

---

**europeana**  
think culture

---

# Europeana Data & Interoperability Issues

Antoine Isaac

Using slides from Valentine Charles, Wibke Kolbmann

And work of Operations team: Jan Molendijk, Susanna Summa, Robina Clayphan, Alicia Ackerman, Ewa Glowacz

---

# The problem

---

- Aggregating data from many, very different providers (sectors, domains)
  - Each with their metadata tradition
    - Centuries!
  - Many have very limited resources
-

# Europeana's AP

---

- Europeana Semantic Elements (ESE)  
<http://version1.europeana.eu/web/guest/technical-requirements/>
  - Based on Dublin Core
  - With some adhoc fields
- 
-

---

# Descriptive metadata

---

**dc:subject**

**dc:creator**



**dc:title**

---

---

---

# Supporting Europeana's specific functions

---

europeana:object

europeana:type



europeana:isShownAt

---

---

Source	Element	Element Refinement(s)
DC	<a href="#">title</a>	<a href="#">alternative</a>
DC	<a href="#">creator</a>	
DC	<a href="#">subject</a>	
DC	<a href="#">description</a>	<a href="#">tableOfContents</a>
DC	<a href="#">publisher</a>	
DC	<a href="#">contributor</a>	
DC	<a href="#">date</a>	<a href="#">created</a> ; <a href="#">issued</a>
DC	<a href="#">type</a> *	
DC	<a href="#">format</a>	<a href="#">extent</a> ; <a href="#">medium</a>
DC	<a href="#">identifier</a>	
DC	<a href="#">source</a>	
DC	<a href="#">language</a> *	
DC	<a href="#">relation</a>	<a href="#">isVersionOf</a> ; <a href="#">hasVersion</a> ; <a href="#">isReplacedBy</a> ; <a href="#">replaces</a> ; <a href="#">isRequiredBy</a> ; <a href="#">requires</a> ; <a href="#">isPartOf</a> ; <a href="#">hasPart</a> ; <a href="#">isReferencedBy</a> ; <a href="#">references</a> ; <a href="#">isFormatOf</a> ; <a href="#">hasFormat</a> ; <a href="#">conformsTo</a>
Europeana		<a href="#">isShownBy</a> ; <a href="#">isShownAt</a>
DC	<a href="#">coverage</a>	<a href="#">spatial</a> ; <a href="#">temporal</a>
DC	<a href="#">rights</a> *	
DC terms	<a href="#">provenance</a>	
Europeana	<a href="#">country</a>	
Europeana	<a href="#">dataProvider</a>	
Europeana	<a href="#">hasObject</a>	
Europeana	<a href="#">language</a> *	
Europeana	<a href="#">object</a>	
Europeana	<a href="#">provider</a>	
Europeana	<a href="#">rights</a> *	
Europeana	<a href="#">type</a> *	
Europeana	<a href="#">unstored</a>	
Europeana	<a href="#">uri</a>	
Europeana	<a href="#">userTag</a>	
Europeana	<a href="#">year</a>	

# Occurrence recommendations

## Strongly recommended

dc:title  
dcterms:alternative  
dc:creator  
dc:contributor  
dc:date  
dcterms:created  
dcterms:issued

## Recommended

dc:coverage  
dcterms:spatial  
dcterms:temporal  
dc:description  
dcterms:isPartOf  
dc:language  
dc:publisher  
dc:source  
dc:subject  
dc:type<sup>7</sup>

## Additional elements

dc:format  
dcterms:extent  
dcterms:medium  
dc:identifier  
dc:rights  
dcterms:provenance  
dc:relation  
dcterms:conformsTo  
dcterms:hasFormat  
dcterms:isFormatOf  
dcterms:hasVersion  
dcterms:isVersionOf  
dcterms:hasPart  
dcterms:isReferencedBy  
dcterms:references  
dcterms:isReplacedBy  
dcterms:replaces  
dcterms:isRequiredBy  
dcterms:requires  
dcterms:tableOfContents

