Cross organization data harmonization and access, linked data decentralized approach

Lec Maj
Yale University
Emmanuelle Delmas-Glass
Yale University
Joshan Mahmud
British Museum
Yale Center for British Art Motivation

• Supplement current Yale collection catalog
• Combine British Museum and Yale Center for British Art data
• Larger goal: global access to British Art collections
• Allow each institution to manage their own data
• Expose data as it is cataloged
• Better data:
  ➢ Harmonize vocabularies
  ➢ Utilize ontologies

“Cross organization data harmonization and access, linked data decentralized approach” Maj, Mahmud, Delmas-Glass – CIDOC 2012
Yale Previous Data Aggregation

YALE ODAI
Yale Office of Digital Assets and Infrastructure
http://odai.yale.edu/

OAI-PMH
Open Archives Initiative Protocol for Metadata Harvesting
http://www.openarchives.org/pmh/

LIDO
Lightweight Information Describing Objects (used by YCBA Art Collections)
http://www.lido-schema.org/schema/v1.0/lido-v1.0-schema-listing.html

Apache SOLR
Open-source server based on the Lucene Java search library
http://lucene.apache.org/solr/

VuFind
Resource portal for browsing and searching collections via web
http://vufind.org/

Website
VuFind Collection Search

Plugins (HTML, Javascript, PHP)

Custom SOLR based tools

“Cross organization data harmonization and access, linked data decentralized approach” Maj, Mahmud, Delmas-Glass – CIDOC 2012
Previous Data Exchange Architecture

Application Layer
- Art Project (Web)
- FloorPlans (Maps for Android)
- Goggles
- Google+ Social

Data Aggregation and Harmonization Layer
- Artstor
- ArtPaintings
- Google

Data Layer
- Data Provider
- Collection Data
- LIDO XML via OAI-PMH

VuFind
- Plugins (HTML, Javascript, PHP)
- Search
- Maps
- FloorPlans

Custom SOLR based tools

Social and Mobile

Proprietary API

Proprietary API

Proprietary API

Proprietary API

Proprietary API

Proprietary API

Proprietary API

Maj, Mahmud, Delmas-Glass – CIDOC 2012

“Cross organization data harmonization and access, linked data decentralized approach”
Architecture problem

• Aggregation is time consuming and could be error prone
• Yale units have to come together on harmonization / mapping decisions
• Not all data is indexed, however there is a field in the index that contains the full LIDO XML
• Only large institutions have the resources to harvest and maintain aggregations

“Cross organization data harmonization and access, linked data decentralized approach” Maj, Mahmud, Delmas-Glass – CIDOC 2012
British Museum Motivation

• Collection Online was developed and allowed public access to collection through internet
• Did not address the requirements of researchers, academics & cultural heritage technologists
• Semantic Web & RDF presented a format on the internet to give those people more power to do what they want
British Museum First Iteration

• Was on full collection
• Performed by Seme4 (University of Southampton)
• Used CIDOC CRM
• Worked in conjunction with Martin Doerr
• Issues:
  – No inferencing
  – Too many extensions
  – Did not take other institutions into account

“Cross organization data harmonization and access, linked data decentralized approach” Maj, Mahmud, Delmas-Glass – CIDOC 2012
British Museum Second Iteration

• Performed in conjunction with YCBA, ResearchSpace (Rembrandt data)
• Allowed better modelling and harmonisation
• Good knowledge exchange & technology exchange

“Cross organization data harmonization and access, linked data decentralized approach” Maj, Mahmud, Delmas-Glass – CIDOC 2012
**Architecture and Process**

