

SemJoconde

Toward a Semantic Representation Of the Joconde Database

Jean-claude.moissinac@telecom-paris.fr

Telecom Paris

Authors: Jean-Claude Moissinac, François Rouzé, Bastien Germain, Piyush Wadhera

October 2020

Presenter

- Jean-Claude Moissinac
- Associate Professor
- Telecom Paris



- Research lead for the Data&Musée project
 - Aiming to discover knowledge from data provening from museums and monuments enriched by the Linked Open Data

SemJoconde Goal

- Contribute to LOD with Knowledge Graphs about French Cultural Heritage
- Get an enriched semantic version of the Joconde database
- Part of the Data&Musée project

Data sources

- Joconde Database
 - Collected by the French Ministry of Culture
 - Data about 580035 creations, part of the French cultural heritage
- Wikidata
- Wikidata:Joconde project
 - https://www.wikidata.org/wiki/Wikidata:WikiProject_Vocabulaires_Joconde/
 - Expert linking between Joconde database vocabulary and Wikidata entities
 - 5/30/2020, validated links:
 - 2560 creators, 153 domains, 47 places, ...

Joconde database content

- For each creation
 - **Ref**: an unique string, internal ID of the work
 - **Name of creators**: anonymous, "RODIN Auguste" ...
 - **Title**: "Tête d'homme barbu"
 - **Domains**: "Néolithique", "photographie",
 - **Techniques**: "bas relief"
 - **Period**: "2e quart 20e siècle"
 - **Dimensions**: "H. 27; l. 6.1 ; P. 4.2"
 - **Location**: "Grenoble ; musée Stendhal"

Source fields and ground truth

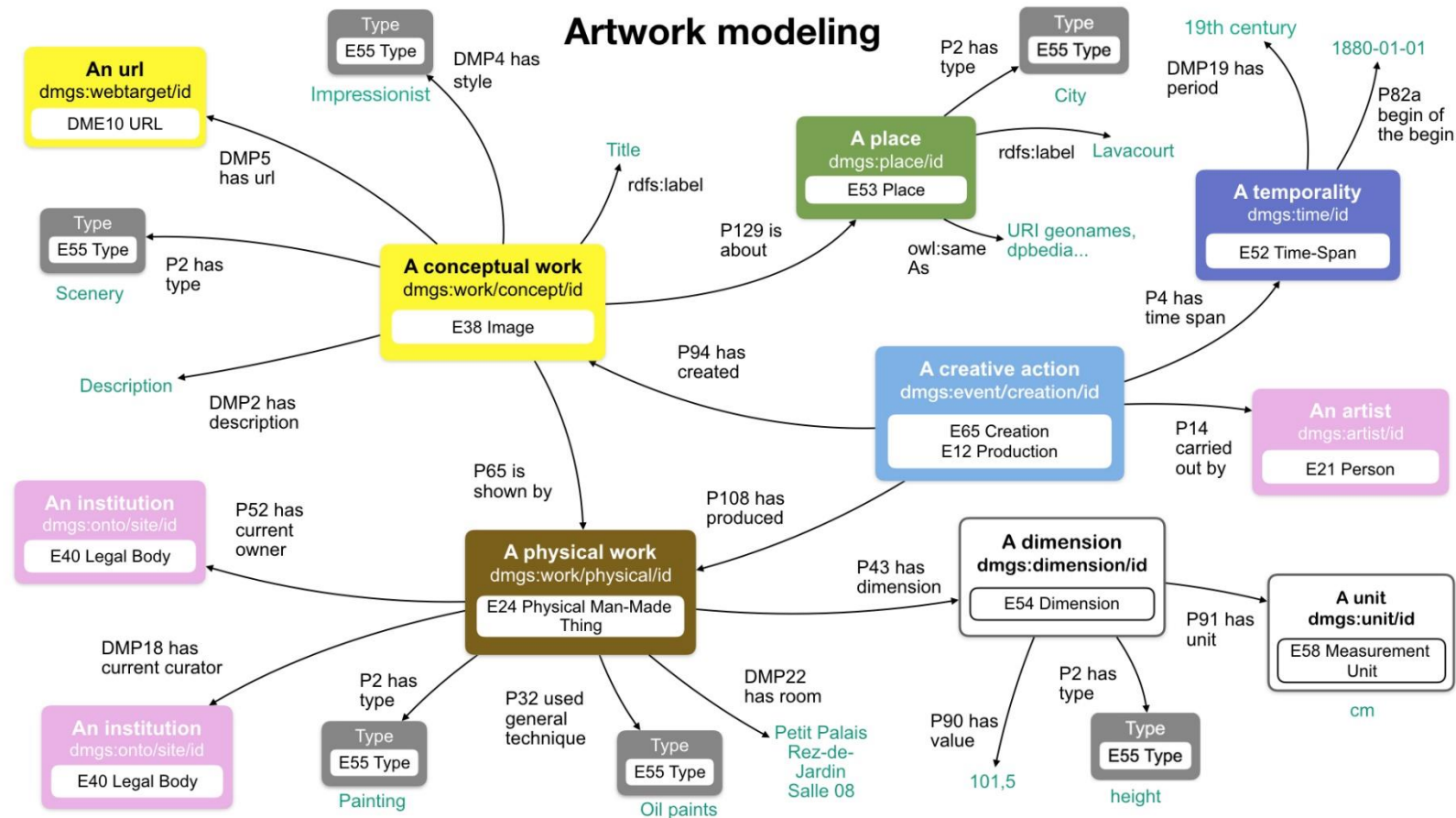
Category (field)	Validated in ground truth	Source	% of validated links
Creators	2560	37828	6.7 %
Domains	153	153	100 %
Object types	77	5766	1.3 %
Places	35	3593	0.9 %
Epochs	500	831	60.1 %
Periods	60	346	17.3 %



