOID

“Description vocabularies”

Semantics involved in describing datasets



The dataset resource

“Value vocabularies”

Dataset

Name
Owner
Type of data
Topic(s)
Data standards used
Data structure
Place of collection
Date of collection
Distribution(s)
[…]

Metadata vocabulary or ontology

for describing **datasets**, e.g. **DCAT or DataCube**

Authority data

Data of type **Organization**, e.g. VIAF

KOS / classification

Concepts describing **Types of data**

KOS / thesaurus

Concepts suitable for organizing by **Topic**, e.g. AGROVOC

Dataset structure

Dimensions
Attributes
Measures
Value lists

Metadata vocabulary or ontology

for describing **data structures,** e.g. DataCube or STAT-DCAY

Schema

for describing **geospatial entities**, e.g. GML

Distribution

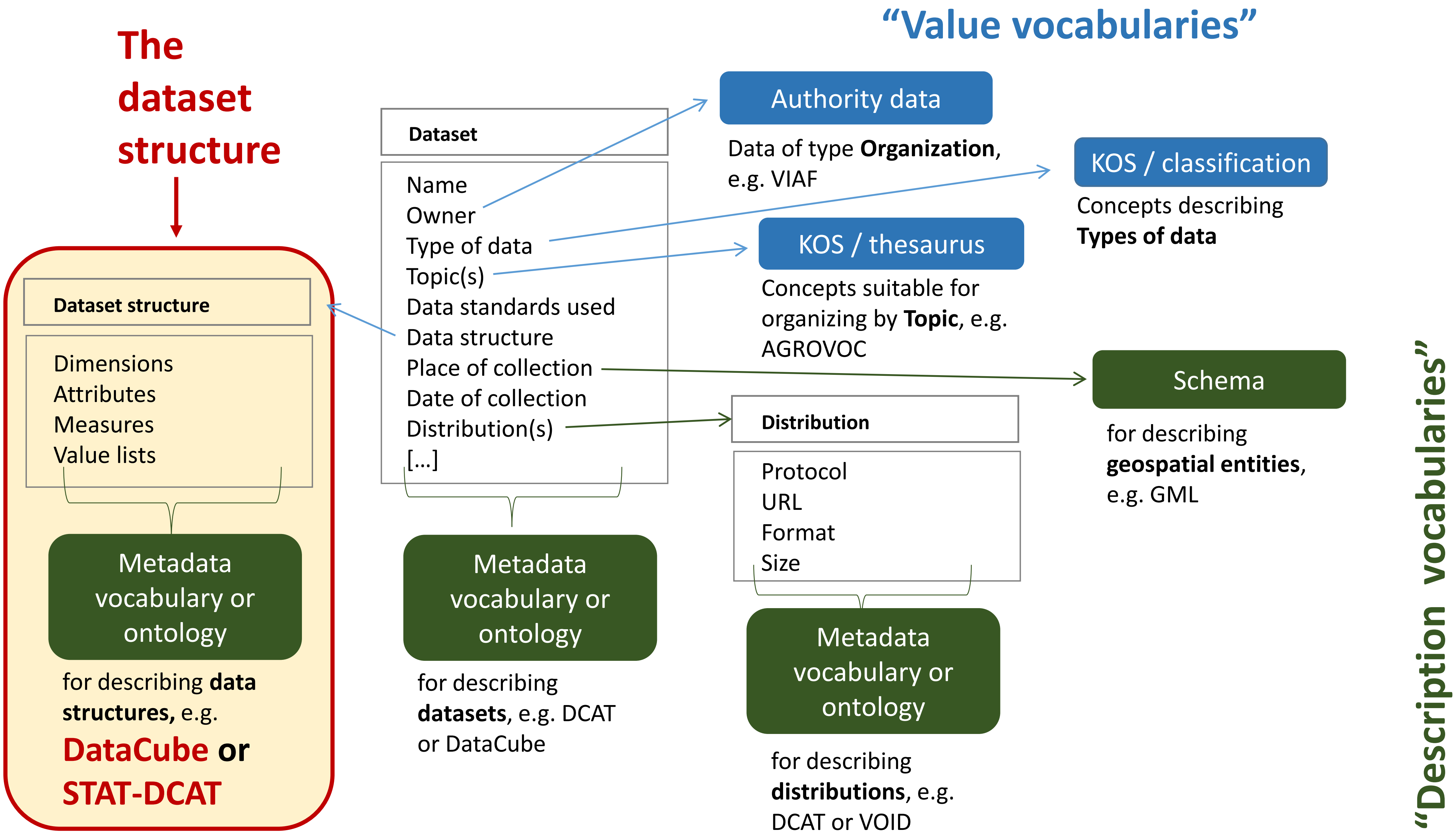
Protocol
URL
Format
Size

Metadata vocabulary or ontology

for describing **distributions**, e.g. DCAT or VOID

“Description vocabularies”

Semantics involved in describing datasets



The dataset structure

Dataset structure

Dimensions
Attributes
Measures
Value lists

Metadata vocabulary or ontology

for describing **data structures,** e.g. **DataCube or STAT-DCAT**

Dataset

Name
Owner
Type of data
Topic(s)
Data standards used
Data structure
Place of collection
Date of collection
Distribution(s)
[…]

Metadata vocabulary or ontology

for describing **datasets**, e.g. DCAT or DataCube

"Value vocabularies"