## Europeana elements

europa:country  
europa:dataProvider  
europa:hasObject  
europa:isShownAt  
europa:isShownBy  
europa:language  
europa:object  
europa:provider  
europa:rights<sup>7</sup>  
europa:type<sup>1</sup>  
europa:unstored  
europa:uri  
europa:usertag  
europa:year

# Some control in Europeana fields

Element	Who is responsible
europeana:dataProvider	You should provide this element when it is possible to do so..
europeana:isShownAt or europeana:isShownBy	You <b>must</b> provide <b>at least one</b> of these elements.
europeana:object	You should provide this element if appropriate to your data
europeana:provider	You <b>must</b> provide this element.
europeana:rights	You should provide this element when it is possible to do so.
europeana:type	You <b>must</b> provide this element
Europeana:unstored	You can provide this element if appropriate to your data
europeana:country	Europeana is responsible for providing all these elements.
europeana:hasObject	
europeana:language	
europeana:uri	
europeana:usertag	
europeana:year	

- Occurrence
- Allowed values

# Problems specific to the simplicity and (non-)flexibility of the AP

- Ambiguity of fields
  - Events and roles
  - Techniques and materials related to the object
    - *FilmManifestation/Duration* - **dcterms:extent**  
(1 min, 4 min, 1 min 30 sec, 2 min, 3 min, 4 min 8 sec)
    - *FilmManifestation/CarrierAspect* - **dc:format**  
(1:1,33, 16:9 86 1:1,37, 1:2,35)
    - *FilmManifestation/CarrierFormat* – **dc:format**  
(35 mm, DVD, 16 mm, VHS)
    - *FilmManifestation/Colour* – **dc:format**  
(Black and White, Colour, tinted)
    - *FilmCopy/CarrierType* - **dc:format**  
(Positive, Digital file, n/a, Video tape, Duplicate negative, Reversal positive, Negative)
    - *FilmCopy/CarrierMaterial* - **dcterms:medium**  
(Acetate, Polyester, Nitrate)

# Problems specific to the simplicity and (non-)flexibility of the AP

---

- Ambiguity of fields
  - Semantic overload of elements
    - Tweaking mapping to fit Europeana display for hierarchical objects
- 
-

# Problems specific to the simplicity and (non-)flexibility of the AP

---

- Ambiguity of fields
  - Semantic overload of elements
  - Violation of one-to-one principle: multiple resources described in one record
    - Mix between digital data and original object data
- 
-

---

No	Value	Original field name	Solution adopted
1	Title Neolithic arrowhead from Southern Iraq	Name	dc:title
2	Unknown	Actor	dc:creator
3	Neolithic	Period	dc:date
4	image/jpeg	Format	dc:format
5	1.5 cm X 3cm	Dimension	dcterms:extent
6	flint	Material	dcterms:medium
7	Iraq	Origin	<b>dc:subject</b>
8	London; BM	Current Location	<b>europena:unstored</b>

---

# Problems specific to the simplicity and (non-)flexibility of the AP

---

- Ambiguity of fields
  - Semantic overload of elements
  - Violation of one-to-one principle: multiple resources described in one record
  - Lack of control for values
    - Especially harmful in cross-domain multilingual environment
- 
-

# Value issues

---

- AP uses simple string values
    - No vocabulary encoding scheme or syntax encoding scheme
  - No handling of elements from controlled vocabularies
    - Notations difficult to exploit
      - 1.712 (SHIC)
    - Cannot exploit synonyms, etc.
  - No handling of complex values
    - Dealing with coordination of concepts  
<dc:subject>Maria Nugent, Journal, Diary, Jamaica<dc:subject>
    - Multiple subjects or coordinated ones?
- 
- No standard syntax for dates and names