**Generic Overview**

- **Collection Management System** – where you keep your collection data, could be several systems that manage your collections
- **XML export** – scripts, SQL, native XML exporter build into the collection system
- **XML Data Files** – files generated by XML exporter from collection system
- **RDFer** – British Museum converter of XML data files into RDF data files utilizing configuration file to map collection data into recommended CRM standard representation.
- **RDF XML Data Files** – files generated by RDFer that will be inserted into a Triple Store
- **Triple Store** – is a purpose-built database for storage and retrieval of triples, a triple being a data entity composed of subject-predicate-object, like: “*Painting was painted by Constable*”
- **SPARQL** – RDF Query Language for Triple Store databases; v1.1 support for Federation
Museum Specific Examples
Architecture and Process

Yale Center for British Art (YCBA)

- Gallery Systems TMS
- LIDO XML Data Files
- RDF XML Data Files
- OWLIM SE Triple Store

British Museum (BM)

- System Simulation LTD Index+
- Index+ XML Data Files
- RDF XML Data Files
- OWLIM SE Triple Store

- **COBOAT** – CogApp application sponsored by Online Computer Library Center (OCLC)
- **Batch Scripts** – Tools specific to British Museum export of data from Index+ to XML data files.
- **LIDO** – Lightweight Information Describing Object standard
- **RDF** – Resource Description Framework XML files utilizing various ontologies (CRM, SKOS, etc.)
XML + CRM rules → RDFer → RDF

“Cross organization data harmonization and access, linked data decentralized approach” Maj, Mahmud, Delmas-Glass – CIDOC 2012
Proposed Data Exchange Architecture

Device Layer

Possible Partner Application Layer

Federated Virtualization Layer (TBD)

Linked Data Repositories with End-Points

"Cross organization data harmonization and access, linked data decentralized approach" Maj, Mahmud, Delmas-Glass – CIDOC 2012
Data Mapping

- Art Object (IDs, dimensions, images, etc.)
- Subject Terms (Places and Concepts)
- Events
  - Creation (date, creator, period, culture, technique)
  - Provenance
  - Exhibition History
  - Publication
  - Acquisition*
  - Framing History*
  - Conservation*
  - Loans?*

* Pending Internal approval, cataloging, modeling, etc.

“Cross organization data harmonization and access, linked data decentralized approach” Maj, Mahmud, Delmas-Glass – CIDOC 2012
YCBA & BM Data Differences

- YCBA Provenance is a large paragraph of text
- BM Provenance there is a large list of previous owners
- YCBA Exhibition History is a list of exhibitions
- BM Exhibitions is a single text field for each object

“Cross organization data harmonization and access, linked data decentralized approach” Maj, Mahmud, Delmas-Glass – CIDOC 2012
Thesaurus modeling for enabling federated queries

Joseph Mallord William Turner, 1775-1851, *Inverary Pier, Loch Fyne: Morning*, ca. 1845, oil on canvas, Yale Center for British Art, Paul Mellon Collection
controlled vocabularies

• Yale Center for British Art:
  – Getty Art and Architecture Thesaurus
  – Getty Thesaurus of Geographic Names
  – ICONCLASS
  – Library of Congress Subject Headings
  – Oxford Dictionary of National Biography
  – Getty Union List of Artists Names

• British Museum:
  – Local object names and materials thesauri

OAR

RT PADDLE
BTCHANDLERY

OBELISK

RT SCULPTURE
BTMONUMENT

OBI

RT WAIST-BAND
BTSASH
George Stubbs, 1724-1806, British, *Horse Frightened by a Lion*, between 1762 and 1768, Oil on canvas, Yale Center for British Art, Paul Mellon Collection.

Subject terms and tags:

- frightened
- scared
- fear
- anthropomorphism
- danger
- anthropomorphic
- rocks
- landscape
- animal art
- horse (animal)
- lion

"Cross organization data harmonization and access, linked data decentralized approach" Maj, Mahmud, Delmas-Glass – CIDOC 2012
YCBA frame vocabulary

Yale Center for British Art

Collections Exhibitions Events Research Education Architecture Visiting

Home -> Collections -> Search -> Results -> British, Louis XV frame

Search Details

Creator Unknown framemaker
Title British, Louis XV frame
Date 2nd and 3rd quarters of 18th century
Medium Carved wood, later oil gilding over original gilding and sanded rail

