Provenance e Paradata negli *OpenData*: il modello 3D-ICONS

*Andrea D’ Andrea*
Centro Interdipartimentale di Servizi di Archeologia
Università degli Studi di Napoli “L’Orientale”
dandrea@unior.it

This project has been funded with support from the European Commission’s CIP ICT PSP programme
OPEN DATA

➢ ... is a radical change of mentality encouraged by the need to share data;
➢ ... encloses several meanings (classifications, applications) characterized by limitation of ©;

But ...

we have not guidelines. There is an ideological approach which enhances the free spirit of the resource rather than his real reuse or usability in other applications. We need OPEN-METADATA to improve the quality of our data
THE 3D-ICONS PROJECT

➢ 3D-ICONS will bring exciting and engaging 3D models and constructions to Europeana end users
   – Challenge for Cultural Heritage sector to bring their 3D models used for scientific & research purposes into the public domain

● CISA-UNO Coordinator (2012-2015)
   ● 16 partners from 11 different countries
   ● Budget: 5M€ for 3 years
WHAT IS EUROPEANA?

➢ A huge database of information about digital content (metadata) held by European cultural heritage institutions
  - Funded by the European Commission
  - Supplied by a range of Best Practice and other Pilot projects as well as directly by (larger) institutions

➢ Catalyst for new content services
  - Europeana services and tools, standards, Linked Open Data, Hackathons – new content-based apps.
CURRENT STATUS OF EUROPEANA

- March 2013
  - over 26 million objects
  - from more than 2,200 institutions
  - from 34 countries.

Below is a list of the top 15 countries providing data to Europeana.eu:

<table>
<thead>
<tr>
<th>Country</th>
<th>Total</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total of all records</td>
<td>26,787,078</td>
<td></td>
</tr>
<tr>
<td>Germany</td>
<td>4,175,834</td>
<td>15.6%</td>
</tr>
<tr>
<td>France</td>
<td>2,728,849</td>
<td>10.2%</td>
</tr>
<tr>
<td>Spain</td>
<td>2,525,987</td>
<td>9.4%</td>
</tr>
<tr>
<td>Netherlands</td>
<td>2,519,664</td>
<td>9.4%</td>
</tr>
<tr>
<td>Sweden</td>
<td>2,128,580</td>
<td>7.9%</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>1,825,601</td>
<td>6.8%</td>
</tr>
<tr>
<td>Norway</td>
<td>1,563,843</td>
<td>5.8%</td>
</tr>
<tr>
<td>Poland</td>
<td>1,457,931</td>
<td>5.4%</td>
</tr>
<tr>
<td>Italy</td>
<td>1,446,758</td>
<td>5.4%</td>
</tr>
<tr>
<td>Europe</td>
<td>1,315,071</td>
<td>4.9%</td>
</tr>
<tr>
<td>Ireland</td>
<td>983,593</td>
<td>3.7%</td>
</tr>
<tr>
<td>Finland</td>
<td>698,599</td>
<td>2.6%</td>
</tr>
<tr>
<td>Denmark</td>
<td>605,012</td>
<td>2.3%</td>
</tr>
<tr>
<td>Austria</td>
<td>501,914</td>
<td>1.9%</td>
</tr>
<tr>
<td>Belgium</td>
<td>424,472</td>
<td>1.6%</td>
</tr>
</tbody>
</table>

Figures are correct as of 18 March 2013.
The Europeana Data Exchange Agreement (DEA) is the central element of the Europeana Licensing Framework. The DEA structures the relationship between Europeana and its data providers.

The DEA sets out two simple principles:

- For all descriptive metadata provided to Europeana, data providers grant Europeana the right to publish the metadata under the terms of the Creative Commons CC0 1.0 Universal Public Domain Dedication. This means that all metadata provided to Europeana can be re-used by third parties without any restrictions.

- Each digital object (and the associated preview) that is available via Europeana needs to carry a rights label that describes its copyright status. Data providers grant Europeana the right to publish previews provided to Europeana. Previews may not be re-used by third parties unless the rights label related to the object allows such re-use.
INNOVATIVE ASPECTS

➢ 3D-ICONS will use CARARE aggregation service
  – ORE OAI aggregation
  – Linked Open Data

➢ Production pipeline
  – Flexibility in data creation
  – Business model
    • Cost analysis, IPR, organizational models
  – Advanced semantics/rich metadata
  – Quality assurance
“Provenance of a resource is a record that describes entities and processes involved in producing and delivering or otherwise influencing that resource”

The systematic large-scale production of digital scientific objects, the diversity of the processes involved and the complexity of describing historical relationships among them, requires the need for an innovative knowledge management system.