Authority data

Data of type **Organization**, e.g. VIAF

KOS / classification

Concepts describing **Types of data**

KOS / thesaurus

Concepts suitable for organizing by **Topic**, e.g. AGROVOC

Schema

for describing **geospatial entities**, e.g. GML

Distribution

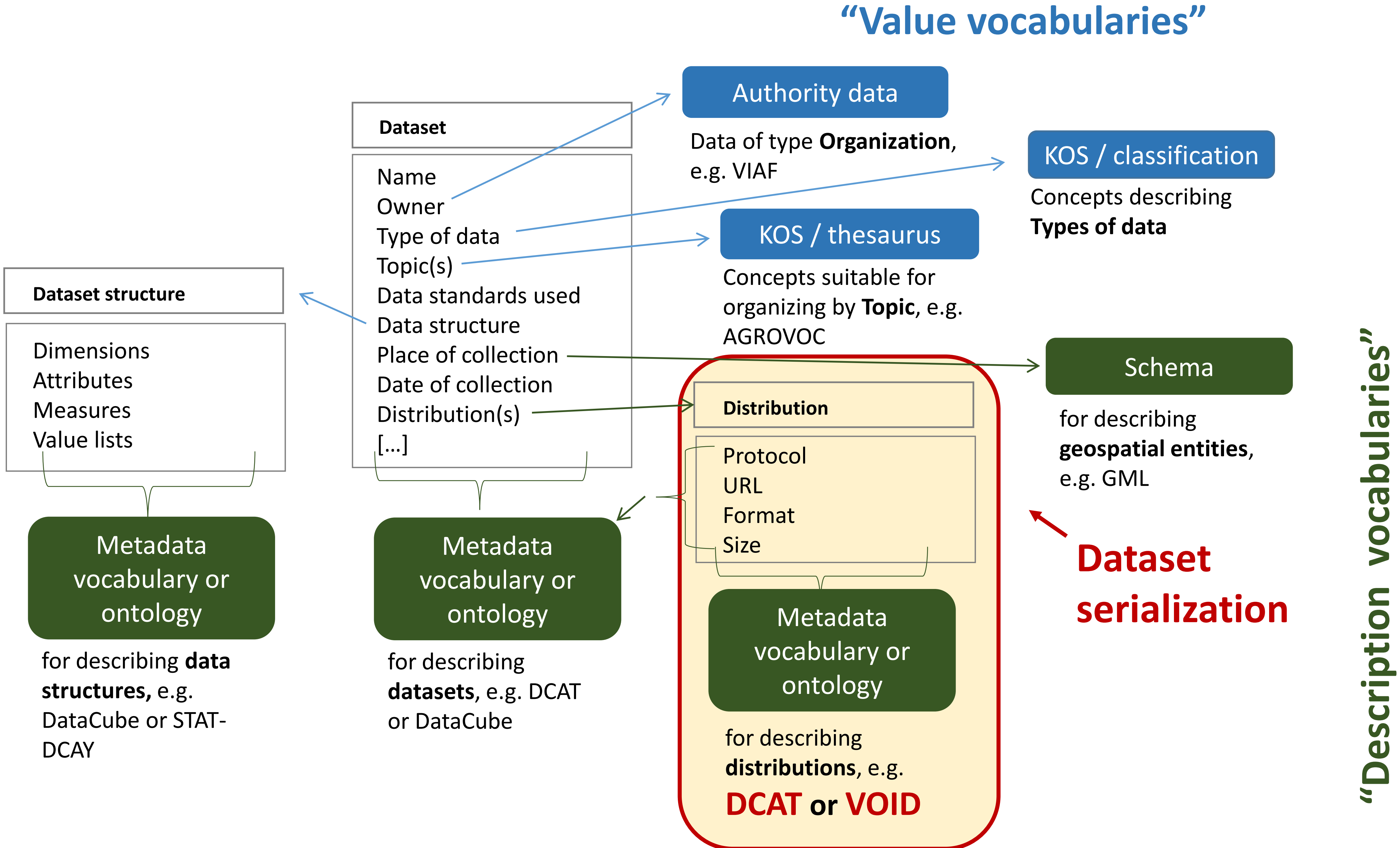
Protocol
URL
Format
Size

Metadata vocabulary or ontology

for describing **distributions**, e.g. DCAT or VOID

"Description vocabularies"

Semantics involved in describing datasets



“Value vocabularies”

Dataset

Name
Owner
Type of data
Topic(s)
Data standards used
Data structure
Place of collection
Date of collection
Distribution(s)
[…]