# Lack of flexibility & low granularity of ingestion format

---

- Some original data is lost
- 
-

# Original record

## ▼ Dates

? LATE POST MEDIEVAL (Certain), Post 1649 AD - Ante 1660 AD

## ▼ Further information

? Method of Manufacture:	Struck or hammered	? Length:	not defined	? Quantity:	1
? Material:	Silver	? Width:	not defined	? Preservation:	
Decoration:		? Thickness:	0.35 mm	? Completeness:	Complete
? Surface Treatment:		? Diameter:	12.9 mm		
? Weight:	0.42 grams	Evidence of reuse:			

## ▼ Discovery Information

? Found by: UNKNOWN OR ACCESS RESTRICTED, 16.04.2006  
? Method of Discovery: Metal detector

## ▼ Additional Information

? Primary Identifier: UNKNOWN OR ACCESS RESTRICTED  
? Secondary Identifier: not defined  
? Recorded by: UNKNOWN OR ACCESS RESTRICTED  
? Subsequent Action: Returned to finder.

## ▼ Coin-specific information

Ruler: Commonwealth (Post medieval: 1649-1660 )  
Ruler Qualifier: Certain  
Denomination: penny  
Denomination Qualifier: Certain  
Obverse Description: Shield of St George within wreath  
Reverse Description: Co-joined shield of St George and Ireland, mark of value 1  
Die Axis Measurement: 4 o'clock

# Delivered by the aggregator to Europeana



**Title:** COIN

**Date:** 2006-08-15

**Creator:** Clifford, Trista - Portable Antiquities Scheme

**Description:** A silver commonwealth penny, 1649-1660 AD.

**Language:** en-GB

**Source:** Portable Antiquities

**Rights:** Copyright 2005 Portable Antiquities Scheme

**Provider:** CultureGrid ; Uk

[Less](#)

**Identifier:** [http://www.findsdatabase.org.uk/hms/pas\\_obj.php?type=finds&id=00144E1F8CC0156F](http://www.findsdatabase.org.uk/hms/pas_obj.php?type=finds&id=00144E1F8CC0156F)

**Type:** Image

**Relation:** Portable Antiquities

# Data quality improvement: which approach to choose?

---

## First level – Data Provider

- Basic errors even for their own standards/norms

## Second level – Aggregators/projects

- First standardization/harmonization of data of one community

## Third level – Metadata enrichment by Europeana

- Requires highly standardized and consistent data
  - Will augment existing data, not replace it
-

# Data quality improvement: which approach to choose?

---

Mostly a matter of policy setting, agreement and hard work from stakeholders

- What is wished for / possible at any given level

Can tools help?

- Perhaps for data normalization, but will be quite adhoc “recipes” specific to one domain, or even one collection
  - Better mapping functions and tools
-

# Data quality improvement streams

---

- Use and occurrence of metadata elements
  - Consistency and standardization of data values
  - Richness and flexibility for ingestion format
- 
-

# Standardization of formats

---

- For dates and names, technical data
  - Use of ISO norms?
    - E.g., ISO 8601 for dates
      - 9th August 2005 becomes 2005-08-09
      - 16th February 1331 to 4th May 1406 becomes 1331-02-16/1406-05-04
-

# Adding mandatory occurrence rules

---

- Priority is to populate fields
    - Easier / more important to have data rather than no data
  - rights info + (institutional) provenance
  - One of dc:subject, dc:type, dc:spatial, dc:coverage
  - dc:title or dc:description
  - dc:language (controlled)
-

# Working on a richer data model

---

- Europeana Data Model (EDM)

<http://group.europeana.eu/web/europeana-project/technicaldocuments/>

---

---

# EDM requirements & principles

---

1. Distinction between “provided object” (painting, book, program) and digital representation
  2. Distinction between object and metadata record describing an object
  3. Allow for multiple records for same object, containing potentially contradictory statements about an object
  4. Support for objects that are composed of other objects
  5. Standard metadata format that can be specialized
  6. Standard vocabulary format that can be specialized
  7. EDM should be based on existing standards
-

