This project is funded under the
ICT Policy Support Programme part of the
Competitiveness and Innovation Framework Programme.

Project

Project Acronym:	AthenaPlus
Grant Agreement number:	325098
Project Title:	Access to cultural heritage networks for Europeana

Deliverable

Deliverable name:	**Linking of metadata to external data sources**
Deliverable number:	**D4.6**
Delivery date:	October 2015
Dissemination level:	Public
Status	Final
Authors (organisation)	Regine Stein (UNIMAR) Nikolaos Simou (NTUA)
Contributors (organisation)	Karin Glasemann
Reviewers (organisation)	Gordon McKenna (CT)

Revision History

Revision	Date	Author	Organisation	Description
	2015-07	Regine Stein Nikolaos Simou	UNIMAR NTUA	Outline and scope of the deliverable
V0.1	2015-10-26	Regine Stein Nikolaos Simou	UNIMAR NTUA	Draft version
V0.2	2015-10-30	Gordon McKenna	CT	Review
V1.0	2015-11-23	Regine Stein	UNIMAR	Final version
		Maria Teresa Natale	ICCU	Formal check

Statement of originality

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Project Coordinator: Istituto centrale per il catalogo unico delle biblioteche italiane
Address : Viale Castro Pretorio 105 – 00185 Roma
Phone number : +3906 06 49210 425
E-mail : info@athenaplus.eu
Project WEB site address : http://www.athenaplus.eu

Table of Contents

1 EXECUTIVE SUMMARY 5

2 INTRODUCTION 6
2.1 Background 6
2.2 Role of this Deliverable in the Project 6

3 ATHENAPLUS REVIEW ON LINKED OPEN DATA SOURCES 7
3.1 Recommendations from the AthenaPlus review 7
3.2 Evaluation of LOD Sources: Update 8

4 PREREQUISITES FOR LINKING TO EXTERNAL DATA SOURCES 14
4.1 Background from Europeana 14
4.2 Requirements from Content Providers 14
4.3 Use of Underlying Terminologies 17

5 PRESENTATION OF SELECTED EXTERNAL DATA SOURCES 18
5.1 Selection Criteria 18
5.2 Presentation of Selected Sources 18
5.2.1 Final Selection Process 18
5.2.2 Getty Union List of Artist Names (ULAN) 19
5.2.3 Getty Thesaurus of Geographic Names (TGN) 20

6 LINKING ATHENAPLUS CONTENT 21
6.1 Web Services for linking to ULAN and TGN vocabularies 21
6.1.1 Data model definition for services 21
6.1.2 Approach and evaluation functions 22
6.1.3 Use of the services 22
6.2 Evaluation of the linking process 24
6.2.1 Linking to ULAN vocabulary 25
6.2.2 Linking to TGN vocabulary 30
6.3 Providing linking results to Europeana 32
6.4 Linking to other external sources 33

7 CONCLUSIONS 34

APPENDIX 1: REFERENCES 35

APPENDIX 2: TERMS AND ABBREVIATIONS 36

APPENDIX 3: ULAN INPUT JSON 37

APPENDIX 4: TGN INPUT JSON 38

APPENDIX 5: ULAN OUTPUT JSON 40

APPENDIX 6: TGN OUTPUT JSON 42

1 EXECUTIVE SUMMARY

The main objective of this deliverable is to present the approach taken by the AthenaPlus project to encourage activities and support partners in the provision of semantically richer metadata to the cultural heritage community, and particularly Europeana, by linking the metadata to external data sources.

As this deliverable builds on the Review of Linked Open Data sources carried out earlier in the AthenaPlus project (Deliverable D4.2), Section 3 first presents a summary of the recommendations from that review as well as an update on the suggested LOD sources identified as possible candidates for linking.

Section 4 then describes the prerequisites for linking to external data sources. Taking into account considerations from both Europeana's side, as expressed by their taskforces on a Multilingual and Semantic Enrichment Strategy and on Metadata Quality, and from the data providers' side, the following important factors were identified for a successful, high-quality and sustainable enrichment of metadata with links to external sources:

1) The enrichment should be adopted as early as possible in the process of metadata production.
2) The enrichment should be based not only on string matching mechanisms, but rather exploit further contextual information.
3) The enrichment workflow should allow the data providers to validate and use the enrichment results according to their own specific criteria.

The core strategy therefore followed was to support the enrichment of underlying local terminologies with links to the external sources, with the ultimate goal to feed these links into the metadata provided, both updating existing metadata and enriching any newly produced metadata.

The final selection criteria and the selection process arising from this strategy are presented in Section 5 – finally two Getty vocabularies, the Union List for Artist Names for agents, and the Getty Thesaurus for Geographic names for places were chosen from the list of possible targets. They particularly satisfy the following criteria:

- They are published with an Open License.
- They support RDF and SPARQL 1.1.
- Many partners aim at linking to the Getty vocabularies anyway as they are domain-specific and most reliable sources in terms of quality and sustainability.

The actual linking process has been implemented as a web service which is described in Section 6. The evaluation of the web services suggests that the linking process to ULAN is of high validity, though only using the life dates of an agent as additional context information to the preferred and alternative names. Decreasing confidence values as well as the number – one or multiple – of matches returned by the service clearly correlate with imprecise or incomplete data in either the source or the target record(s). For the linking process to TGN, however, further refinements are suggested as the context information as currently exploited – continent, nation, and other places the actual place in question is part of – is not sufficiently ensuring unique and valid results.

In general, the strategy of enriching underlying terminologies in the local metadata production environment with links to external sources proves to be very efficient for increasing the number of links and thereby the quality of the metadata on cultural heritage objects eventually published as Linked Open Data.

Also Europeana benefits in the most efficient and sustainable way from the linking results, i.e. by harvesting records that have been annotated by the semi-automatic way here described that guarantees good results.

2 INTRODUCTION

2.1 Background

Linking of metadata to other, external sources is a very important factor in terms of quality in a Linked Open Data Environment. According to the four rules for Linked Data as proposed by Tim Berners-Lee it relates to the fourth rule:[1]

1.	Use URIs as names for things;
2.	Use HTTP URIs so that people can look up those names;
3.	When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL);
4.	Include links to other URIs, so that they can discover more things.

or to the fifth star in terms of his five star rating system as a "road to good linked data"[2].