Metadata vocabulary or ontology

for describing **datasets**, e.g. DCAT or DataCube

Dataset structure

Dimensions
Attributes
Measures
Value lists

Metadata vocabulary or ontology

for describing **data structures,** e.g. DataCube or STAT-DCAY

Authority data

Data of type **Organization**, e.g. VIAF

KOS / classification

Concepts describing **Types of data**

KOS / thesaurus

Concepts suitable for organizing by **Topic**, e.g. AGROVOC

Schema

for describing **geospatial entities**, e.g. GML

Distribution

Protocol
URL
Format
Size

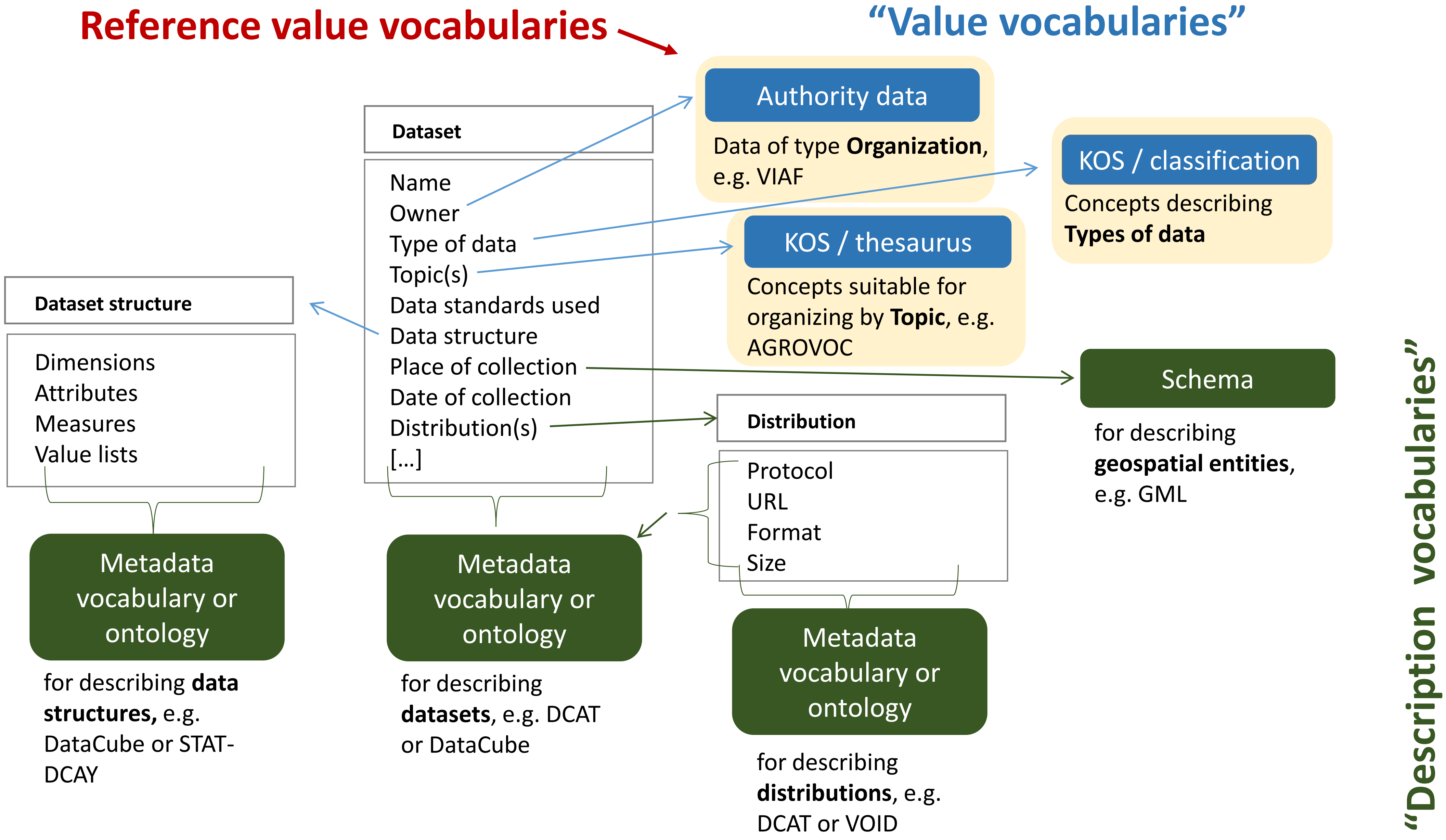
Metadata vocabulary or ontology

for describing **distributions**, e.g. **DCAT or VOID**

Dataset serialization

“Description vocabularies”

Semantics needed to describe datasets



Reference value vocabularies

“Value vocabularies”

Authority data

Data of type **Organization**, e.g. VIAF

KOS / classification

Concepts describing **Types of data**

KOS / thesaurus

Concepts suitable for organizing by **Topic**, e.g. AGROVOC

Dataset

Name
Owner
Type of data
Topic(s)
Data standards used
Data structure
Place of collection
Date of collection
Distribution(s)
[...]

Dataset structure

Dimensions
Attributes
Measures
Value lists

Distribution

Protocol
URL
Format
Size

Schema

for describing **geospatial entities**, e.g. GML

Metadata vocabulary or ontology

for describing **data structures,** e.g. DataCube or STAT-DCAY

Metadata vocabulary or ontology

for describing **datasets**, e.g. DCAT or DataCube

Metadata vocabulary or ontology

for describing **distributions**, e.g. DCAT or VOID

“Description vocabularies”

Semantics of the values

Thematic metadata

KOS / thesaurus

Examples:
- **AGROVOC**
- **CABI thesaurus**

Geographic metadata

Authority data

Examples:
- **GeoNames**
- **FAO Geopol Ontology**

Dimensions

Code list

Examples:
- **ICASA** variables
- **CF conventions** RDF

Publisher metadata

Authority data

Examples:
- **VIAF** registry
- **Library of Congress**
- **ORCID**

- Standardization of the values, e.g. for “thematic coverage” or “dimensions” of datasets, “format” or “protocol used” of distributions etc.
- The value should be standardized, possibly a URI
- The value should be part of an authority list / code list

RDF dataset vocabularies normally treat these values as resources, so identifiable by URIs, BUT…

a) Often strings are used

b) Often a local concept URI is used

c) THERE AREN’T AGREED KOSs FOR EVERYTHING!

Examples of relevant value vocabularies

Not for everything we would need!

- Domain
  - Agricultural concepts, topics: AGROVOC > GACS (or agreed subsets)
  - Crop names: AGROVOC, Crop Ontology
  - Soil types: USDA Soil Taxonomy, INSPIRE Registry
  - Dimensions / variables: ICASA variables (→ RDF?), CF conventions RDF
- Cross-domain
  - Authority lists of organizations, projects: VIAF, CERIF, ORCID?
  - Geospatial / geopolitical data: GeoNames, FAO Geopolitical Ontology
  - Data formats / data standards? AgriSemantics Map of Standards
  - File formats: IANA types (→ RDF?), W3C formats
  - Agreed list of types of data?
  - Units of measure?
  - Authority list of licenses (OpenDefinition list?)

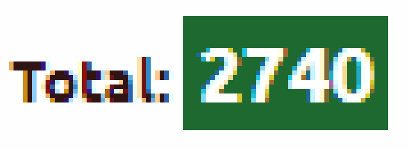
The CIARD RING

The CIARD RING is a **federated and curated catalog of agri-food datasets** and data services

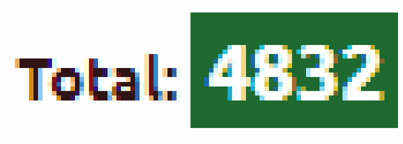
http://ring.ciard.net

- a **primary catalog** (providers can catalog individual data services and datasets directly in the RING) exposing all metadata as RDF
- a **federated catalog** (it harvests dataset metadata from other catalogs)