Dimensions
Credit Line Yale Center for British Art
Accession Number B1974.3.11FR
Collection Frames

Link to Full Record http://collections.britishart.yale.edu/vufind/Record/3414330

Cross-section Ogee with frieze
Ornaments Leaf & dart back; gadrooned rail; acanthus sight
Alteration probably not

Public Domain | Download

“Cross organization data harmonization and access, linked data decentralized approach” Maj, Mahmud, Delmas-Glass – CIDOC 2012
YCBA frame thesaurus (local)
YCBA local terms nested in AAT and TGN
AAT unique identifiers

Research

Art & Architecture Thesaurus® Online
Full Record Display

Click the icon to view the hierarchy.

ID: 300033618

Terms:

paintings (visual works) (preferred, C,U,English-P,D,U,PN)
paintings (visual work) (C,U,LC,English,AD,U,SN)
schilderingen (C,U,Dutch-P,D,U,U)
schildering (C,U,Dutch,AD,U,U)
peintures (visual works) (C,U,French-P,D,U,PN)
peinture (visual work) (C,U,French,AD,U,SN)
pinturas (obra visual) (C,U,Spanish-P,D,U,PN)
pintura (obra visual) (C,U,Spanish,AD,U,SN)

Note: Unique works in which images are formed primarily by the direct application of pigments suspended in oil, water, egg yolk, molten wax, or other liquid, arranged in masses of color, onto a generally two-dimensional surface.
LIDO: concept source and conceptID

```
<lido:descriptiveMetadata xml:lang="eng">
  <lido:objectClassificationWrap>
    <lido:objectWorkTypeWrap>
      <lido:objectWorkType>
        <lido:conceptID lido:source="AAT" lido:type="Object name">300033618</lido:conceptID>
        <lido:term>painting</lido:term>
      </lido:objectWorkType>
    </lido:objectWorkTypeWrap>
    <lido:classificationWrap>
      <lido:classification>
        <lido:conceptID lido:source="YCBA" lido:type="Genre">10</lido:conceptID>
        <lido:term>historical subject</lido:term>
      </lido:classification>
    </lido:classificationWrap>
  </lido:objectWorkTypeWrap>
  <lido:classificationWrap>
    <lido:classification>
      <lido:conceptID lido:source="AAT" lido:type="Classification">300033618</lido:conceptID>
      <lido:term>Painting</lido:term>
    </lido:classification>
  </lido:classificationWrap>
  <lido:objectIdentificationWrap>
    <lido:titleWrap>
```

“Cross organization data harmonization and access, linked data decentralized approach” Maj, Mahmud, Delmas-Glass – CIDOC 2012
### Dereferenceable URIs

**at Yale Center for British Art**

http://collection.britishart.yale.edu/id/ney/tt/192793

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>is ?P129_is_about</code> of</td>
<td>[<a href="http://collection.britishart.yale.edu/id/object/10342">http://collection.britishart.yale.edu/id/object/10342</a>]</td>
</tr>
<tr>
<td></td>
<td>[<a href="http://collection.britishart.yale.edu/id/object/1176">http://collection.britishart.yale.edu/id/object/1176</a>]</td>
</tr>
<tr>
<td></td>
<td>[<a href="http://collection.britishart.yale.edu/id/object/1200">http://collection.britishart.yale.edu/id/object/1200</a>]</td>
</tr>
<tr>
<td></td>
<td>[<a href="http://collection.britishart.yale.edu/id/object/134">http://collection.britishart.yale.edu/id/object/134</a>]</td>
</tr>
<tr>
<td></td>
<td>[<a href="http://collection.britishart.yale.edu/id/object/1416">http://collection.britishart.yale.edu/id/object/1416</a>]</td>
</tr>
<tr>
<td></td>
<td>[<a href="http://collection.britishart.yale.edu/id/object/203">http://collection.britishart.yale.edu/id/object/203</a>]</td>
</tr>
</tbody>
</table>

| `?:inScheme`      | [http://collection.britishart.yale.edu/300192793] |
| `?:prefLabel`     | audience |
|                   | audiences |
| `?:type`          | [http://erlangen-crm.org/current/E55] |
|                   | [http://www.w3.org/2004/02/skos/core] |

---

This page shows information obtained from the SPARQL endpoint at http://collection.britishart.yale.edu/id/ney/tt/192793.