★	Available on the web (whatever format) *but with an open licence, to be Open Data*
★★	Available as machine-readable structured data (e.g. excel instead of image scan of a table)
★★★	as (2) plus non-proprietary format (e.g. CSV instead of excel)
★★★★	All the above plus, Use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff
★★★★★	All the above, plus: Link your data to other people's data to provide context

While one might see the linking to external sources as the "cherry on the cake", because it appears as the last point in both lists, it can actually be argued that it is the core idea of Linked Data – the idea of connecting the data silos spread on the web in order to contextualize the data and make it available in a much wider sense.[3] The great value can only become fully effective when linking to external sources becomes a matter of course in the process of publishing data on the web, instead of being an additional and detached activity.

2.2 Role of this Deliverable in the Project

This deliverable describes the approach taken by the AthenaPlus project to encourage the provision, and provide semantically richer metadata to the cultural heritage community, and particularly Europeana, by performing metadata enrichment steps prior to the actual publication of the metadata. It builds on D4.2 Review on Linked Open Data sources for the selection of external sources that will be the targets of the linking process.

Being part of Work package 4 *Terminologies and Semantic enrichment*, it specifically focuses on the use of underlying local terminologies for the metadata enrichment process in order to enhance the quality of enrichment results. Special consideration is given to the question of maintaining the enrichment results for future updates of the metadata and also possibilities to re-use them for newly released or produced metadata.

---

[1] Berners-Lee, Tim. (2006–2009). *Design Issues: Linked Data*. World Wide Web Consortium (W3C). http://www.w3.org/DesignIssues/LinkedData.html

[2] ibid.

[3] All hyperlinks in this document were retrieved between September 28 and November 8, 2015.

3 ATHENAPLUS REVIEW ON LINKED OPEN DATA SOURCES

3.1 Recommendations from the AthenaPlus review

Following the review on Linked Open Data sources as presented in D4.2, the main conclusions are summarized as follows (see section 6 Conclusions):

- The analysis of LOD sources had led to 27 possible data sources as candidates for linking with AthenaPlus content. 18 of these data sources were suggested by partners in the survey, and an additional nine sources were found through exploration of the Datahub[4].

- The selection of candidates has been based on the recommendations and criteria of the W3C, in particular on the five star rating system for LOD publishing. However, the starting point to include only data sources that support a SPARQL endpoint and that are published with a CC0 license could not be maintained: and the criteria were applied softly.

- Since quality criteria turned out to be highly important for the partners the criteria for linking resources must be carefully developed.

- In general, it can be expected that establishing links will be the more reliable the closer the linking process is tied to the source data in terms of its semantics. It was therefore a highly recommended strategy to include links, e.g. URIs for resources in LOD datasets, already in the actual metadata production phase.

In addition, from the First Review of the AthenaPlus project the following feedback was given by the reviewers: “A set of 27 Linked Open Data sources have been identified to establishing links to AthenaPlus vocabularies. However, the project may need to select a smaller number to avoid spending resources on data sources that will not be used for enrichments.” (Technical Review Report 13/05/2015, p.7)

According to these recommendations we have decided to reduce the number of suggested LOD sources, in a first step, by strictly applying the following two criteria to the data sources, as these are crucial not only for the linking process, but also for the re-use of linking results in Europeana and other contexts:

1) Published with an Open License
   The dataset must have a clear license statement which satisfies the Open Definition.[5] This means in particular that one is able to re-use the data also commercially.[6]

2) Supporting RDF and SPARQL, preferably SPARQL 1.1
   The dataset must be available according to “standards from W3C (RDF and SPARQL)”, the fourth star in the five star rating system for LOD publications.

---

[4] The linked open data information site: http://www.datahub.io

[5] See http://opendefinition.org/okd/

[6] For an overview which licenses are considered as open or non-open according to this requirement, see section 4.2. in Deliverable D2.1 “Best Practice Report on Cultural Heritage linked data and metadata standards” of the Linked Heritage project: http://www.linkedheritage.org/getFile.php?id=229

3.2 *Evaluation of LOD Sources: Update*

Building on the overview of suggested LOD source in D4.2, section 5.3 “LOD sources at a glance”, the datasets were revisited and the above criteria applied. Datasets compliant with the criteria are marked in green. In addition, information about the last updates made to a dataset on the *Datahub* was taken as indicator if the dataset is maintained and/or evolving since 2013 when the review for D4.2 was carried out.

No	Name	Link	Description	Licence	Protocols	Compliant with criteria 1) Open license 2) RDF/SPARQL	Last update on *Datahub*
1	Amsterdam Museum	http://semanticweb.cs.vu.nl/lod/am/	The Amsterdam Museum dataset in Europeana Data Model RDF	CC-BY-SA	Dump SPARQL	1) yes 2) yes	Aug 2014
2	BNB (British National Bibliography)	http://datahub.io/dataset/bluk-bnb	British National Bibliography (BNB) published as Linked Data by the British Library	CC0 1.0	Dump SPARQL	1) yes 2) yes	Feb 2014
3	BNE (Spanish National Library)	http://datos.bne.es/	Open bibliographic linked data from the Spanish National Library	CC0 (1.0)	DUMP, SPARQL	1) yes 2) yes	N/A
4	BNF (Bibliothèque nationale de France)	http://data.bnf.fr/semanticweb-en	Open bibliographic linked data from the French National Library	Various (http://data.bnf.fr/licence)	DUMP, RESTful API	1) no 2) no	Apr 2014
5	British Museum Collection	http://collection.britishmuseum.org	Linked Data access to the collection data of the British Museum’s Online Collection	British Museum’s Open Data Licence 1.0	SPARQL	1) yes 2) yes	N/A
6	CLAROS	http://data.clarosnet.org/	CLassical Art Research Online. Service of the universities of Oxford, Cologne, and Paris, for the art of ancient Greece and Rome.	Not specified	SPARQL	1) no 2) yes	N/A

No	Name	Link	Description	Licence	Protocols	Compliant with criteria 1) Open license 2) RDF/SPARQL	Last update on *Datahub*
7	Cultura italia	http://dati.culturaitalia.it/	Portal of Ministero per i Beni e le Attività Culturali. Metadata from museums and other local, regional and national cultural heritage organizations.	CC0 (1.0)	OAI-PMH, SPARQL	1) yes 2) yes	N/A
8	DBpedia	http://dbpedia.org/	Community effort to extract structured information from Wikipedia to make this information available on the Web.	CC-BY-SA 3.0 GNU Free Documentation License	SPARQL	1) yes 2) yes	N/A
9	Dewey Decimal Classification (DDC)	http://dewey.info/	DDC Summaries (the top three levels of the DDC) of Edition 22 in 11 languages and all assignable numbers of Abridged Edition 14 in three languages.	CC-BY-NC 2.0	Dump, SPARQL	1) no 2) yes	N/A
10	DNB (Deutsche Nationalbibliografie)	http://www.dnb.de/EN/nationalbibliografie	The Linked Data Service of the German National Library (Deutsche Nationalbibliothek, DNB)	CC0 1.0	Dump	1) yes 2) no	May 2015
11	Freebase	http://freebase.com/	Freebase is a big collection of structured data (knowledge graph), and a Freebase platform for accessing and manipulating that data via the Freebase API.	CC-BY 2.5	Dump, API	Service closed down	N/A
12	Gencat	http://www20.gencat.cat/portal/site/dadesobertes?newLang=en_GB	Portal of Government of Catalonia. website which publishes public data. from different bodies of the Government.	Various	Dump	1) no 2) no	N/A