CIARD RING


Total: 2740

datasets


Total: 4832

services

Federated catalogs so far

datahub

EU Open Data Portal
Access to European Union open data

IFPRI
Dataverse catalog

DATA.GOV.UK Beta
Opening up Government

DATA.GOV

Semantics in the RING dataset hub

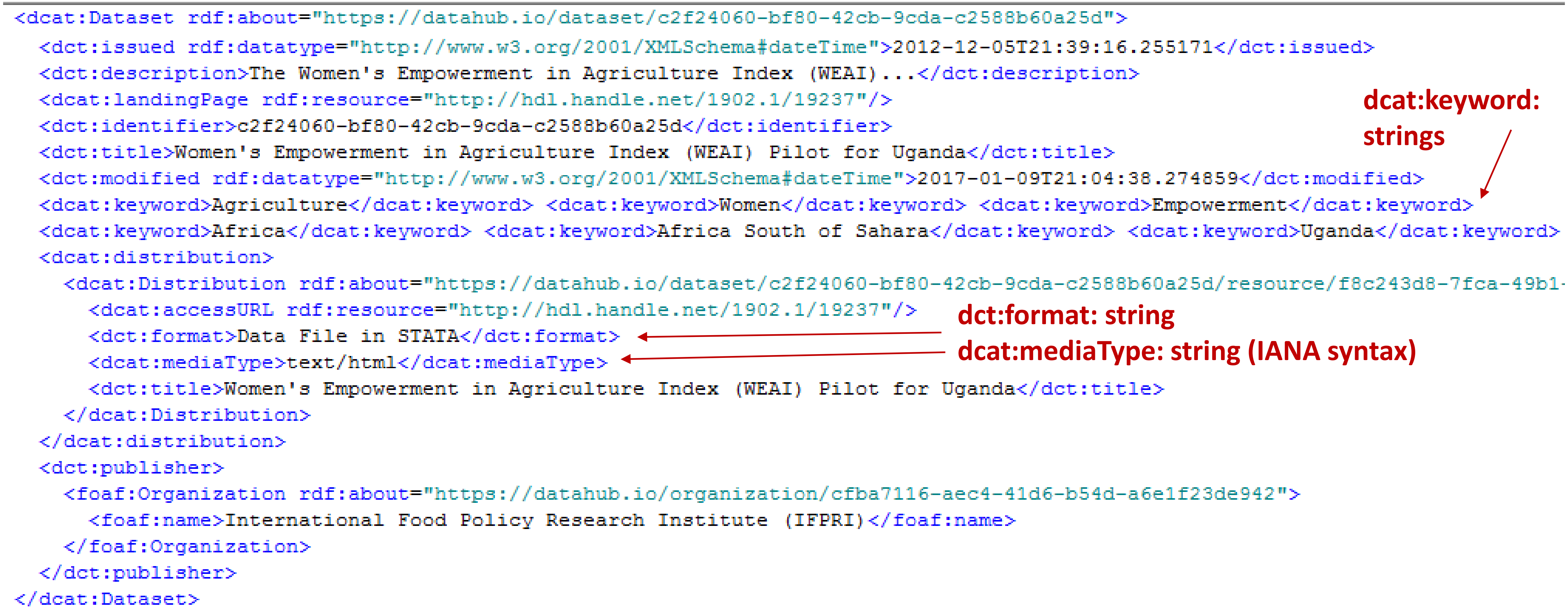
- **Dataset description vocabularies**
  The RING uses a combination of the **DCAT-AP** model + the **VOID** vocabulary and the **DataCube** vocabulary

  → a "RING DCAT profile" will be published

- **Value vocabularies**
  - Domains: **local** RING Domains **SKOS**, based on FAO and USDA top-level classifications of domains
  - Types of data: **local** RING ”Types of data” **SKOS**, aligned with GODAN Ag Sector Package types of data
  - Topics: **AGROVOC**
  - Countries: **FAO Geopolitical Ontology**
  - Data formats / data standards: **AgriSemantics Map of data standards**
  - File formats: "mapped" to IANA types and W3C formats when applicable

Examples of semantics in federated datasets - 1

- IFPRI dataset in Datahub (DCAT RDF)



```
<dcat:Dataset rdf:about="https://datahub.io/dataset/c2f24060-bf80-42cb-9cda-c2588b60a25d">
  <dct:issued rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2012-12-05T21:39:16.255171</dct:issued>
  <dct:description>The Women's Empowerment in Agriculture Index (WEAI)...</dct:description>
  <dcat:landingPage rdf:resource="http://hdl.handle.net/1902.1/19237"/>
  <dct:identifier>c2f24060-bf80-42cb-9cda-c2588b60a25d</dct:identifier>
  <dct:title>Women's Empowerment in Agriculture Index (WEAI) Pilot for Uganda</dct:title>
  <dct:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2017-01-09T21:04:38.274859</dct:modified>
  <dcat:keyword>Agriculture</dcat:keyword> <dcat:keyword>Women</dcat:keyword> <dcat:keyword>Empowerment</dcat:keyword>
  <dcat:keyword>Africa</dcat:keyword> <dcat:keyword>Africa South of Sahara</dcat:keyword> <dcat:keyword>Uganda</dcat:keyword>
  <dcat:distribution>
    <dcat:Distribution rdf:about="https://datahub.io/dataset/c2f24060-bf80-42cb-9cda-c2588b60a25d/resource/f8c243d8-7fca-49b1-
      <dcat:accessURL rdf:resource="http://hdl.handle.net/1902.1/19237"/>
      <dct:format>Data File in STATA</dct:format>
      <dcat:mediaType>text/html</dcat:mediaType>
      <dct:title>Women's Empowerment in Agriculture Index (WEAI) Pilot for Uganda</dct:title>
    </dcat:Distribution>
  </dcat:distribution>
  <dct:publisher>
    <foaf:Organization rdf:about="https://datahub.io/organization/cfba7116-aec4-41d6-b54d-a6e1f23de942">
      <foaf:name>International Food Policy Research Institute (IFPRI)</foaf:name>
    </foaf:Organization>
  </dct:publisher>
</dcat:Dataset>
```