Model

- Based on CIDOC-CRM
 - CIDOC = ICOM International Committee for Documentation
 - Developing standards for information management in museums
 - CRM = Conceptual Reference Model
 - Used by: Europeana, British Museum, Paris Musées Collections, ...
- With added properties like
 - DMP5_has_url, for a link to a significant web page
 - DMP19_has_period; ex.: 19th century

Artwork Model

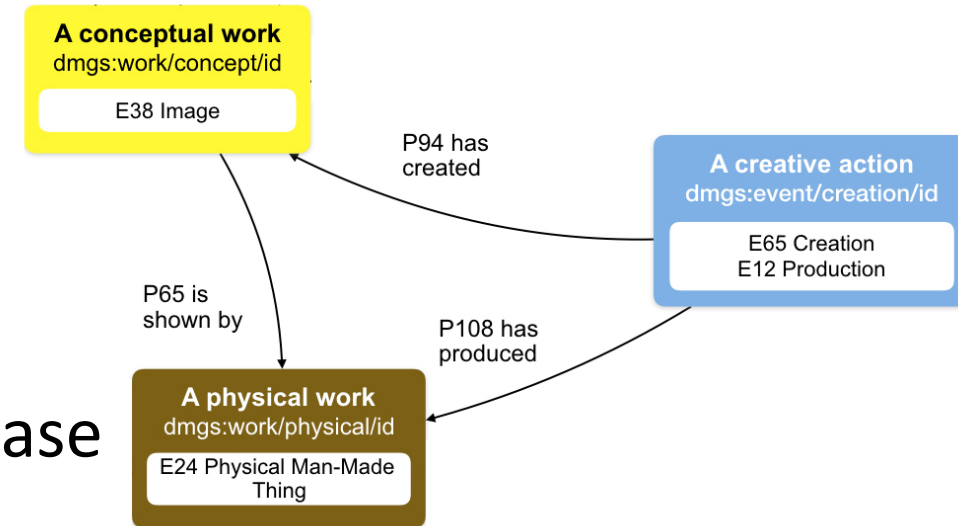


Process

- Create three URIs for each work
- List the possible values for each field
- Clean values
 - Ex.: sculpture=Sculpture
- Count the number of works of each value
- Try to link each value with Wikidata
- Expert check of link for the most used values
- Check links against a ground truth to evaluate quality of the resulting representation

3 for each work

- URI=Universal Resource Identifier
- Based on the REF field of the Joconde database
 - Example: REF = [AE037477](#)
- Creation event
 - <http://datamusee.givingsense.eu/onto/event/creation/AE037477>
- Physical work
 - <http://datamusee.givingsense.eu/onto/work/physical/AE037477>
- Conceptual work:
 - <http://datamusee.givingsense.eu/onto/work/concept/AE037477>



Entity linking, a simple algorithm

- Giving a label
- Produce variants: lowercase, uppercase, title case, push the first word in the last place
- Use the Wikidata search service with each variant
 - Within a sparql query
 - With a target for types of entities (see sample in the next slide)
- If only one link found, get it
- If several links found
 - If the label for one link matches the searched label in lowercase, get the link
 - Else get nothing



Example: link about a creator

- Leonard de Vinci
 - Variants: leonard de vinci, LEONARD DE VINCI, Leonard De Vinci, De Vinci Leonard
- Possible types:
 - Wikidata entities
 - Painter (Q1028181), sculptor (Q1281618), drawer (Q15296811), artist (Q483501), photographer (Q33231)... (see article)

Techniques

- 9697 terms used
- Hundred most used covers 88% of works
- Useful Wikidata types:
 - Artistic technique Q11177771
 - Painting material Q3300034
- Exact match between these types and terms: 45
- Coverage of these 45 terms: 43%

Creators

No known creator	16.83 % of works
Creator known as «Anonymous »	10.72 % of works
Identified creators	426.885 works
Creators producing more than 10 works	5217
Validated links in the ground truth	2560
Productive creators in the ground truth	1325 (covering 25.04 % of works)
Links found by our algorithm	1180
Exact links found	1177

Evaluation about creators

Measure	Value
Number of creators validated by the Wikidata Joconde project and targets of the evaluation	1315
Number of creators for which our algorithm finds a Wikidata entity	1180
Number of exact links found	1177
Precision relative to creators	99.74
Recall relative to creators	88.83
F1 score	93.97



Museums and other sites

Number of sites in the Joconde database	3593
Sites linked in the ground truth	0
«Richest» sites (with more than 100 works)	304
Coverage of the richest sites	98.43 % (580035 works)
Found links by our algorithm	155
Manually checked links (first sites)	76
Exact links in the checked links	100 % (76)

Measure	Value
Searched links	76
Found links	42
Exact links	42
Precision	100
Recall	55.26
F1	71.18

Conclusion

- Closed to 600.000 works described by triples
- 1918315 links to wikidata
- Publicly available: <https://github.com/datamusee/semjoconde>
- Asserted level of quality
- Possibly useful for future works
 - Machine learning
 - French cultural heritage vocabulary

Questions

- Jean-claude.moissinac@telecom-paris.fr