No	Name	Link	Description	Licence	Protocols	Compliant with criteria 1) Open license 2) RDF/SPARQL	Last update on *Datahub*
13	GeoNames	http://www.geonames.org/	Geographical database which covers all countries and contains over eight million placenames	CC BY 3.0	Dump RESTful web service	1) yes 2) no	N/A
14	Getty AAT	http://www.getty.edu/research/tools/vocabularies/aat	Structured, multilingual vocabulary including terms, descriptions, and other information for generic concepts related to art, architecture, and other cultural heritage, and conservation.	ODC-BY	Dump, SPARQL	1) yes 2) yes	July 2015
15	Getty TGN	http://www.getty.edu/research/tools/vocabularies/tgn/	Structured vocabulary containing names and other information about places	ODC-BY	Dump, SPARQL	1) yes 2) yes	July 2015
NEW	Getty ULAN	http://www.getty.edu/research/tools/vocabularies/ulan/	Structured vocabulary containing names, biographies, and other information for people and corporate bodies.	ODC-BY	Dump, SPARQL	1) yes 2) yes	July 2015
16	GND	http://www.dnb.de/EN/gnd	Integrated Authority File (GND) of German National Library (Deutsche Nationalbibliothek). Contains data records representing on persons, corporate bodies, congresses, geographic entities	CC0 1.0	Dump SRU	1) yes 2) no	May 2015
17	ICONCLASS	http://www.iconclass.org/help/lod	Multilingual classification system for cultural content	ODbL 1.0	RDF dump, HTTP content negotiation	1) yes 2) no	N/A

No	Name	Link	Description	Licence	Protocols	Compliant with criteria 1) Open license 2) RDF/SPARQL	Last update on *Datahub*
18	LOC Subject Headings	http://id.loc.gov	Authority Files provide authoritative data for subject headings (LCSH) and for names (LCNAF) of persons, organizations, events, places, and titles.	Public Domain	Dump	1) yes 2) no	N/A
19	Muninn World War I	http://rdf.muninn-project.org/	The Muninn Project is a multi-disciplinary, multi-national, academic research project investigating millions of records pertaining to the First World War in archives around the world.	ODC-By 1.0	SPARQL	1) yes 2) yes	Sep 2014
20	NSZL	http://nektar.oszk.hu/wiki/Semantic_web	Authority Files of the Hungarian National Library (National Széchényi Library)	Not specified	Dump SPARQL	1) no 2) yes	N/A
21	Public Library of Veroia	http://libver.math.auth.gr/	Bibliographic Data of the Public Library of Veroia Linked Open Data Project	CC-BY-SA	SPARQL	1) yes 2) yes	Nov 2014
22	RAMEAU subject headings (STITCH)	http://www.cs.vu.nl/STITCH/rameau/	SKOS representation of the RAMEAU book indexing vocabulary, maintained by the French National Library (BnF)	Not specified	Dump	1) no 2) no	N/A
23	ReLoad / LODLAM	http://labs.regesta.com/progettoReload/	Italian Repository for Linked Open Archival Data (Reload)	CC-BY	SPARQL	1) yes 2) yes	Sep 2014

No	Name	Link	Description	Licence	Protocols	Compliant with criteria 1) Open license 2) RDF/SPARQL	Last update on *Datahub*
24	SOCH (K-samsök)	http://www.ksamsok.se/in-english/api/	Portal of Swedish National Heritage Board. Objects harvested from a large number of museums and other local, regional and national cultural heritage organizations.	Other (Open)	SOCH API	1) yes 2) no	N/A
25	VIAF	http://viaf.org/viaf/data/	OCLC dataset and service - built in cooperation with 20 national libraries and other partners - that virtually combines multiple LAM name authority files into a single name authority service	ODC-By 1.0	Dump	1) yes 2) no	June 2015
26	V&A	http://www.vam.ac.uk/api	RESTful interface to the collections of the Victoria and Albert Museum	Other (Non-Commercial)	RESTful API	1) no 2) no	N/A
27	Wikidata	http://www.wikidata.org	Database project of the Wikimedia Foundation to provide support for Wikipedia, Wikimedia projects, and others.	CC0 1.0	Dump, SPARQL	1) yes 2) several, but marked as experimental	May 2015

From the list of 27 candidates gathered in D4.2 there are 11 sources compliant with the criteria of openness and standard-based technologies. One source has been added as it is of obvious interest for the project with a wide range of museum partners: Getty ULAN, the Union List of Artist Names. The reason why it had not been included in the original list was the argument that ULAN is fully covered by VIAF, though VIAF does not comply with the criteria of providing a SPARQL endpoint, ULAN is now evaluated separately.

So in the course of the past two years, as expected, encouraging developments in the field can be observed. Compared to October 2013, four more sources satisfy the criteria. These are the three Getty vocabularies – AAT, TGN, and ULAN – in the meanwhile released as LOD, and an open license statement is now assigned to the Repository Linked Open Archival Data. In the case of Wikidata several SPARQL endpoints are available, but so far their status is explicitly set to experimental. In addition, it can be stated that for most of these sources – 8 out of 12 – the information on the *Datahub* was updated during this period, and many of them even recently, which indicates some level of maintenance of the sources.

