



europaena  
think culture

# Validation of Europeana data: application profile, OWL ontology, or else?

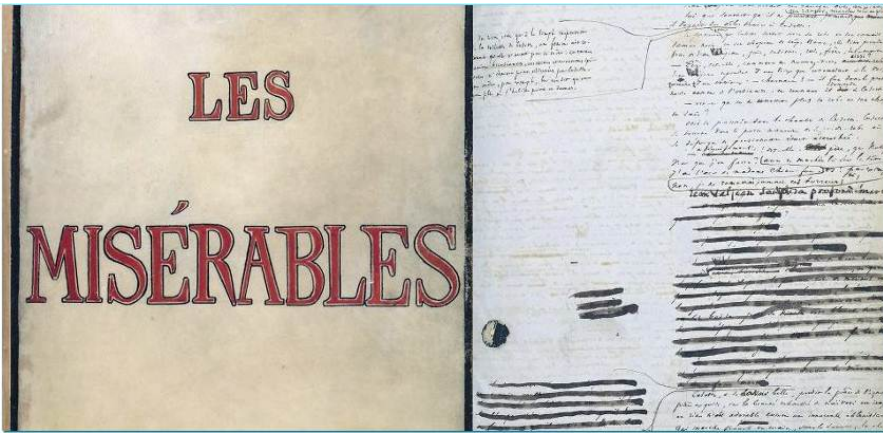
Antoine Isaac

Application Profiles as an alternative to OWL Ontologies  
Dublin Core Conference  
5 September 2013