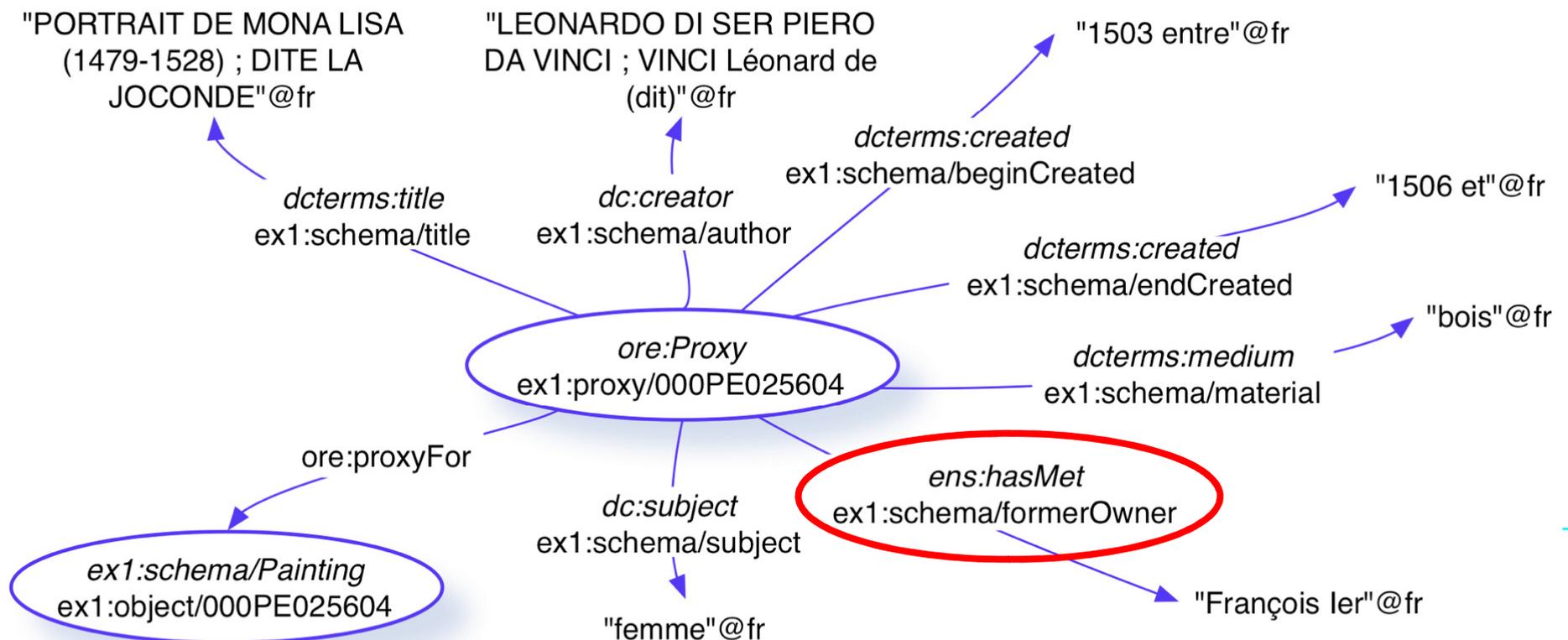
# EDM basics

---

- OAI ORE for organization of metadata about an object
  - Dublin Core for descriptive metadata representation
  - SKOS for vocabulary representation
- 
-

# A flexible model: different semantic grains

- Keep data expressed as close as possible to original model
- Using mappings to more interoperable level