4 PREREQUISITES FOR LINKING TO EXTERNAL DATA SOURCES

4.1 Background from Europeana

The aim of linking Europeana data to external data sources is of highly strategic importance for Europeana and has been addressed from various perspectives in several taskforces:

1) Multilingual and Semantic Enrichment Strategy (October 2013 to March 2014)
"*The semantic enrichment of metadata in Europeana is a core concern as it will improve access to the material, define relations among objects and enable multilingual access to content. To ensure enrichments enfold their whole potential and act as facilitators of access, a semantic and multilingual enrichment strategy is needed.*"[7]

2) Metadata quality (January 2014 to May 2015)
"*This Task Force is designed with metadata quality in mind and aims to explore how we can improve the quality of metadata to Europeana, how this provision will improve the end user experience and lastly produce a set of best practice guidelines on metadata quality.*"[8]

3) Evaluation and Enrichments (January 2015 to October 2015)
"*Automatic enrichments can be very beneficial for enabling retrieval across languages and adding context to resources accessible via Europeana.*"[9]

The mission statements of the three taskforces already express the value of semantic enrichment and the significance of metadata quality for positioning Europeana as an important hub in the Linked Open Data cloud. Some important conclusions drawn by the first two taskforces can be summarized as follows:[10]

- Many recommendations target the source metadata and suggest solutions to empower providers to deliver high quality content.[11]

- The use of linked open data vocabularies and the provision of relevant URIs in the metadata submitted to Europeana is highly encouraged, and advice is given for the enrichment of the data prior to submitting it to Europeana in order to provide the best possible data. [12]

4.2 Requirements from Content Providers

The survey on Linked Open Data carried out in July 2013 among AthenaPlus partners and presented in D4.2 included questions about the partners' plans to connect their content with new LOD sources (section 3.4.3) as well as their expectations and criteria (section 3.4.6). The answers are of course an important source for eventually defining the criteria and select the external sources to link to.

To the question "*Does your organisation plan to connect with new LOD sources in the near future?*" 50% of responses, that is 14 partners, expressed the intention to connect with new LOD sources, and out of these positive responses again 50%, i.e. 7 partners intended to connect with the Getty thesauri when they will be published as LOD, which has happened in the meantime. Other LOD sources listed more than once were: DBpedia, Geonames, LOC Subject Headings, SOCH, VIAF, and Wikidata. Furthermore, it was mentioned that sources should be trustworthy and provide a SPARQL endpoint.

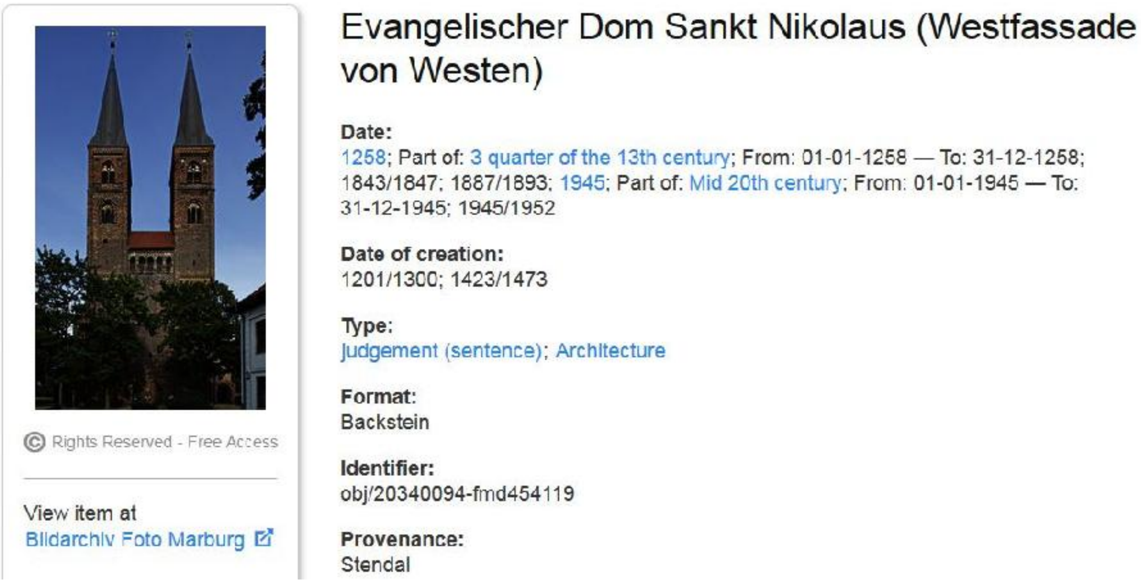
Criteria were listed as follows: need for trustful sources; provider (and data) need to be reliable; relevance, thoroughness and quality of the information; regular updates; popularity; persistent URIs. Also important is the simplicity of the linking process which should be more formalized.

---

[7] http://pro.europeana.eu/get-involved/europeana-tech/europeanatech-task-forces/multilingual-and-semantic-enrichment-strategy
[8] http://pro.europeana.eu/get-involved/our-network/task-forces/metadata-quality
[9] http://pro.europeana.eu/get-involved/europeana-tech/europeanatech-task-forces/evaluation-and-enrichments
[10] The report of the Task Force on Evaluation and Enrichment was published on October 29, 2015, therefore not yet available at the time of writing this deliverable.
[11] Taskforce on a Multilingual and Semantic Enrichment Strategy, Final Report p.30
[12] Taskforce on Metadata Quality, Final Report p.52

Corresponding to the fact that quality criteria are most important it can also be observed that automatic linking results are often not appreciated by the content providers, and they wish to validate the linking results themselves.

The following example may illustrate the problems with automatic enrichment based on string matching only.



Evangelischer Dom Sankt Nikolaus (Westfassade von Westen)

Date:
1258; Part of: 3 quarter of the 13th century; From: 01-01-1258 — To: 31-12-1258; 1843/1847; 1887/1893; 1945; Part of: Mid 20th century; From: 01-01-1945 — To: 31-12-1945; 1945/1952

Date of creation:
1201/1300; 1423/1473

Type:
Judgement (sentence); Architecture

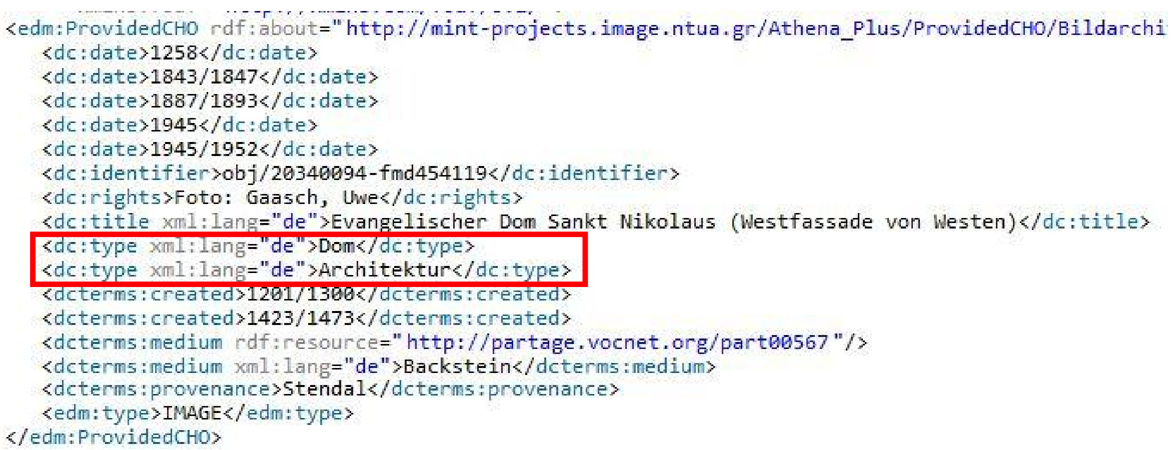
Format:
Backstein

Identifier:
obj/20340094-fmd454119

Provenance:
Stendal

Example for incorrect enrichment result[13]