# Europeana.eu, Europe's cultural heritage portal

Text



Video



Image



3D



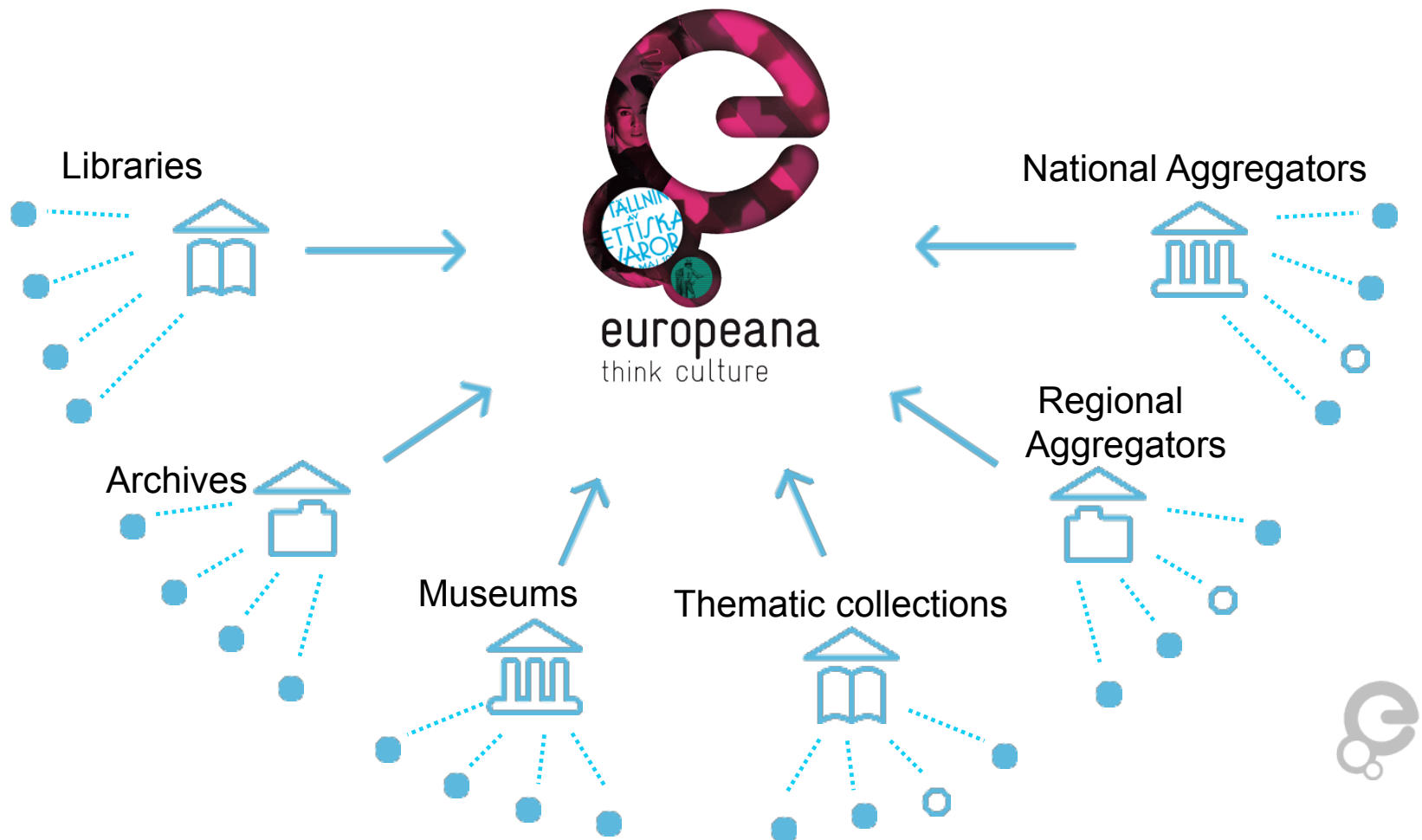
Sound





# Europeana's aggregation network

29M objects from 2,200 European galleries, museums, archives and libraries



---

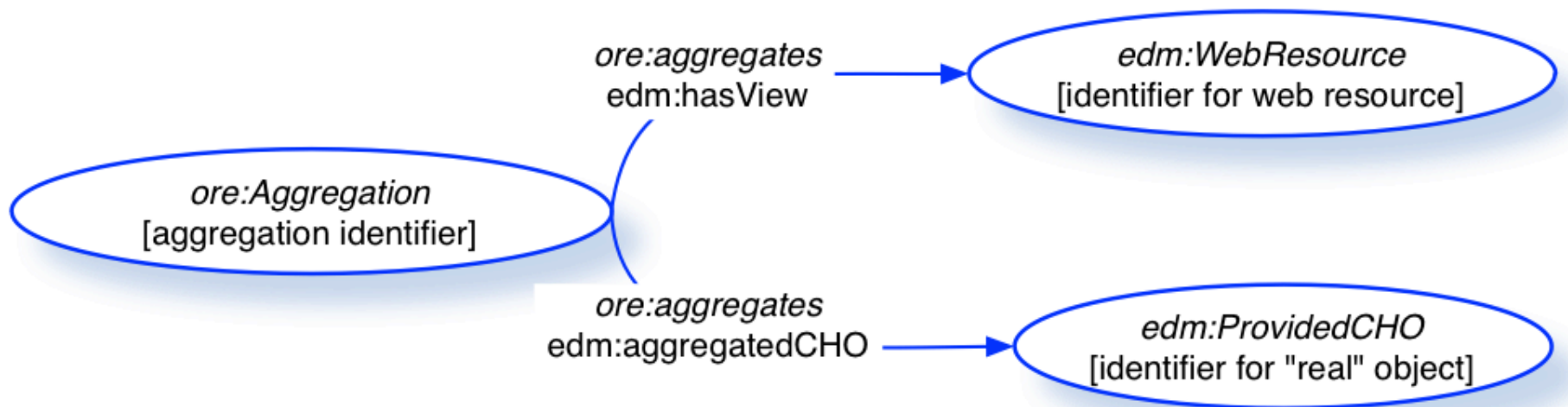
# EDM rationale

1. Distinguish “provided objects” (painting, book, movie, etc.) from their digital representations
2. Distinguish object from its metadata record
3. Allow multiple records for a same object, containing potentially contradictory statements about it
4. Support for objects that are composed of other objects
5. Support for contextual resources, including concepts from controlled vocabularies





# EDM basic pattern


- A data provider submits to Europeana a “bundle” of an object and its digital representation(s)





# Europeana Data Model: an example



 Attribution-NonCommercial-ShareAlike 

View item at  
[Cité de la musique](#) 

 Share

 Cite on Wikipedia

## Clavecin

**Description:** 2 claviers : C / E à C<sup>'''</sup>, 45 notes \* 3 rangs de Cordes : 2 x 8" + 1 x 4" \* Table en épicéa (?) \* T \* Collection Geneviève Thibault de Chambure