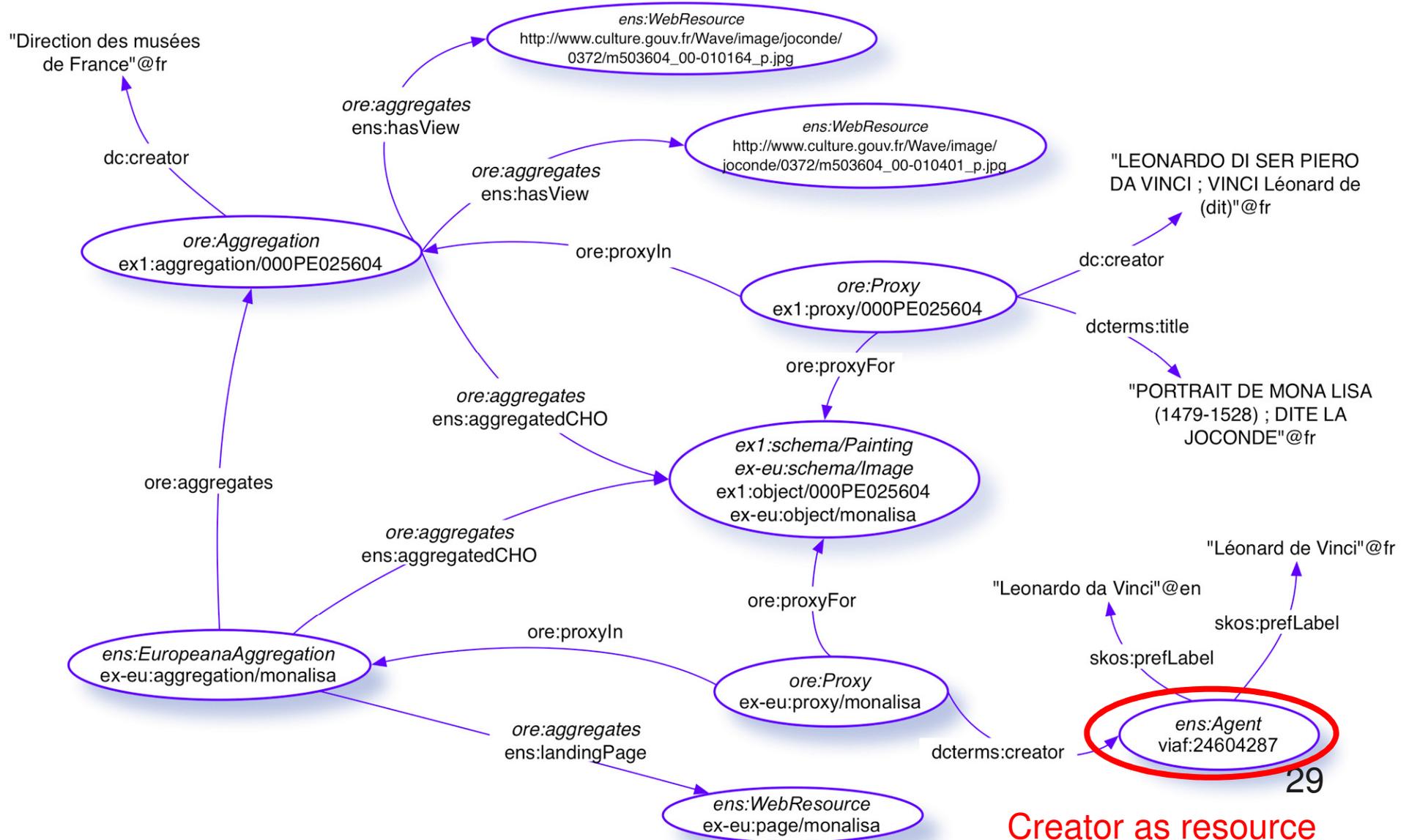
# Advanced modeling in EDM

---

- Relations between provided objects
    - Part-whole links for complex (hierarchical) objects
    - Derivation and versioning relations
  - Relations to contextual entities: events, persons, places...
- 
-



# Representation of contextual entities as resources





**Thanks!**

---

---

