



# An observational study of equivalence links in cultural heritage linked data for agents

Nuno Freire, Hugo Manguinhas, Antoine Isaac

Danse de trols faunes et trols bacchantes  
Hopfer, Hieronymus  
1500-1563, Bibliothèque Municipale De Lyon  
France, Public Domain

TPDL 2020 – Theory and Practice in Digital Libraries 2020



Co-financed by the European Union  
Connecting Europe Facility

# Introduction

- We conducted an observational study of the virtual graph formed by equivalence relations between entities of eight open Knowledge Bases (KBs)
  - We studied entities of type agent (persons, organizations)
  - ... in cultural heritage data

# The eight KBs in our study :

- DBpedia
- data.bnf.fr (BnF)
- datos.bne.es (BNE)
- Library of Congress Names (NAF)
- The Union List of Artist Names (ULAN)
- Gemeinsame Normdatei (GND)
- Virtual International Authority File (VIAF)
- Wikidata

# Introduction (cont.)

- We measured the quantity of equivalences that this graph could provide for a dataset from Europeana containing references to agents in descriptions of cultural heritage objects.
- This study is informative for designing innovative applications, such as the case of Europeana who seeks to acquire agent name variants/translations or extra biographical information.

# The study

Tak met vier mangolia's  
Anonymous  
1910-1925, Rijksmuseum  
Netherlands, Public Domain



# The equivalence graph

- We considered the transitive closure of the set of equivalence statements
  - It forms a virtual graph with entities from all the knowledge bases as nodes
  - We considered all statements where the property was one of
    - owl:sameAs
    - skos:exactMatch
    - skos:closeMatch
    - schema:sameAs

# The two parts of the study

- We conducted two studies of the equivalence graph
  - First, we measured the amount of stated equivalence relations between KBs
  - Second, we analysed the entity type *agent*

# The amounts of equivalence statements involving each knowledge base

Considering all equivalence properties

KB	As subject of equivalences		As object of equivalences		Total statements
	Statements	to KBs	Statements	from KBs	
VIAF	25,118,745	7	21,666,779	6	46,785,524
GND	11,313,935	4	9,454,213	5	20,768,148
NAF	6,101,051	1	14,216,491	6	20,317,542
Wikidata	4,624,309	6	9,785,342	4	14,409,651
<u>DBpedia</u>	7,396,520	3	977,907	5	8,374,427
<u>BnF</u>	4,505,773	5	3,124,674	3	7,630,447
BNE	997,183	5	698,329	2	1695,512
ULAN	249,812	2	383,593	2	633,405

Considering only skos:closeMatch properties

KB	As subject of equivalences		As object of equivalences	
	Statements	to KBs	Statements	from KBs
GND			25,952	1
NAF			150,224	2
Wikidata			6	1
<u>BnF</u>	67,746	3		
BNE			16,118	1
ULAN	124,554	2		



# The study of the entity type *agent*

- We created a set of URIs referring to agents from the dataset of Europeana
  - containing 286,090 unique agent URIs
- This set was then used to initiate the crawling iterations of the equivalence graph

# The results of the 4 crawling iterations of the Europeana set of agent URIs

KB	Initial Europeana set (a)	New equivalences found after each iteration (b)				% of the initial Europeana set with equivalences (c)
		1 <sup>st</sup> Crawl	2 <sup>nd</sup> Crawl	3 <sup>rd</sup> Crawl	4 <sup>th</sup> Crawl	
<u>DBpedia</u>	2	4,407	34,968	47,031	47,410	16.57%
<u>BnF</u>	2,010	6,282	9,803	53,280	54,554	19.07%
<u>BNE</u>	30,449	3,321	9,952	12,471	12,934	4.52%
<u>NAF</u>	0	11,935	15,554	77,702	78,207	27.34%
<u>ULAN</u>	7,451	1,737	3,439	12,137	12,701	4.44%
<u>GND</u>	242,297	7,684	8,596	14,939	15,100	5.28%
<u>VIAF</u>	2,174	13,095	170,057	173,608	173,613	60.68%
<u>Wikidata</u>	0	1,651	92,588	98,450	98,813	34.54%
<b>Total</b>	<b>284,383</b>	<b>50,112</b>	<b>344,957</b>	<b>489,618</b>	<b>493,332</b>	-
<b>Δ from previous crawl</b>	-	-	<b>588%</b>	<b>42%</b>	<b>0.76%</b>	-

a - number of URIs of each KB in the Europeana set

b - number of equivalences found after each iteration

c - percentage of the Europeana URIs that after the 4<sup>th</sup> iteration have an equivalence to the KB considered.

# Conclusions and Future work

Chat "regardant" à travers une longue-vue et  
autre chat perché dessus

Agence Rol. Agence photographique,  
Bibliothèque national de France  
France, Public Domain

14.045

ROL 14045 - K138688

# Conclusions

- We observed that agents in KBs are highly interlinked.
  - The majority of equivalences are expressed with exact equivalence predicates (like owl:sameAs), while matches with uncertainty (skos:closeMatch) are a minority of 0.3%
  - Although each KB is not directly linked to all other KBs, all KBs are a source and a target of equivalence links.
- Crawling of the agent URIs used in Europeana required only three iterations to collect 99% of the equivalences
  - VIAF is the KB with the highest number of agent equivalences (60.7%), followed by Wikidata (34.5%).

# Future work

- The detection of possibly incorrect equivalences, since this study has detected some quality issues in the (owl:sameAs) links.
- To estimate recall issues, i.e. whether many new links could be created across KBs via automatic or manual alignment



**Thank you for your attention**

**nuno.freire@tecnico.ulisboa.pt**

**Arrival of a Portuguese ship**  
**Anonymous**

1660 - 1625, Rijksmuseum  
Netherlands, Public Domain

## **Acknowledgments**

Fundação para a Ciência e a Tecnologia (FCT): UID/CEC/50021/2013

European Commission contract number 30-CE-0885387/00-80.