**Creator:** <http://www.mimo-db.eu/InstrumentMaker/Person/593>;  
Cristofori

**Coverage:** <http://sws.geonames.org/3176959/>

**Date:** fin 17e

**Type:** <http://www.mimo-db.eu/InstrumentsKeywords/2251>;  
<http://www.mimo-db.eu/HornbostelAndSachs/6461>

**Identifier:** #CM:0161930

**Data provider:** Cité de la musique

**Provider:** MIMO - Musical Instrument Museums Online

**Providing country:** MUL

[Auto-generated tags](#) ▶

### Search also for:

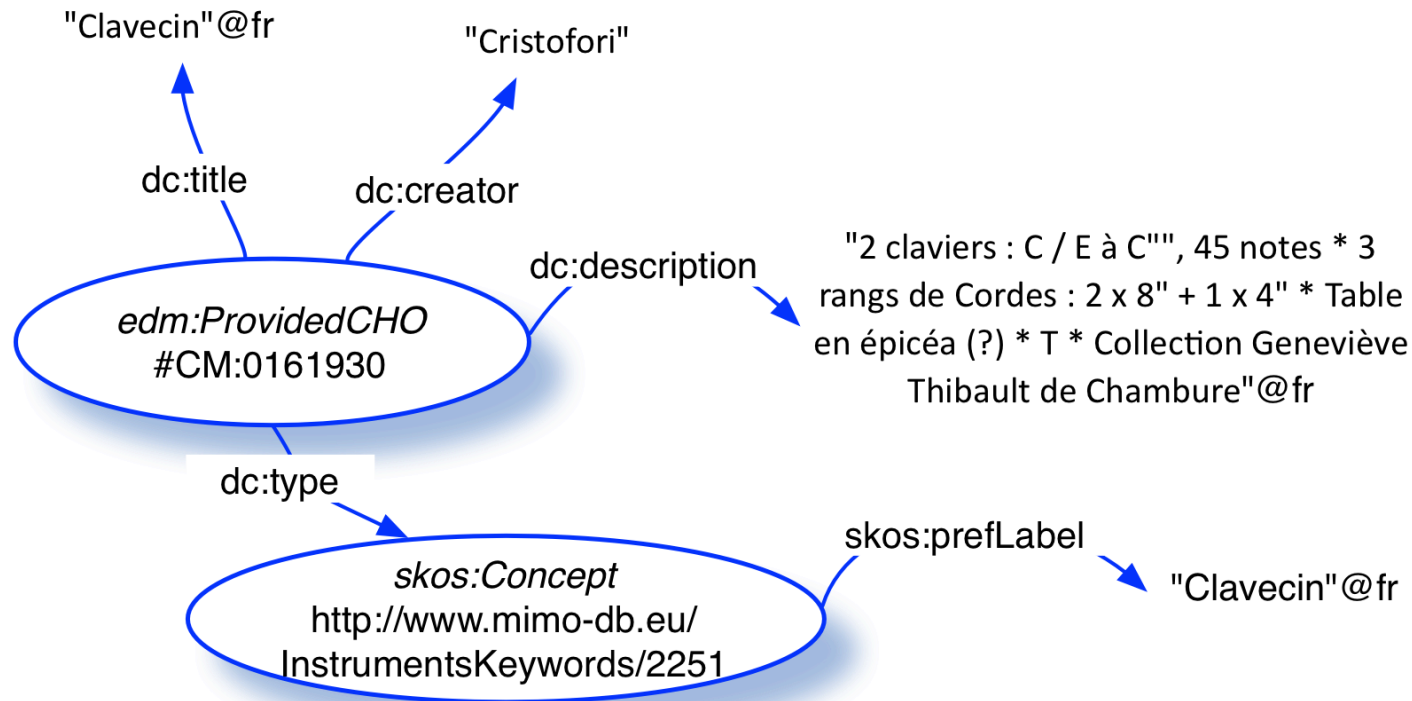
**Title**  
[Clavecin](#) (1308)