As type of the provided cultural heritage object which is the cathedral “Sankt Nikolaus” in Stendal, Germany, the term “judgement (sentence)” appears – a user will be surprised and unsatisfied by the apparently wrong metadata delivered to Europeana as there is no evidence of which information is coming from the content provider, and which is automatically enriched. The content provider on the other hand is at least confused if not upset about the change of meaning in his metadata delivered: The EDM record delivered to Europeana in German language contained as type “Dom”, in English: “cathedral”.



```
<edm:ProvidedCHO rdf:about="http://mint-projects.image.ntua.gr/Athena_Plus/ProvidedCHO/Bildarchi
   <dc:date>1258</dc:date>
   <dc:date>1843/1847</dc:date>
   <dc:date>1887/1893</dc:date>
   <dc:date>1945</dc:date>
   <dc:date>1945/1952</dc:date>
   <dc:identifier>obj/20340094-fmd454119</dc:identifier>
   <dc:rights>Foto: Gaasch, Uwe</dc:rights>
   <dc:title xml:lang="de">Evangelischer Dom Sankt Nikolaus (Westfassade von Westen)</dc:title>
   <dc:type xml:lang="de">Dom</dc:type>
   <dc:type xml:lang="de">Architektur</dc:type>
   <dcterms:created>1201/1300</dcterms:created>
   <dcterms:created>1423/1473</dcterms:created>
   <dcterms:medium rdf:resource="http://partage.vocnet.org/part00567"/>
   <dcterms:medium xml:lang="de">Backstein</dcterms:medium>
   <dcterms:provenance>Stendal</dcterms:provenance>
   <edm:type>IMAGE</edm:type>
</edm:ProvidedCHO>
```

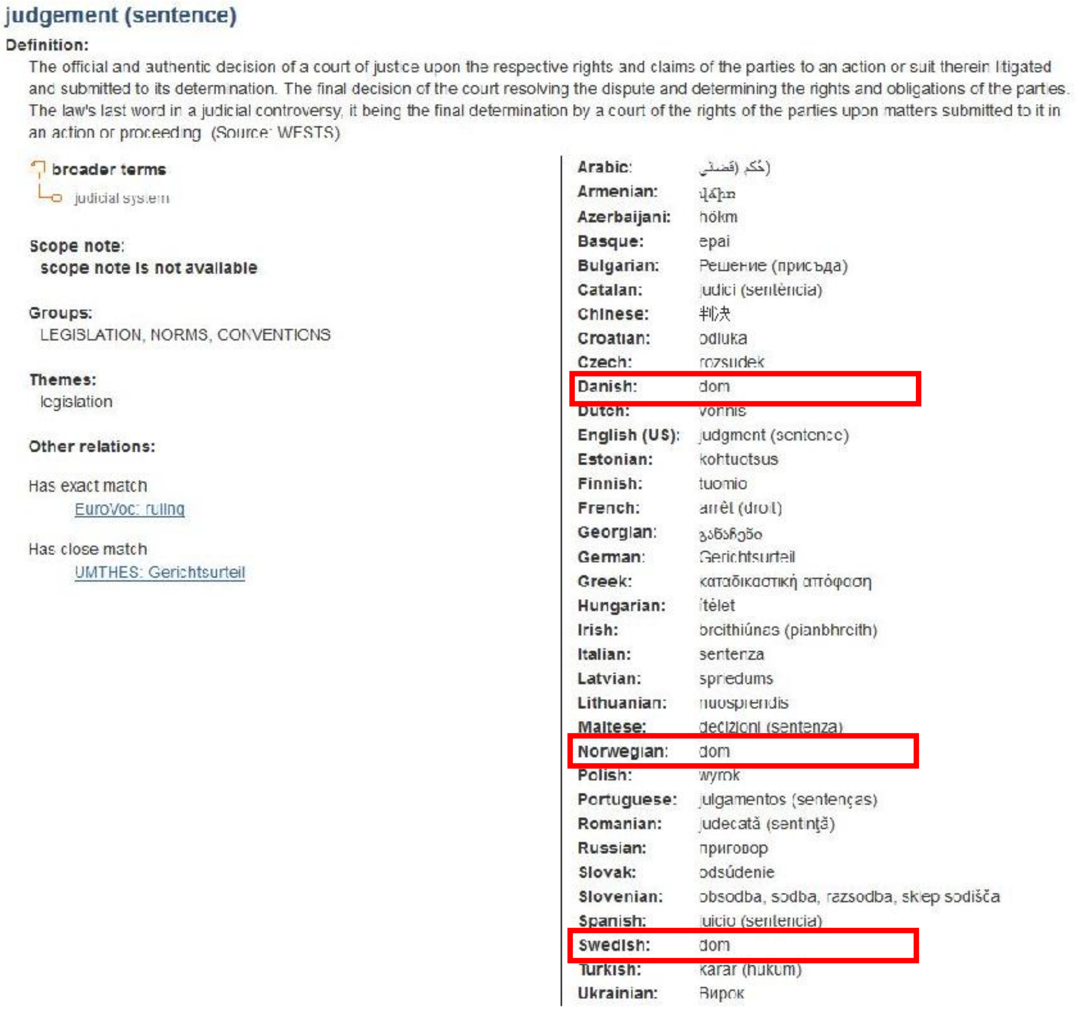
Metadata delivered to Europeana - EDM record of same example

The explanation for the misleading change in type information about the object can be tracked by looking up the record on Europeana’s side. It reveals the enrichment of the record during ingestion with the concept “judgement (sentence)” [14] from the GEMET Thesaurus.