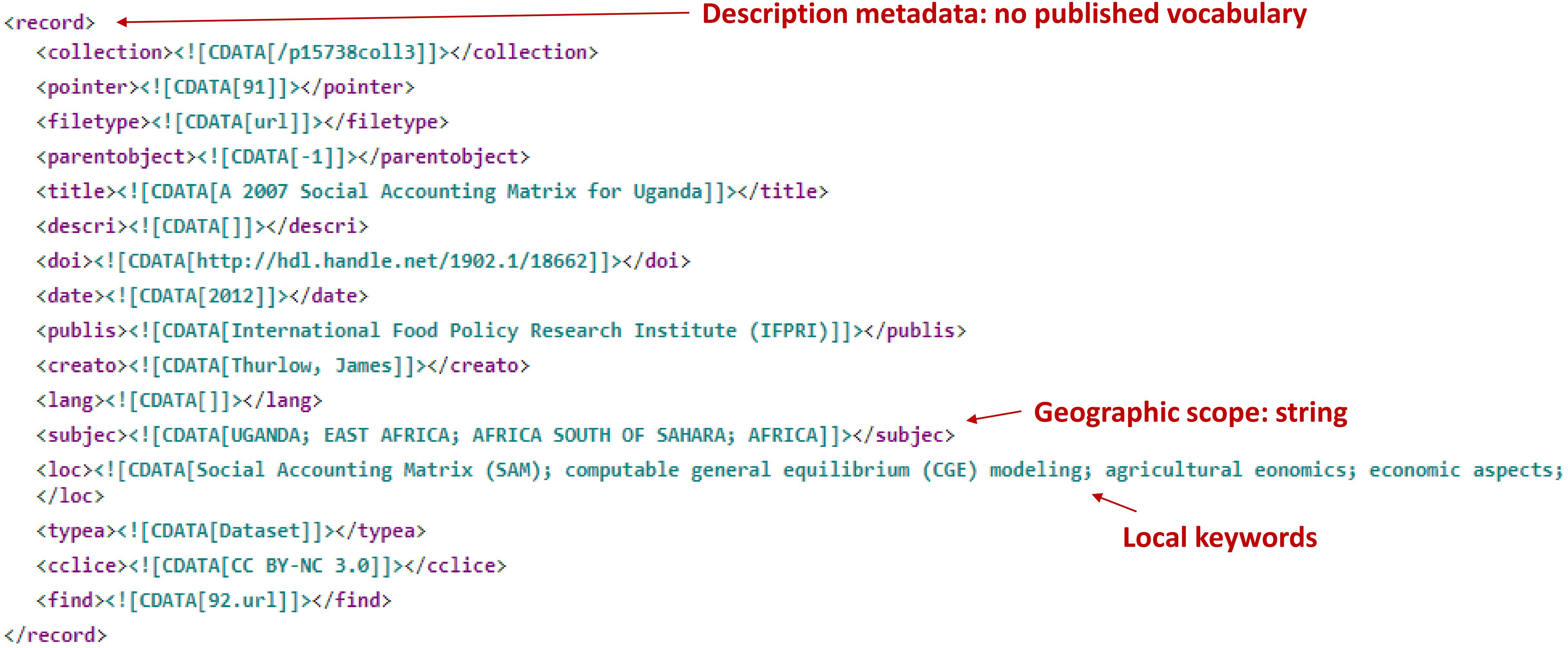
dcat:keyword: strings

dct:format: string

dcat:mediaType: string (IANA syntax)

Examples of semantics in federated datasets - 2

- IFPRI dataset in Dataverse (OAI-PMH XML response)



```
<record>
   <collection><![CDATA[/p15738coll3]]></collection>
   <pointer><![CDATA[91]]></pointer>
   <filetype><![CDATA[url]]></filetype>
   <parentobject><![CDATA[-1]]></parentobject>
   <title><![CDATA[A 2007 Social Accounting Matrix for Uganda]]></title>
   <descri><![CDATA[]]></descri>
   <doi><![CDATA[http://hdl.handle.net/1902.1/18662]]></doi>
   <date><![CDATA[2012]]></date>
   <publis><![CDATA[International Food Policy Research Institute (IFPRI)]]></publis>
   <creato><![CDATA[Thurlow, James]]></creato>
   <lang><![CDATA[]]></lang>
   <subjec><![CDATA[UGANDA; EAST AFRICA; AFRICA SOUTH OF SAHARA; AFRICA]]></subjec>
   <loc><![CDATA[Social Accounting Matrix (SAM); computable general equilibrium (CGE) modeling; agricultural eonomics; economic aspects;
   </loc>
   <typea><![CDATA[Dataset]]></typea>
   <cclice><![CDATA[CC BY-NC 3.0]]></cclice>
   <find><![CDATA[92.url]]></find>
</record>
```

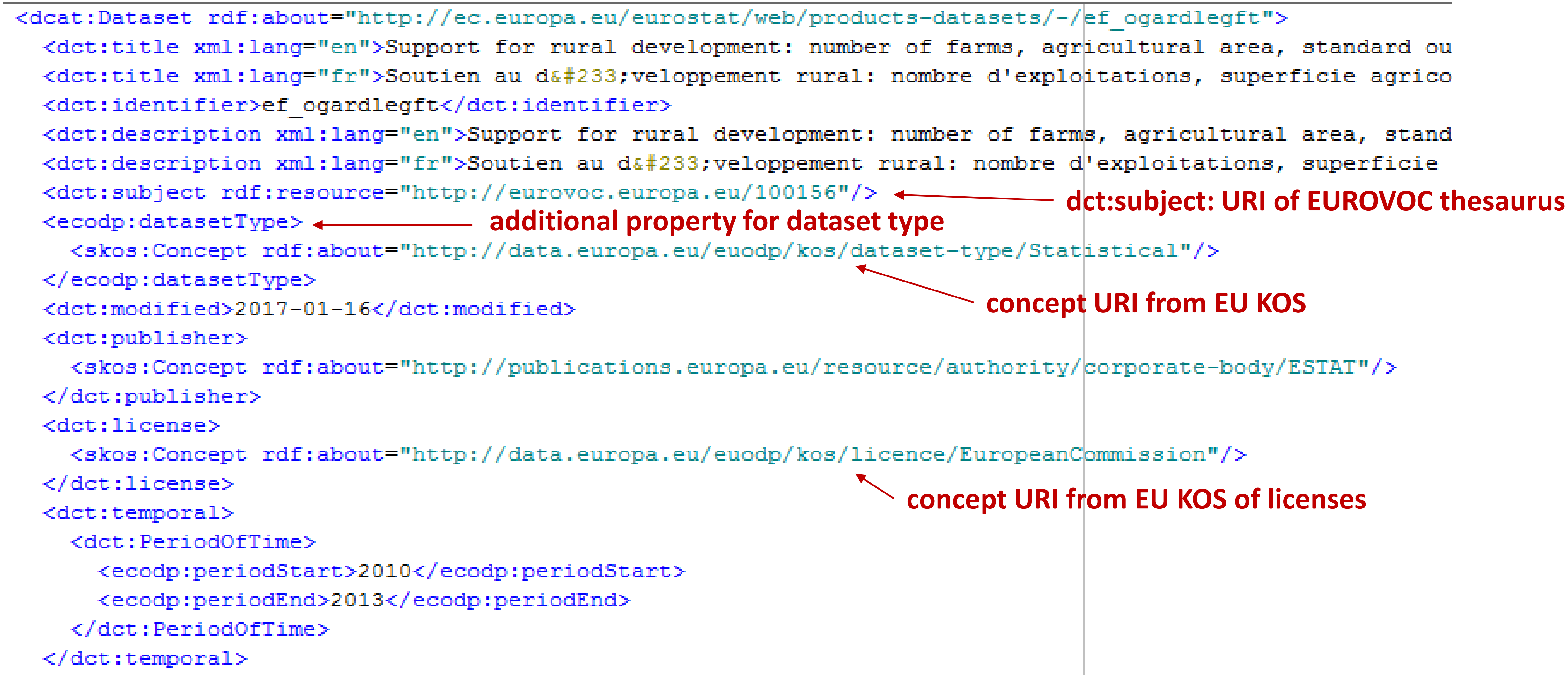
Description metadata: no published vocabulary

