Linked Data in Cultural Heritage

Open up **data silos**

- Improve reusability data
- Support integration collections
- Identifiers for things
- Cross-referencing
- Lins across collections
- Shared views & context of objects
- Data models for interoperability

*Researchers & Collection Managers using it for **deep analysis** of objects and collections as a whole*
Rijksmuseum Amsterdam

Collection
- Collection of ~1,000,000 objects
- Artworks on display ~8,000
- Dutch Masters like Rembrandt

Online Collection
- Accessible through API
- 597,193 object records
- 207,441 works have CC0 image

Images are released in the public domain for users & developers

https://www.rijksmuseum.nl/en/api
Digitisation projects

Professional catalogers and photographers

- Register artworks
- Provide annotations
- Digitise artworks
- Publish them online

~40,000 new object records a year

time consuming & costly endeavour

Versioning of data
Data from collection management is harvested daily & loaded in a database serving the website.
• A subset of **245 metadata fields** (597 in total) are included in the output of collection management

• Fields **no longer used or contain sensitive data**, e.g. insurance values are excluded

• The selected fields are transformed to form field names which **better reflect their content**, **omit empty values** and **generate links to other databases** maintained by the Rijksmuseum (XSLT)
Conversion to Linked Data

Relevant metadata fields of a collection object are mapped to the **Europeana Data Model** that most closely resembles the values of the field.

The output of the API is used to obtain a complete harvest of the data, which is in turn loaded into a triple store (run on a monthly basis with links to downloads of older versioned datadumps).
modelling the complete collection & integrating it with other collections from other institutions required the ability to model different (potentially conflicting) metadata records from different sources describing the same artwork
Europeana Data Model

The Rijksmuseum dataset was one of the first entries in the Europeana Thought Lab Images converted to comply with the VRA data model, 46K

The data model is designed with reuse of existing classes and properties in mind. It includes elements from the Dublin Core metadata initiative and the Object Reuse and Exchange definition of the Open Archives Initiative.
**Links to external datasets**

**Iconclass**
- Concepts about subjects, themes and motifs in Western art
- Links artworks to subject

**Art & Architecture Thesaurus (AAT)**
- Concepts about art styles, materials and agents
- Links artworks to type and format

**Short-Title catalogue Netherlands (STCN)**
- retrospective national bibliography of the Netherlands maintained by the National Library of the Netherlands.
- books that are the source of objects in the print collection of the Rijksmuseum
Links to external datasets

"Rijksmuseum" provided CHO
SK-A-3276
Concept
71077

"Rembrandt Harmensz. van Rijn" provided CHO
Agent
PEOPLE.5706

"Jeremiah Lamenting the Destruction of Jerusalem" @en

Vocabularies

Concept 300015050
prefLabel

"Old Testament" @en
term

Concept 1000014078-en
term

Concept 71
prefLabel

"oil paint" @en

Concept 71077
broader

"Jeremiah lamenting over the destruction of Jerusalem" @en

term

 aggregations

isShownBy

aggregated CHO

creator

WebResource

aggregation

COL.5242

dataProvider

isShownBy

aggregated CHO

creator

ProvidedCHO

SK-A-3276

title

"Jeremiah Lamenting the Destruction of Jerusalem" @en
22,846,996 triples describing 351,814 objects
207,441 with graphical depiction

Ten sub-collections are maintained:
• sculptures (29,782 objects)
• historical items (19,936 objects)
• paintings (3,949 objects)
• Asian art (3,722 objects)
• prints, drawings & photos (280,047 objects)