---

“Cross organization data harmonization and access, linked data decentralized approach” Maj, Mahmud, Delmas-Glass – CIDOC 2012
YCBA tool to add unique identifiers to local terms

<table>
<thead>
<tr>
<th>Term</th>
<th>LocalTerm</th>
<th>CN</th>
<th>SourceTermID</th>
<th>ScopeName</th>
</tr>
</thead>
<tbody>
<tr>
<td>military engineers</td>
<td>AUT.AAE.AAB.AAA.AA</td>
<td>25086</td>
<td>Practitioners of military engineering</td>
<td></td>
</tr>
<tr>
<td>assistants</td>
<td>AUT.AAE.AAB.AAA.AA</td>
<td>25898</td>
<td>Persons who act as subordinates</td>
<td></td>
</tr>
<tr>
<td>architectural societies</td>
<td>AUT.AAE.AAB.AAA.AA</td>
<td>26011</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Architectural Baroque moulding</td>
<td>AAT.02.002</td>
<td>-1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Architectural cassetta</td>
<td>AAT.02.004</td>
<td>-1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>architectural societies</td>
<td>AUT.AAE.AAB.AAA.AA</td>
<td>26011</td>
<td></td>
<td></td>
</tr>
<tr>
<td>architectural critics</td>
<td>AUT.AAE.AAB.AAA.AA</td>
<td>25520</td>
<td>Those who specialize in architectural critics</td>
<td></td>
</tr>
<tr>
<td>architectural critics</td>
<td>AUT.AAE.AAB.AAA.AA</td>
<td>25520</td>
<td>Those who specialize in architectural critics</td>
<td></td>
</tr>
<tr>
<td>architectural critics</td>
<td>AUT.AAE.AAB.AAA.AA</td>
<td>25520</td>
<td>Those who specialize in architectural critics</td>
<td></td>
</tr>
<tr>
<td>architectural historians</td>
<td>AUT.AAE.AAB.AAA.AA</td>
<td>25540</td>
<td>Those who specialize in architectural historians</td>
<td></td>
</tr>
<tr>
<td>architectural critics</td>
<td>AUT.AAE.AAB.AAA.AA</td>
<td>25520</td>
<td>Those who specialize in architectural critics</td>
<td></td>
</tr>
<tr>
<td>architectural critics</td>
<td>AUT.AAE.AAB.AAA.AA</td>
<td>25520</td>
<td>Those who specialize in architectural critics</td>
<td></td>
</tr>
<tr>
<td>architectural historians</td>
<td>AUT.AAE.AAB.AAA.AA</td>
<td>25540</td>
<td>Those who specialize in architectural historians</td>
<td></td>
</tr>
<tr>
<td>architectural critics</td>
<td>AUT.AAE.AAB.AAA.AA</td>
<td>25520</td>
<td>Those who specialize in architectural critics</td>
<td></td>
</tr>
<tr>
<td>architectural critics</td>
<td>AUT.AAE.AAB.AAA.AA</td>
<td>25520</td>
<td>Those who specialize in architectural critics</td>
<td></td>
</tr>
<tr>
<td>architectural historians</td>
<td>AUT.AAE.AAB.AAA.AA</td>
<td>25540</td>
<td>Those who specialize in architectural historians</td>
<td></td>
</tr>
<tr>
<td>firms</td>
<td>AUT.AAE.AAB.AAB.AA</td>
<td>25971</td>
<td>Partnerships of two or more firms</td>
<td></td>
</tr>
</tbody>
</table>