**Who**  
[Cristofori](#) (46)

**Provider**  
[Cité de la musique](#) (23075)  
[MIMO - Musical Instrument Museums Online](#) (36390)



# Provided Cultural Heritage Object (CHO) and descriptive metadata



# Web Resources – digital representations



*edm:WebResource*  
<http://www.mimo-db.eu/media/CM/IMAGE/CMIM000030829.jpg>

edm:rights

<http://creativecommons.org/licenses/by-nc-sa/3.0/>



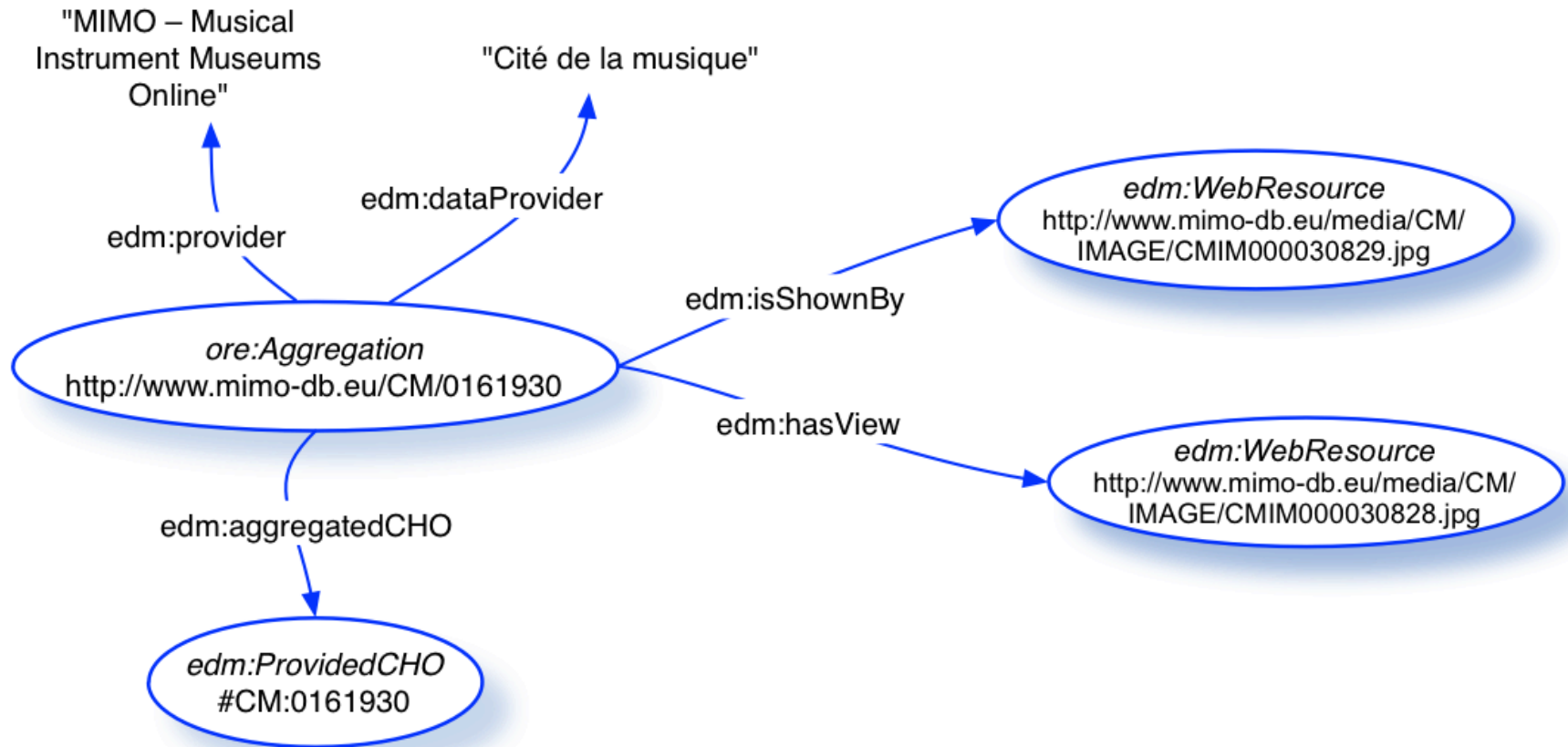
*edm:WebResource*  
<http://www.mimo-db.eu/media/CM/IMAGE/CMIM000030828.jpg>

edm:rights

<http://creativecommons.org/licenses/by-nc-sa/3.0/>



# Aggregations – Bundling it all together



---

# EDM Specs

<http://pro.europeana.eu/edm-documentation>

- EDM Definition:
- Mapping Guidelines and templates
- XML Schema
- OWL ontology



---

# EDM Definitions

High level definition of classes and properties

`edm:aggregatedCHO`

- **Definition:** This property associates an ORE aggregation with the cultural heritage object(s) (CHO for short) it is about.
- **Subproperty of:** `ore:aggregates`, `dc:subject`, `P129_is_about`
- **Domain:** `ore:Aggregation`
- **Range:** `edm:ProvidedCHO`