Geographic scope: string

Local keywords

Examples of semantics in federated datasets – 3a

- EuroStat dataset in EU Data Portal (DCAT RDF) (1)



```
<dcat:Dataset rdf:about="http://ec.europa.eu/eurostat/web/products-datasets/-/ef_ogardlegft">
  <dct:title xml:lang="en">Support for rural development: number of farms, agricultural area, standard ou
  <dct:title xml:lang="fr">Soutien au d&#233;veloppement rural: nombre d'exploitations, superficie agrico
  <dct:identifier>ef_ogardlegft</dct:identifier>
  <dct:description xml:lang="en">Support for rural development: number of farms, agricultural area, stand
  <dct:description xml:lang="fr">Soutien au d&#233;veloppement rural: nombre d'exploitations, superficie
  <dct:subject rdf:resource="http://eurovoc.europa.eu/100156"/>
  <ecodp:datasetType>
    <skos:Concept rdf:about="http://data.europa.eu/euodp/kos/dataset-type/Statistical"/>
  </ecodp:datasetType>
  <dct:modified>2017-01-16</dct:modified>
  <dct:publisher>
    <skos:Concept rdf:about="http://publications.europa.eu/resource/authority/corporate-body/ESTAT"/>
  </dct:publisher>
  <dct:license>
    <skos:Concept rdf:about="http://data.europa.eu/euodp/kos/licence/EuropeanCommission"/>
  </dct:license>
  <dct:temporal>
    <dct:PeriodOfTime>
      <ecodp:periodStart>2010</ecodp:periodStart>
      <ecodp:periodEnd>2013</ecodp:periodEnd>
    </dct:PeriodOfTime>
  </dct:temporal>
```

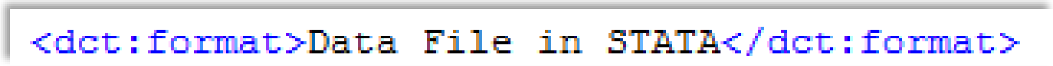
dct:subject: URI of EUROVOC thesaurus

additional property for dataset type

concept URI from EU KOS

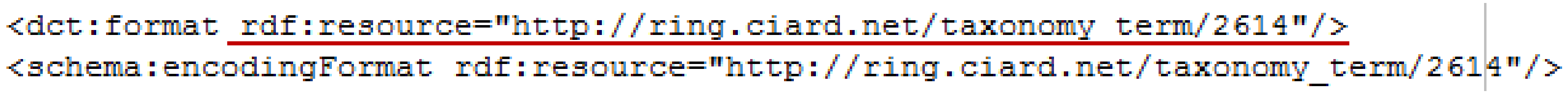
concept URI from EU KOS of licenses

RING: enriching and linking semantics

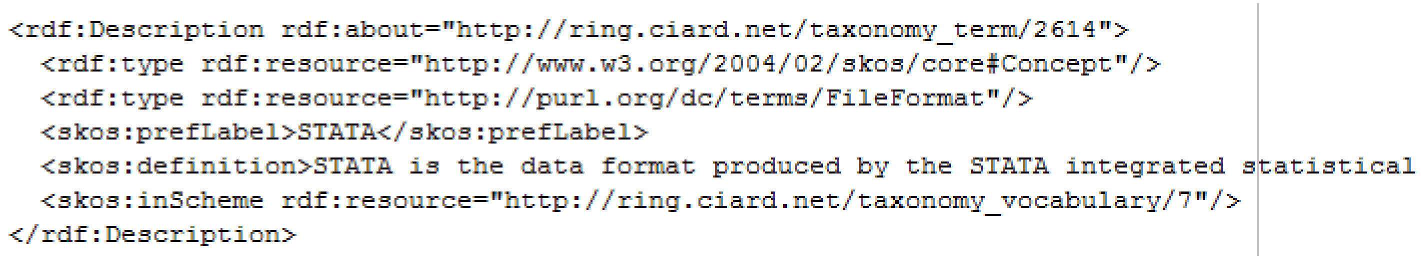


```
<dct:format>Data File in STATA</dct:format>
```

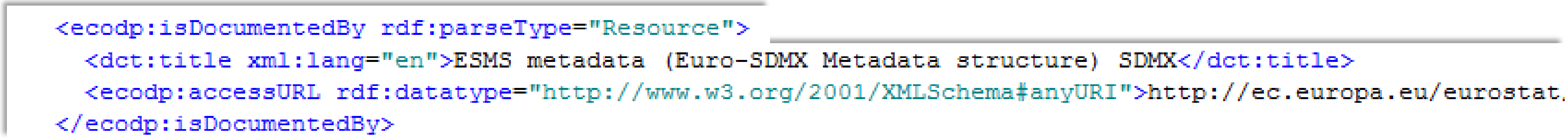
Match or partial match with synonym in local KOS >> Becomes a RING resource of type skos:Concept and dc:FileFormat with local URI



```
<dct:format rdf:resource="http://ring.ciard.net/taxonomy_term/2614"/>
<schema:encodingFormat rdf:resource="http://ring.ciard.net/taxonomy_term/2614"/>
```

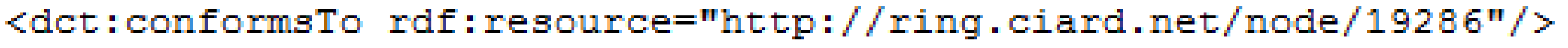


```
<rdf:Description rdf:about="http://ring.ciard.net/taxonomy_term/2614">
  <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
  <rdf:type rdf:resource="http://purl.org/dc/terms/FileFormat"/>
  <skos:prefLabel>STATA</skos:prefLabel>
  <skos:definition>STATA is the data format produced by the STATA integrated statistical
  <skos:inScheme rdf:resource="http://ring.ciard.net/taxonomy_vocabulary/7"/>
</rdf:Description>
```