---

[13] See http://www.europeana.eu/portal/record/2048077/Athena_Plus_ProvidedCHO_Bildarchiv_Foto_Marburg_obj_20340094_fmd454119.html

[14] See http://www.eionet.europa.eu/gemet/concept/11556



judgement (sentence)

Definition:

The official and authentic decision of a court of justice upon the respective rights and claims of the parties to an action or suit therein litigated and submitted to its determination. The final decision of the court resolving the dispute and determining the rights and obligations of the parties. The law's last word in a judicial controversy, it being the final determination by a court of the rights of the parties upon matters submitted to it in an action or proceeding. (Source: WESTS)

broader terms
judicial system

Scope note:
scope note is not available

Groups:
LEGISLATION, NORMS, CONVENTIONS

Themes:
legislation

Other relations:

Has exact match
EuroVoc: ruling

Has close match
UMTHES: Gerichtsurteil

Arabic:	حُكم (قضائي)
Armenian:	վճիռ
Azerbaijani:	hökm
Basque:	epai
Bulgarian:	Решение (присъда)
Catalan:	judici (sentència)
Chinese:	判决
Croatian:	odluka
Czech:	rozsudek
Danish:	dom
Dutch:	vonnis
English (US):	judgment (sentence)
Estonian:	kohtuotsus
Finnish:	tuomio
French:	arrêt (droit)
Georgian:	განაჩენი
German:	Gerichtsurteil
Greek:	καταδικαστική απόφαση
Hungarian:	ítélet
Irish:	breithiúnas (pianbhreith)
Italian:	sentenza
Latvian:	spriedums
Lithuanian:	nuosprendis
Maltese:	decizjoni (sentenza)
Norwegian:	dom
Polish:	wyrok
Portuguese:	julgamentos (sentenças)
Romanian:	judecată (sentinţă)
Russian:	приговор
Slovak:	odsúdenie
Slovenian:	obsodba, sodba, razsodba, sklep sodišča
Spanish:	juicio (sentencia)
Swedish:	dom
Turkish:	karar (hukum)
Ukrainian:	Вирок

Concept “judgement (sentence)” from the GEMET Thesaurus

It turns out that the string “dom” has a completely different meaning in German language on one hand and in Scandinavian languages, e.g. Danish, Norwegian, and Swedish, on the other hand. While the English equivalent for the German concept “Dom” will be “Cathedral (building)” – the Getty Art and Architecture Thesaurus provides this term for its respective concept[15] – the English equivalent for “dom” in Danish, Norwegian, and Swedish is indeed “judgement (sentence)”.

It is surprising that for the automatic enrichment apparently not even the language information as given in the provided metadata through language tags (xml:lang=”de”) is taken into account.

Such change of meaning in the metadata obviously not only results in a misleading presentation of the object both in the portal as well as in any other application re-using the Europeana data through their APIs, but also foils the actual intention of the enrichment efforts, which is facilitating discovery and access to the object. It is therefore affecting the quality of Europeana from both the content provider’s and the user’s perspective.

---

[15] See http://vocab.getty.edu/page/aat/300007501

4.3 Use of Underlying Terminologies

Taking into account the considerations of the previous sections, we conclude the following as important factors for a successful enrichment of the AthenaPlus content with links to external sources:

1) The enrichment should be adopted as early as possible in the process of metadata production.
2) The enrichment should be based not only on string matching mechanisms, but rather exploit further contextual information.
3) The enrichment workflow should allow the data providers to validate the enrichment results, and preferably leave the decision to them if the enrichments will be eventually included in the metadata records that are submitted to Europeana.

A promising strategy to achieve more reliable results in enrichment processes is, instead of relying only on string matching mechanisms, the use of underlying terminologies in the matching rules.

The actual metadata delivered to Europeana is typically derived, often with some intermediate steps of exporting and transforming the data, from descriptions of the cultural heritage object held in documentation systems of the data providers. In the case of museum or museum-like content which is particularly represented in the AthenaPlus project, the collection management systems will often support locally maintained controlled vocabularies and authorities. However, when exporting the metadata about a cultural heritage object for submission to Europeana, the contextual information about referenced entities like concepts, agents, or places present in the local system is usually not included in the records.

It is therefore a valid strategy to focus on the underlying terminologies themselves and perform enrichments directly on the local terminologies by linking these sources to external sources. The ultimate goal of this strategy is to feed back the links into the local vocabulary and authority records of the collection management system, thereby achieving several positive effects at the same time:

- The approach gives full control over the enrichment to the data provider who can decide if and to what extent, according to their own evaluation of the enrichment results, they wish to integrate the links.
- The links are integrated at the source of the metadata production. Hence, the enrichment results will be kept in the most sustainable way and can give the maximum possible coverage.
- The links will enrich both existing cultural heritage object descriptions, and newly produced or updated descriptions in the daily work of the data provider. Thereby any content update, as well as new content delivery to Europeana, will benefit from the enrichment.
- Finally, any further enrichment of the cultural heritage object metadata on Europeana’s side is supported in the best possible way as the links to external sources delivered with the metadata can subsequently be used for such additional enrichment processes, by providing context information which is accessible according to LOD principles.

An important precondition for following this strategy is obviously the definition of a workflow that can easily be followed by the data providers. They will need to provide their terminology records as a data dump in a predefined format, and be able to enrich in return their records with the linking results, preferably in an automated way.

5 PRESENTATION OF SELECTED EXTERNAL DATA SOURCES

5.1 *Selection Criteria*

The re-evaluation of the list of candidate LOD sources suggested in D4.2 Review on Linked Open Data sources as presented in section 3.2 led to a reduced list of 12 sources. This is still a large number of sources, covering a wide range of different scopes, so we developed further criteria for the final selection of sources, based on the prerequisites discussed in the previous section.

As we aim at underlying terminologies for the linking process an important criteria is obviously for which metadata elements usually present in the records submitted to Europeana such terminologies are available.

The relevant entities, e.g. EDM classes, are: agents (edm:Agent), concepts (skos:Concept), cultural heritage objects (edm:ProvidedCHO), places (edm:Place), time spans (edm:TimeSpan).

The following considerations are taken into account:

- Underlying terminologies are mostly available for agents, concepts, and places, but not so often for time spans. The cultural heritage objects are the focus in itself.
- Matching agents and places is more promising than concepts or cultural heritage objects because it is fairly independent of language issues, and information to be used in the linking process is usually more standardized than for other entities.
- A provider’s local agent authorities will often include: Preferred name, alternative names, gender, life dates, nationality.
- A provider’s local place authorities will often include: Preferred name, alternative names, country, continent, and further part-of-place relationships.