“Cross organization data harmonization and access, linked data decentralized approach” Maj, Mahmud, Delmas-Glass – CIDOC 2012
St. Paul’s cathedral is in London

Turner was born in London

Canaletto, *St. Paul’s Cathedral*, ca. 1754, oil on canvas, Yale Center for British Art, Paul Mellon Collection

Attributed to Richard Dighton, *Portrait of JMW Turner 1828-1838*, British Museum

“Cross organization data harmonization and access, linked data decentralized approach” Maj, Mahmud, Delmas-Glass – CIDOC 2012
OCLC unique identifiers

The pursuit of happiness: a view of life in Georgian England: an exhibition selected from the Paul Mellon Collection, Yale Center for British Art, New Haven, April 19 to September 18, 1977

Author: J.H. Plumb, Edward J. Hyman, Nancy L. Pressly. Yale Center for British Art
Publisher: [New Haven]: The Center, ©1977.
Edition/Format: [Book: English]
Rating: [not yet rated] (5 with reviews - Be the first)

OCLC Number: 3447120

<lido:objectID lido:source="OCLC" lido:type="local">3447120</lido:conceptID>
Modeling Authority Records

- YCBA – Object Authority
  - Object [YCBA:499]
  - Publication [YCBA:191]
  - Exhibition History [URI]
  - Creation [URI]

- OCLC - Book Authority
  - Object [OCLC:3447120]

- VIAF - Personal Names
  - Object [VIAF:5203004]

- Library of Congress - Names
  - Object [LC:n85157944]

- BM – Object Authority
  - Object [BM:12]
    - Loan Out

“Cross organization data harmonization and access, linked data decentralized approach” Maj, Mahmud, Delmas-Glass – CIDOC 2012
Harmonizing vocabularies

• Thesauri modeled using SKOS Concept & CRM E42 Type

• Owl:sameAs property:

```xml
<rdf:Description rdf:about="#James_Mallord_William_Turner">
  <owl:sameAs rdf:resource="#JMWTurner"/>
</rdf:Description>
```

• Use URIs of linked data vocabularies

• WordNet to provide automated URIs
Cross organization data harmonization and access, linked data decentralized approach

Maj, Mahmud, Delmas-Glass – CIDOC 2012
Visual SPARQL Tools

“Cross organization data harmonization and access, linked data decentralized approach” Maj, Mahmud, Delmas-Glass – CIDOC 2012
Rembrandt Federated Search Results (Sofia - OWLIM)

- Rembrandt Database: Saskia van Uylenburgh in Arcadian Costume
  http://www.remmbrandtdatabase.org/obj/22578
- Rembrandt Database: Man in Oriental Costume
  http://www.remmbrandtdatabase.org/obj/58650

Yale Federated Search Results (New Haven - OWLIM)

- Yale Database: Costume Design
  http://collection.britishart.yale.edu/id/object/10631
- Yale Database: The Countess of Coningsby in the Costume of the Charlton Hunt
  http://collection.britishart.yale.edu/id/object/1084
- Yale Database: Costume Study for a Portrait of a Man
  http://collection.britishart.yale.edu/id/object/12582
- Yale Database: Costume Study for the Portrait of a Lady
  http://collection.britishart.yale.edu/id/object/14141
- Yale Database: Study of a Gentleman in 17th Century Costume
  http://collection.britishart.yale.edu/id/object/15103
- Yale Database: Figures in Totemic Costume and Totem Pole, Number 149
  http://collection.britishart.yale.edu/id/object/20671

Courtesy: Dominic Oldman @ British Museum, Research Space
Semantic Development - Future

• Ability to query data via unified vocabularies
• References between detailed representation of data ie; allow bounding box map queries
• Guidelines for museums, similar to what libraries are doing
• Add Library and Archive data to the mix
• Other institution SPARQL endpoints / data