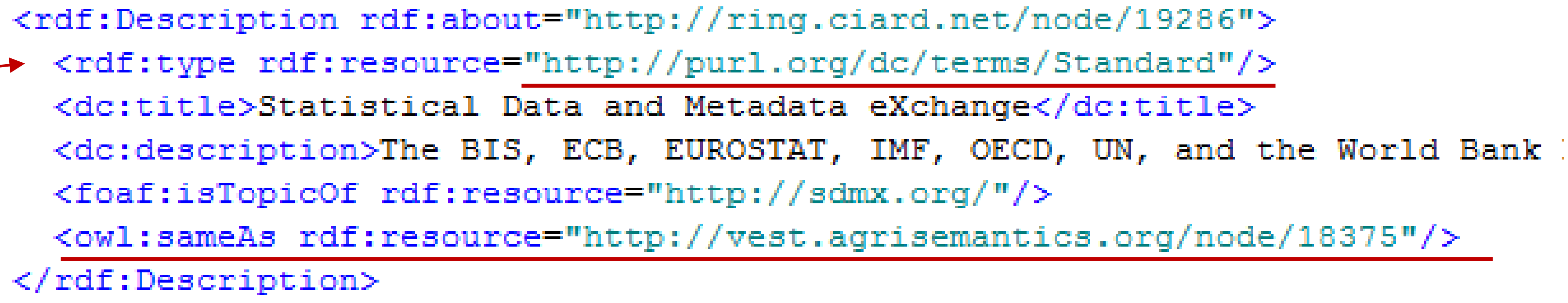


```
  <ecodp:isDocumentedBy rdf:parseType="Resource">
    <dct:title xml:lang="en">ESMS metadata (Euro-SDMX Metadata structure) SDMX</dct:title>
    <ecodp:accessURL rdf:datatype="http://www.w3.org/2001/XMLSchema#anyURI">http://ec.europa.eu/eurostat
  </ecodp:isDocumentedBy>
```

Becomes dct:conformsTo as in DCAT with resource of type dc:Standard linked to URI of same standard in AgriSemantics Map of Data Standards

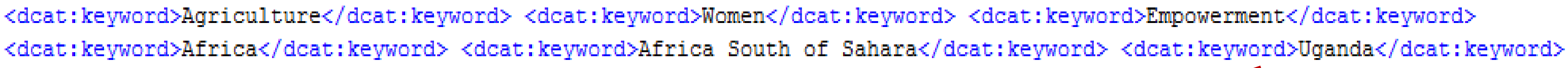


```
<dct:conformsTo rdf:resource="http://ring.ciard.net/node/19286"/>
```

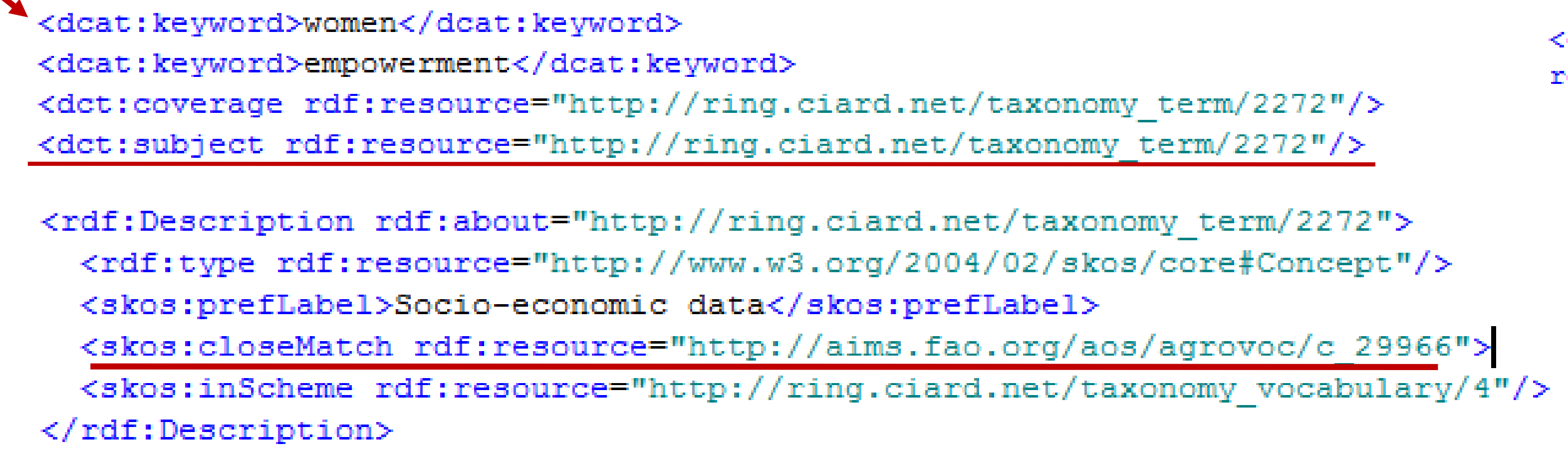


```
<rdf:Description rdf:about="http://ring.ciard.net/node/19286">
  <rdf:type rdf:resource="http://purl.org/dc/terms/Standard"/>
  <dc:title>Statistical Data and Metadata eXchange</dc:title>
  <dc:description>The BIS, ECB, EUROSTAT, IMF, OECD, UN, and the World Bank
  <foaf:isTopicOf rdf:resource="http://sdmx.org/"/>
  <owl:sameAs rdf:resource="http://vest.agrisemantics.org/node/18375"/>
</rdf:Description>
```

RING: linking semantics

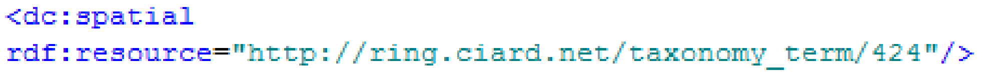


```
<dcat:keyword>Agriculture</dcat:keyword> <dcat:keyword>Women</dcat:keyword> <dcat:keyword>Empowerment</dcat:keyword>
<dcat:keyword>Africa</dcat:keyword> <dcat:keyword>Africa South of Sahara</dcat:keyword> <dcat:keyword>Uganda</dcat:keyword>
```