# EDM Definitions

→ Avoids adding semantics to re-used classes and properties

Except for mapping purposes, hierarchies of classes and properties for inference

```
dc:contributor rdfs:subPropertyOf edm:hasMet .
```

→ Borderline case of axioms not in formal version of original specs

```
ore:proxyIn
```

- Obligation & Occurrence: A proxy may be in 1 to many aggregations, and an aggregation may have 0 to many proxies in it



---

# EDM Definitions

## First hints at data constraints

`edm:dataProvider`

- **Obligation & Occurrence:** Mandatory for Europeana (Minimum: 1, Maximum: 1)

`edm:currentLocation`

- **Domain:** The set of cultural heritage objects that Europeana collects descriptions about, represented in the EDM by ProvidedCHOs and ORE proxies for these CHOs.

`edm:aggregatedCHO`

- **Obligation & Occurrence:** In Europeana, an aggregation aggregates exactly one CHO





# Data validation: Europeana requirements

EDM is RDF-oriented: unbounded web of information, etc.

But Europeana needs to enforce constraints on the data it receives

- Data that meets basic Europeana function requirements
  - An Aggregation should always have an edm:aggregatedCHO
  - There must be exactly one edm:type -- the value must be TEXT, VIDEO, SOUND, IMAGE or 3D
- Data quality criteria
  - A ProvidedCHO should have at least a dc:title or a dc:description

*We need specs for validation that are easily shareable, both for humans and machines*



# EDM Mapping Guidelines

- Document written after the EDM Definitions
- Tries to formulate clearer instructions for Europeana providers
- Template-based, e.g. for provider's Aggregation:

<i>property</i>	<i>value type</i>	<i>cardinality</i>
<b>edm:aggregatedCHO</b>	reference (of an item)	min 1, max 1
<b>edm:dataProvider</b>	literal or reference	min 1, max 1
<b>edm:isShownAt</b>	reference	min 0, max 1 -- Either isShownBy OR isShownAt is Mandatory
<b>edm:isShownBy</b>	reference	min 0, max 1 -- Either isShownBy OR isShownAt is Mandatory
<b>edm:object</b>	reference	min 0, max 1
<b>edm:provider</b>	literal or reference	min 1, max 1
<b>dc:rights</b>	reference or literal	min 0, max unbounded
<b>edm:rights</b>	reference	min 1, max 1

# Machine-readable specs as OWL ontologies?

OWL is good for writing constraints, but not for validation!

Quite OK

- “Value types” via owl:ObjectProperty owl:DatatypeProperty in OWL(DL)
- Data ranges (TEXT-VIDEO-SOUND-IMAGE-3D)

Less ok:

- Object domain and ranges
- (qualified) cardinality axioms
  - Including combinations: (either isShownBy OR isShownAt is mandatory)



# OWL?

We've created an OWL version of EDM

<https://github.com/europeana/corelib/blob/master/corelib-solr-definitions/src/main/resources/eu/rdf/>

[...]

```
<owl:equivalentClass>
```

```
  <owl:Restriction>
```

```
    <owl:cardinality rdf:datatype="&xsd;nonNegativeInteger">1</owl:cardinality>
```

```
    <owl:onProperty rdf:resource="&edm;aggregatedCHO"/>
```

```
  </owl:Restriction>
```

```
</owl:equivalentClass>
```

[...]

*But these are not really validation axioms*

*And it's bad practice to add semantics to classes and properties that already exist, such as ore:Aggregation*

*(let's be honest: we were not ready for full RDF/OWL compatibility anyway...)*



# XML Schema

EDM is implemented by an XML Schema (for RDF data!)