Therefore the expected quality of linking results will be the most promising for matching agents and places, and these can particularly provide relevant context to support further co-referencing efforts for the cultural heritage object descriptions. Also they are particularly relevant for the discovery of the cultural heritage objects and their re-contextualization with objects from other collections.

5.2 *Presentation of Selected Sources*

5.2.1 *Final Selection Process*

The list of 12 candidate sources has finally been assessed with the above criteria.

No	Name	Scope	Plans to connect with by AthenaPlus partners[16]
1	Amsterdam Museum	Cultural Heritage objects	No
2	BNB (British National Bibliography)	Bibliographic data	No

---

[16] See section 4.2 in this deliverable – according to results from partner survey presented in D4.2 Review on Linked Open Data sources, p.10

No	Name	Scope	Plans to connect with by AthenaPlus partners[16]
3	BNE (Spanish National Library)	Bibliographic data	No
5	British Museum Collection	Cultural Heritage objects	Yes, one partner
7	Cultura italia	Cultural Heritage objects Concepts	No
8	DBpedia	Not restricted	Yes, more than one partner
14	Getty AAT	Concepts	Yes, more than one partner
15	Getty TGN	Places	Yes, more than one partner
NEW	Getty ULAN	Agents	Yes, more than one partner
19	Muninn World War I	Archival data	No
21	Public Library of Veroia	Bibliographic data	No
23	ReLoad / LODLAM	Archival data	No

The Getty vocabularies turn out to be of primary relevance in terms of all different aspects. Many partners are particularly aiming at the Getty vocabularies as they are domain-specific sources, and furthermore are considered a most reliable source in terms of quality and sustainability. DBpedia is also highly relevant, but will not be included in the linking process: In contrast to the Getty vocabularies its broad scope does not specifically match with the museum-like content as focused from AthenaPlus. Also, Europeana is already performing automatic enrichments with DBpedia, but not with Getty vocabularies, though Europeana explicitly recommends them as target source for enrichments.

5.2.2 *Getty Union List of Artist Names (ULAN)*

Name	The Union List of Artist Names
Author	The Getty Research Institute
Maintainer	Joan Cobb
Link LOD source	http://vocab.getty.edu/ulan/
Source	http://www.getty.edu/research/tools/vocabularies/ulan/index.html

Datahub link	http://datahub.io/dataset/getty-ulan
Description of Content	The Union List of Artist Names® (ULAN) is a structured vocabulary, including names, biographies, and other information for people and corporate bodies related to the design, creation, patronage, collection, conservation, and maintenance of art, architecture, and other cultural materials about artists and architects. Like all of the Getty Vocabularies, the ULAN is compliant with international standards and grows through contribution.
Amount of resources	Around 230,000 agents (persons and groups of persons)
Licence	Open Data Commons Attribution License http://opendefinition.org/licenses/odc-by/
Protocols supported	Dump, SPARQL

5.2.3 *Getty Thesaurus of Geographic Names (TGN)*

Name	The Getty Thesaurus of Geographic Names
Author	The Getty Research Institute
Maintainer	Joan Cobb
Link LOD source	http://vocab.getty.edu/tgn/
Source	http://www.getty.edu/research/tools/vocabularies/tgn/about.html
Datahub link	http://datahub.io/dataset/getty-tgn
Description of Content	Getty Thesaurus of Geographic Names® (TGN) is a structured vocabulary of geographic names intended to provide terminology and other information important to various disciplines that specialize in art, architecture and material culture. The TGN includes names and associated information about places. Places in TGN include administrative political entities (e.g., cities, nations) and physical features (e.g., mountains, rivers). Current and historical places are included. Like all of the Getty Vocabularies, the TGN is compliant with international standards and grows through contribution.
Amount of resources	Around 1,430,000 places
Licence	Open Data Commons Attribution License http://opendefinition.org/licenses/odc-by/
Protocols supported	Dump, SPARQL

6 LINKING ATHENAPLUS CONTENT

6.1 Web Services for linking to ULAN and TGN vocabularies

This section summarizes the functional and technical requirements of the web-service implemented for linking providers' terms to ULAN and TGN. More specifically the scenario employed is as follows. A content provider holds values in his content management system about places and persons and these values can be linked to ULAN and TGN. An export of these values is first made and then using the web service implemented by NTUA, he gets the possible links together with a confidence degree. The next step is to validate the links by picking up the one that is correct and finally by integrating it into his metadata for providing it to Europeana. The following sections describe the data models used for the linking services together with the evaluation of the linking services on the datasets used.

6.1.1 *Data model definition for services*

The validity of linking or resource discovery strongly depends on the information that the data source holds for the resource and also on the information the provider holds about the entity he wishes to link or enrich. So for example the Union List of Artist Names holds information about artists that includes their name and alternative names in many languages, biographical information, events in which the artist participated in and other useful information. On the other hand what a provider holds about an artist in his content management system can vary a lot, being even more or much less expressive than the information provided in ULAN. So amending limited information on providers' side can be, in addition to linking to an external resource, an additional reason why resource discovery is performed – that is enriching providers' metadata with the information hold on domain specific data sources like ULAN and TGN.

Therefore the first step for implementing a web service for linking to ULAN and TGN was to come up with the data models that could sufficiently cover the type of resources to be linked. Ideally the information an organisation holds on an authority includes the following:

- **prefName [1...1]:** The preferred name of the artist. This string value is usually comprised of the first name (first letter in capital) followed by the surname (first letter in capital). However a middle name may sometimes exists and the order the first name and surname may be reversed.
- **altName [0...n]:** The alternative names of the artist. The format and syntax of the values is similar to the prefName. The main difference in this case is the cardinality is not fixed to 1 but to n.
- **gender [0...1]:** The gender of the authority.
- **birthYear [0...n]:** The birth year of the authority. Strangely the cardinality of this value is not restricted to 1 and the reason for this is that sometimes there is no exact year of birth but possible estimations. Another problem with this value is its format and syntax. Different formats for dates exist and since most of can be applicable a normalization will be required.
- **deathYear [0...n]:** The death year defined in similar way as birthYear.

A sample input JSON file for the ULAN service can be found in Appendix 1. Similarly the information an organisation holds about a place includes the following:

- **prefName [1...1]**: The preferred name of the place. Normally the preferred label should be in English. Otherwise the preferred label can be one of the alternative labels.
- **altName [0...n]**: The alternative names of the place.
- **nation [0...1]**: The nation of the place.
- **continent [0...1]**: The continent of the place.
- **partOfPlace [0...n]**: A place of which the given place is part of.