Overview of the predicates that describe collection items

<table>
<thead>
<tr>
<th>predicate</th>
<th>distinct artworks</th>
<th>distinct resources</th>
<th>vocabularies</th>
<th>distinct literals</th>
<th>en literals</th>
<th>nl literals</th>
</tr>
</thead>
<tbody>
<tr>
<td>dc:contributor</td>
<td>89,796</td>
<td>0</td>
<td>-</td>
<td>9,146</td>
<td>0</td>
<td>9,146</td>
</tr>
<tr>
<td>dc:coverage</td>
<td>138,141</td>
<td>0</td>
<td>-</td>
<td>212</td>
<td>106</td>
<td>106</td>
</tr>
<tr>
<td>dcterms:created</td>
<td>340,865</td>
<td>0</td>
<td>-</td>
<td>46,255</td>
<td>19,837</td>
<td>19,837</td>
</tr>
<tr>
<td>dc:creator</td>
<td>349,787</td>
<td>27,904</td>
<td>rma</td>
<td>38,851</td>
<td>97</td>
<td>38,754</td>
</tr>
<tr>
<td>dc:description</td>
<td>217,202</td>
<td>0</td>
<td>-</td>
<td>176,786</td>
<td>2,487</td>
<td>174,299</td>
</tr>
<tr>
<td>dcterms:extent</td>
<td>288,318</td>
<td>0</td>
<td>-</td>
<td>55,241</td>
<td>15,229</td>
<td>40,012</td>
</tr>
<tr>
<td>dc:format</td>
<td>322,152</td>
<td>593</td>
<td>aat, rma</td>
<td>924</td>
<td>322</td>
<td>602</td>
</tr>
<tr>
<td>dcterms:hasPart</td>
<td>487</td>
<td>0</td>
<td>-</td>
<td>13,646</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>dc:identifier</td>
<td>351,814</td>
<td>0</td>
<td>-</td>
<td>703,429</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>dcterms:isPartOf</td>
<td>59,197</td>
<td>0</td>
<td>-</td>
<td>2,754</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>dcterms:isReferencedBy</td>
<td>98,815</td>
<td>0</td>
<td>-</td>
<td>80,770</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>dc:language</td>
<td>351,814</td>
<td>0</td>
<td>-</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>dcterms:provenance</td>
<td>13,239</td>
<td>0</td>
<td>-</td>
<td>1,155</td>
<td>0</td>
<td>1,155</td>
</tr>
<tr>
<td>dc:publisher</td>
<td>351,814</td>
<td>0</td>
<td>-</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>dcterms:spatial</td>
<td>214,752</td>
<td>3,339</td>
<td>rma</td>
<td>3,361</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>dc:subject</td>
<td>221,868</td>
<td>44,840</td>
<td>ic, rma</td>
<td>32,452</td>
<td>0</td>
<td>32,452</td>
</tr>
<tr>
<td>dc:title</td>
<td>351,789</td>
<td>0</td>
<td>-</td>
<td>297,271</td>
<td>6,784</td>
<td>290,487</td>
</tr>
<tr>
<td>dc:type</td>
<td>351,749</td>
<td>3,541</td>
<td>aat, rma</td>
<td>3,700</td>
<td>155</td>
<td>3,545</td>
</tr>
<tr>
<td>edm:type</td>
<td>351,814</td>
<td>0</td>
<td>-</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Types in the Rijksmuseum thesaurus with more than 500 values

<table>
<thead>
<tr>
<th>type</th>
<th>distinct resources</th>
<th>nl labels</th>
<th>en labels</th>
</tr>
</thead>
<tbody>
<tr>
<td>person</td>
<td>38,939</td>
<td>27,904</td>
<td>27,904</td>
</tr>
<tr>
<td>place</td>
<td>17,174</td>
<td>17,174</td>
<td>152</td>
</tr>
<tr>
<td>object name</td>
<td>6,074</td>
<td>6,074</td>
<td>298</td>
</tr>
<tr>
<td>keyword</td>
<td>5,021</td>
<td>5,021</td>
<td>166</td>
</tr>
<tr>
<td>event</td>
<td>1,982</td>
<td>1,982</td>
<td>43</td>
</tr>
<tr>
<td>technique</td>
<td>1,401</td>
<td>1,401</td>
<td>73</td>
</tr>
<tr>
<td>occupation</td>
<td>1,044</td>
<td>1,044</td>
<td>15</td>
</tr>
<tr>
<td>material</td>
<td>882</td>
<td>882</td>
<td>428</td>
</tr>
<tr>
<td>location RMA</td>
<td>808</td>
<td>808</td>
<td>5</td>
</tr>
</tbody>
</table>
Frequency distributions of the top 50 concepts of AAT & Iconclass in Rijksmuseum collection

A small subset of concepts is often used:
- 305 distinct formats
- 124 distinct types
- prints (183,916)
- stereoscopic photographs (3,480)
- plates (1,617)
- art styles are often debatable

Many concepts are often used (ave ~ 27 times):
- 39,578 concepts in the vocabulary
- 10,434 are used to add information to an object
- 351,814 collection objects
- 172,059 have one or more Iconclass annotations
Annotating Artworks

Focus on art-historical information

Occasional lack of expertise regarding subject matter annotations

This print is described as:

- “Bird with blue head”
- “Branch with red leaves”
Experts are adding information

Create links using Accurator annotation tool
http://annotation.accurator.nl/

Organise annotation events

- Bird watching event
- Fashion event
Dialogue about data

Publishing data widens the type of users involved

Engage in a dialogue

› What information is needed?
› Which vocabularies to use?
› Which fields can be used to describe the objects?
Unleashed creativity
Many prints originate from books

- References to these books are added as curators comments

Short-Title catalogue Netherlands

- Retrospective national bibliography in the period 1540-1800
- Includes 139,817 publications

Linking books to prints

- Scan for curator comments containing Title, Author and Year
- 3598 links from prints to 501 publications
Possibilities for integration - Rijksmuseum Website
Possibilities for integration - Dutch Species

Gekraagde roodstaart
_Phoenicurus phoenicurus_

Foto J.M.D. (Jan) van Es, 29 juni 2013, Kamp terrein Westerbork bij Hooghalen

Afbeelding(en): 7

Foto J.M.D. (Jan) van Es
Foto Aat Bender
Foto Jaap Denee
Foto Jaap Denee
Foto Hans van der Meulen
Foto Hans van der Meulen
Foto Hans van der Meulen
Print Kono Bairei
How to use the data

**All at once**
monthly datadumps
[https://datahub.io/dataset/rijksmuseum](https://datahub.io/dataset/rijksmuseum)

**Request based**
OAI API

**Queries**
SPARQL Endpoint
[https://datahub.io/dataset/rijksmuseum](https://datahub.io/dataset/rijksmuseum)