[...]

```
<sequence>
```

```
<element ref="edm:aggregatedCHO" maxOccurs="1" minOccurs="1"/>
```

```
<element ref="edm:dataProvider" maxOccurs="1" minOccurs="1"/>
```

```
<element ref="edm:isShownAt" maxOccurs="1" minOccurs="0"/>
```

```
<element ref="edm:isShownBy" maxOccurs="1" minOccurs="0"/>
```

```
<element ref="edm:object" maxOccurs="1" minOccurs="0"/>
```

```
<element ref="edm:provider" maxOccurs="1" minOccurs="1"/>
```

```
<element ref="dc:rights" maxOccurs="unbounded" minOccurs="0"/>
```

```
<element ref="edm:rights" maxOccurs="1" minOccurs="1"/>
```

```
</sequence>
```

[...]





---

# XML Schema

And Schematron rules:

[...]

```
<sch:pattern name="Either Is shownby or is shownat should be present">  
  <sch:rule context="ore:Aggregation">  
    <sch:assert test="edm:isShownAt or edm:isShownBy">  
      [Error message]  
    </sch:assert>  
  </sch:rule>  
</sch:pattern>
```



---

# XML Schema: not ideal!

- Specific to a syntax
- Document-centric approach to validation  
Back to square one: records!
- Forces us to enumerate the attributes with extra constraints, especially order of elements  
It's really a super-closed world
- Schematron does slightly better, but then we have two constraint languages co-existing in a same implementation



---

# Even worse

- We have several contexts for validating EDM data
- “internal” schema vs. provider schema
  
- XML Schema-level constraints cause a lot of duplication for declarations
- No easy layering of different constraint sets



---

# EDM as a “real” application profile?

- It is an application profile, already: mixing several vocabularies, adding specific constraints
- Documentation includes definitions with constraints and examples
- Interpretation of constraint in APs fit quite well
  - AP constraints are expressed on the data
  - Europeana needs dataset-level validation, mostly



# EDM as a real application profile?

A fragment in DSP XML

<http://dublincore.org/documents/dc-dsp/>

```
<DescriptionTemplate ID="aggregation" standalone="yes">
```

```
  <ResourceClass>ore:Aggregation</ResourceClass>
```

```
  <StatementTemplate minOccurs="1" maxOccurs="1">
    <Property>edm:aggregatedCHO</Property>
  </StatementTemplate>
```

```
  <StatementTemplate minOccurs="1">
    <Property>edm:isShownBy</Property>
    <Property>edm:isShownBy</Property>
  </StatementTemplate>
```

```
</DescriptionTemplate>
```





# Could be converted to other constraint checking formalisms (1/2): SPIN

SPARQL Inferencing Notation

<http://spinrdf.org>

```
ore:Aggregation
  spin:constraint
    [ a sp:Ask ;
      sp:text ""
        # either isShownBy or isShownAt must be present
        ASK WHERE {
          {?this isShownBy ?image } UNION {?this isShownBy ?page }
        }""
    ].
```

*Issue: still looks like adding semantics to ore:Aggregation in general...*



# Could be converted to other constraint checking formalisms (1/2): Stardog ICV

Integrity Constraint Validation

<http://stardog.com/docs/sdp/>

Class: ore:Aggregation

SubClassOf: exactly 1 edm:aggregatedCHO

Class: ore:Aggregation

SubClassOf: min 1 edm:isShownBy or min 1 edm:isShownAt

*Note: this is OWL2's 'Manchester Syntax'*

*Stardog accepts OWL, SWRL and SPARQL, uses SPARQL as back-end*

*Issue: still looks like adding semantics to ore:Aggregation in general...*



---

# Conclusions

Europeana requirements seem to be met by the AP approach, if this AP approach is matched with SPARQL constraints

Much better than to try to partially catch constraints in OWL and XML +Schematron as isolated machine-readable specs

Needs further testing

incl. trying to express all constraints with DSP and SPARQL queries (with or without the help of a higher-level language)

An area that needs maturation

Maintainers (like us) may have validation specs in various forms

DSP is not flying; RDF validation still worthy of workshops!



<https://www.w3.org/2012/12/rdf-val/>



europæana  
think culture

**Thank you!**

[aisaac@few.vu.nl](mailto:aisaac@few.vu.nl)