- To capture and handle all the semantic information linked to the creation, management and documentation of the origins and derivation of digital resources, recently CRM\textsubscript{dig} has been implemented as an extension of the event-model CIDOC-CRM ontology. Scope of the CRM\textsubscript{dig} is to describe all the processes forming a connected graph through the data and things involved in multiple events.
PARADATA

Are:

➤ information about human processes of understanding and interpretation of data objects.
➤ descriptions stored within a structured dataset of how evidence was used to interpret an artefact, or a comment on methodological premises within a research publication.
➤ closely related, but somewhat different in emphasis, to “contextual metadata”, which tend to communicate interpretations of an artefact or collection.
While *Provenance* is a record about *technical process*, *Paradata* is a record about the *human process*.

We represent the process through data, people, places, time, objects involved in multiple events forming a chain of events.

We make explicit the methodological premises and the research targets referred to the digitization process (alternative hypothesis and factual evidences to support the motivations and the reasonings at the basis of the implementation of 3D content).
3D-ICONS SCHEMA (CARARE 2.0

- Is an updating of the CARARE 1.0
- Reuse CRMdig for provenance
- Reuse CIDOC-CRM for paradata
- Is mapped on EDM (Europeana Data Model)
- CIDOC-CRM is aligned to EDM
CARARE SCHEMA CORE

- Actors
- Activity
- Place
- Digital resource
- Heritage asset
- Collection
- CARARE object

Relationships:
- Actors was present at Place
- Digital resource is representation of Heritage asset
- Heritage asset is part of Collection
- Collection is part of CARARE object

Provenance
Paradata
CARARE 2.0
**Activity.** This is information about the events or activities that the monument has taken part in.

**Methods:** the methods used in this specific activity, e.g. open area excavation, sample survey, augering, boring, stratigraphic, restoration, conservation, re-pointing, photogrammetric survey etc. Use of a controlled vocabulary is recommended. *(PROVENANCE)*

**Event type:** general classification of the type of event or activity which took place, e.g. survey, archaeological excavation, digitization, rebuilding. Use of a controlled vocabulary is recommended. *(PROVENANCE)*
Activity ...

**Had General Purpose:** this is a free text description of the general goal or purpose of an Activity. For example this could include practicing, preparing, monitoring, researching, designing, testing etc. *(PARADATA)*

**Consists of:** this is a repeating group of elements which allows the specific activity (or activities) that took place during the overall Event to be described.

**Had Specific Purpose:** a free text note describing the specific goal or purpose of this activity. For example, carrying out 3D data acquisition, restoration of a part of a building, completing a survey, constructing a building, etc. *(PARADATA)*
To define the digitization process adopted to create the 3D final model we added some new relations explaining the relations between the Heritage Asset, Digital Resource and the Activity. The properties reuse those from CIDOC-CRM and in particular from CRM-Dig:

L1B.was_digitized_by – associates a Heritage Asset with an Activity. It is a subproperty of Was Present At.
L20F.has_created – associates an Activity to an Digital Resource. It is a subproperty of Was Present At.
L22F.created_derivative – defines the reuse of a Digital Resource, created by a digitization process, through multiple different processing phases. It is a subproperty of Is Derivative Of.

MAPPING SCHEMA

CARARE > EDM

EDM Event Approach
3D-ICONS ingestion and Service Architecture

Provider 1
Provider 2
Provider N

Mapping to CARARE 2.0 SCHEMA

3D-ICONS Repository

MINT2

MORE2

Europeana

3D ICONS Portal ?

CONCLUSIONS

- CARARE 2.0 will foster increasingly the European Institutions to adopt an clearer approach describing the features of the cultural object, the techniques and the methodologies chosen for the digitization and the motivation at the basis of the creation of the digital object.

- The complete knowledge of the digital resource will allow a more efficient reuse of the archive improving the usability of the resources available on-line. Furthermore it will be more easy to compare models, their complexity, any eventual innovation in their creation and their reliability.
The future of OpenData will be influenced by the quality of the record (and its metadata). Provenance and Paradata can contribute to overcome these obstacles by proposing a form of structuring metadata for digital 3D models.

LinkedOpenData technology extends the capabilities of the OpenData through new WEB formats which will integrate the archives in an infinite interconnections.

All Europeana datasets can be explored and queried through an experimental SPARQL endpoint provided by Ontotext at europeana.ontotext.com.
Thank you
dandrea@unior.it

http://3dicons-project.eu