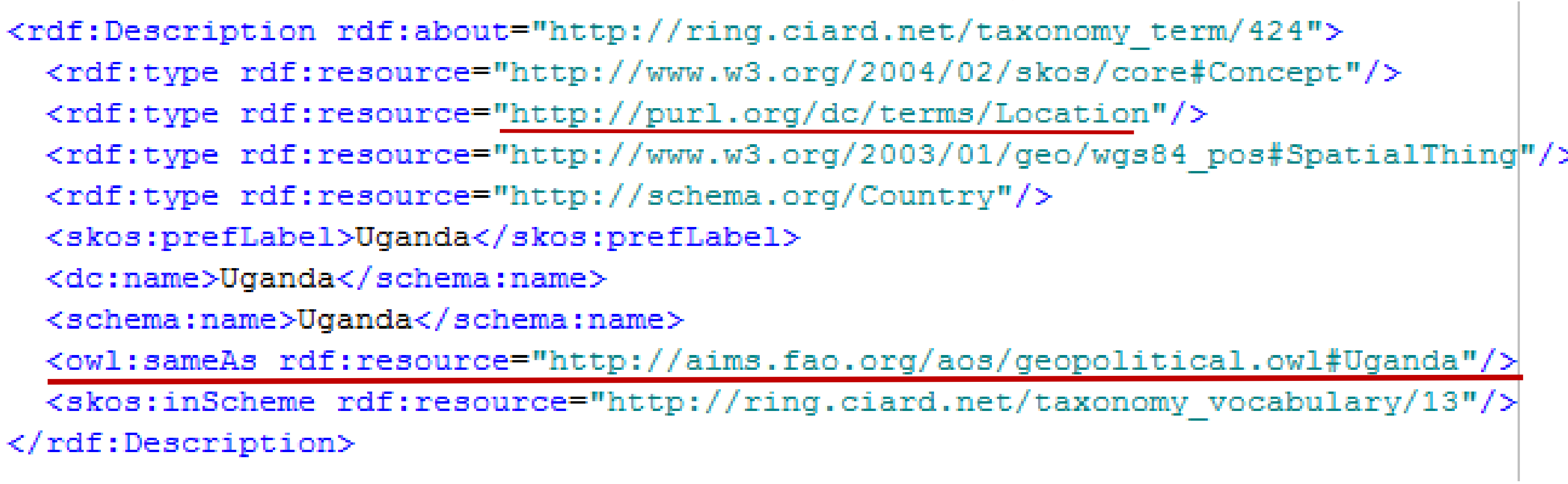
```
<dcat:keyword>women</dcat:keyword>
<dcat:keyword>empowerment</dcat:keyword>
<dct:coverage rdf:resource="http://ring.ciard.net/taxonomy_term/2272"/>
<dct:subject rdf:resource="http://ring.ciard.net/taxonomy_term/2272"/>

<rdf:Description rdf:about="http://ring.ciard.net/taxonomy_term/2272">
  <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
  <skos:prefLabel>Socio-economic data</skos:prefLabel>
  <skos:closeMatch rdf:resource="http://aims.fao.org/aos/agrovoc/c_29966">
  <skos:inScheme rdf:resource="http://ring.ciard.net/taxonomy_vocabulary/4"/>
</rdf:Description>
```

“Women” narrower of “Socio-economic data” local concept, mapped as closeMatch to URI of “socioeconomic development” concept in **AGROVOC**



```
<dc:spatial
rdf:resource="http://ring.ciard.net/taxonomy_term/424"/>
```

“Uganda” matches a local concept of type skos:Concept and dc:Location, mapped to URI of Uganda country in **FAO Geopolitical Ontology**



```
<rdf:Description rdf:about="http://ring.ciard.net/taxonomy_term/424">
  <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
  <rdf:type rdf:resource="http://purl.org/dc/terms/Location"/>
  <rdf:type rdf:resource="http://www.w3.org/2003/01/geo/wgs84_pos#SpatialThing"/>
  <rdf:type rdf:resource="http://schema.org/Country"/>
  <skos:prefLabel>Uganda</skos:prefLabel>
  <dc:name>Uganda</schema:name>
  <schema:name>Uganda</schema:name>
  <owl:sameAs rdf:resource="http://aims.fao.org/aos/geopolitical.owl#Uganda"/>
  <skos:inScheme rdf:resource="http://ring.ciard.net/taxonomy_vocabulary/13"/>
</rdf:Description>
```

Queries can leverage LOD mappings - 1

Example: To get all datasets with geographic coverage of “Uganda" using the Geopolitical Ontology URI for “Uganda"

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX dc: <http://purl.org/dc/terms/>
PREFIX dcat: <http://www.w3.org/ns/dcat#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>

DESCRIBE ?dataset ?distro WHERE {
?dataset rdf:type dcat:Dataset .
?dataset dcat:distribution ?distro .
?dataset dc:spatial ?topic .
?topic owl:sameAs <http://aims.fao.org/aos/geopolitical.owl#Uganda> .
}
```

URI of the “Uganda” in the Geopolitical Ontology

Queries can leverage LOD mappings - 2

Example: To get all datasets on topic "Livestock" using the AGROVOC URI for "Livestock"

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX dc: <http://purl.org/dc/terms/>
PREFIX dcat: <http://www.w3.org/ns/dcat#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>

DESCRIBE ?dataset ?distro WHERE {
?dataset rdf:type dcat:Dataset .
?dataset dcat:distribution ?distro .
?dataset dcat:theme ?topic .
?topic owl:sameAs <http://aims.fao.org/aos/agrovoc/c_4397> .
}
```

URI of the “Livestock” concept in the AGROVOC thesaurus

Queries can leverage LOD mappings - 3

Example: To get all datasets complying with the INSPIRE specification for soil

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX dc: <http://purl.org/dc/terms/>
PREFIX dcat: <http://www.w3.org/ns/dcat#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>

DESCRIBE ?dataset ?distro WHERE {
  ?dataset rdf:type dcat:Dataset .
  ?dataset dcat:distribution ?distro .
  ?distro dc:conformsTo ?standard .
  ?standard owl:sameAs  <http://vest.agrisemantics.org/node/19915> .
}
```

URI that identifies the
INSPIRE specification for Soil

Relevant vocabularies, catalog tools, catalogs

- DCAT: http://www.w3.org/TR/vocab-dcat/
- DCAT AP: https://joinup.ec.europa.eu/asset/dcat_application_profile/home
- STAT-DCAT: https://joinup.ec.europa.eu/asset/stat_dcat_application_profile/home
- DataCube: http://purl.org/linked-data/cube#
- VOID: http://rdfs.org/ns/void-guide
- DDI-RDF Discovery Vocabulary: http://rdf-vocabulary.ddialliance.org/discovery.html
- VIVO Datastar: http://sourceforge.net/projects/vivo/files/Datastar%20ontology/
- CERIF for datasets: https://cerif4datasets.wordpress.com/c4d-deliverables/
- CKAN: http://ckan.org/
- Dataverse: http://dataverse.org/
- Datahub: http://datahub.io/
- DataCite: http://search.datacite.org/ui?q=subject%3Aagriculture
- Re3data: http://www.re3data.org
- OpenAIRE: https://www.openaire.eu/
- CIARD RING: http://ring.ciard.info

Semantic challenges in sharing dataset metadata
and creating federated dataset catalogs.
The example of the CIARD RING

Thank you

Valeria Pesce (GFAR)
valeria.pesce@fao.org