As shown from the way the data model for linking has been defined, it is in both cases required by the prefName as the minimum information while in case providers hold and can provide additional things then they are encouraged to do so since a better result will be established.

6.1.2 *Approach and evaluation functions*

Getty provides a SPARQL endpoint for both ULAN and TGN[17]. However the direct use of this endpoint for the link discovery process was not possible for various reasons. First because the Getty server has been set up in a way that does not support the use of the SPARQL endpoint from external services and secondly because even this was possible SPAQRL provides quite limited functionalities on string matching. Therefore we employed the following approach. SPARQL endpoints have been set up for the ULAN and TGN vocabularies on a local server (Getty provides dump files for both in RDF[18]). After than we extracted the preferred and alternative names in all languages for all the terms and we indexed them using Lucene[19] (two separated indexes have been create one for places and one for artists – the key was the Getty identifier and the value the pref/alt name). Once the indexes where there, for each entry the following have been performed

1. The preferred and alternative names were examined for a fuzzy string searching with a degree of confidence 0.85
2. SPARQL queries then performed for the obtained results for comparing the additional information of the input with this of the output.
3. The following evaluation functions were employed for estimating the confidence.

    a. **ULAN**:
    date_distance(x, y) = 4/(3π)*atan(abs(x-y)/40 - 1) + 1/3

    d1 = cosine_distance(name, ulan_name)
    d2 = date_distance(birthyear, ulan_birthyear)
    d3 = date_distance(deathyear, ulan_deathyear)

    If some of the dates are not provided the respective distance is set to 0.5

    The confidence score is computed as follows:
    confidence = (d1 + ⅔*min(d2,d3)+⅓*max(d2,d3))/2

    Only results with confidence > 0.5 are kept.

    b. **TGN**:

    d1 = 2gram cosine_distance (name, tgn_name)
    d2 = 2gram cosine_distance (nation, tgn_nation)
    d3 = 2gram cosine_distance (continent, tgn_continent)
    d4 = 2gram cosine_distance (partOfPlace, tgn_partOfPlace)

    The confidence score is computed as follows:
    confidence = 1- (⅔*d1 + 1/9*d2 + 1/9 * d3 +1/9 * d4)

    Only results with confidence > 0.5 are kept.

6.1.3 *Use of the services*

The services are made available and can be accessed by performing a post call to http://zenon.image.ece.ntua.gr/fres/service/ulan

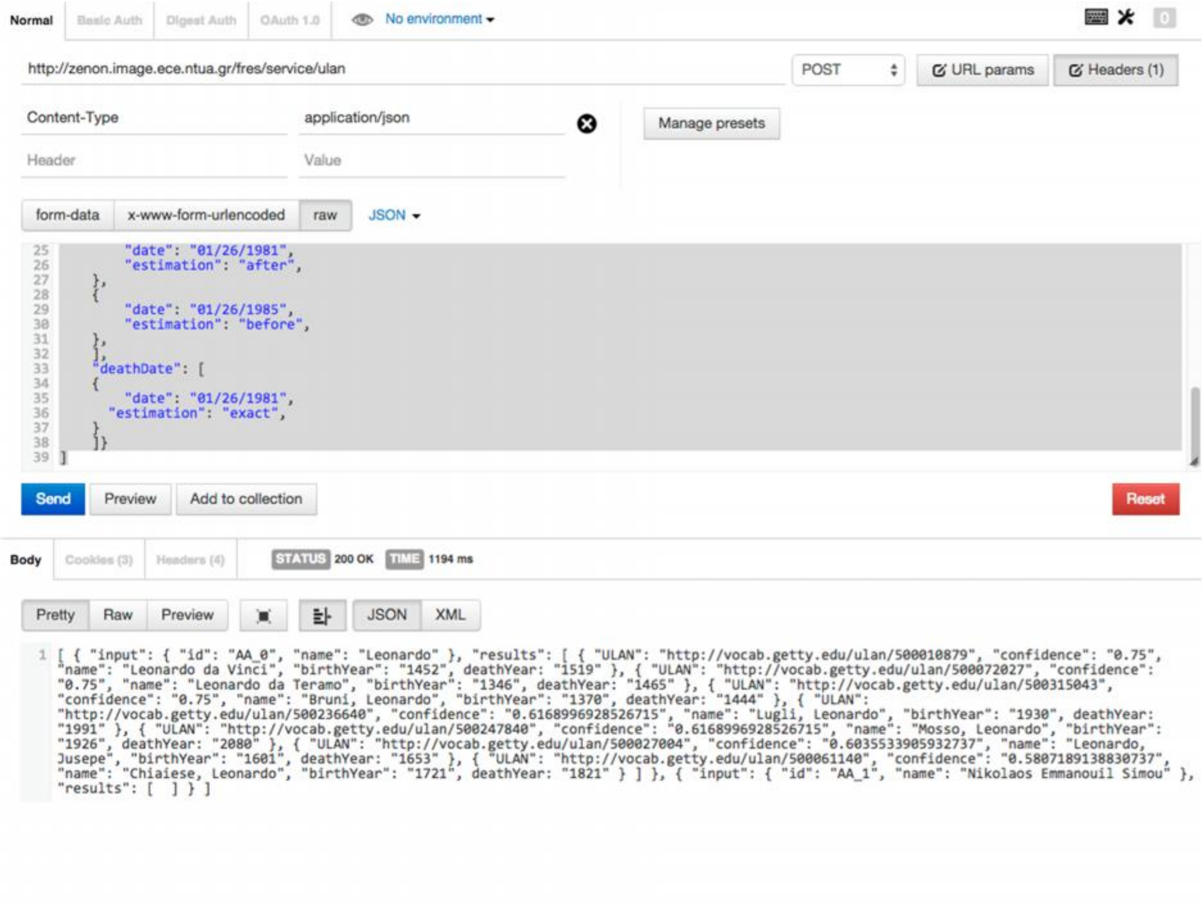
---

[17] http://vocab.getty.edu/sparql
[18] http://vocab.getty.edu/
[19] https://lucene.apache.org/core/

and

http://zenon.image.ece.ntua.gr/fres/service/tgn

for ULAN and TGN respectively. The Content-type has to be set to application/json and the JSON format that can be found in Appendix 1 and 2 should be used. A sample JSON output file can be found at Appendix 3 while the following figure shows how the ULAN linking service can be used by the Google Chrome application POSTMAN rest client plugin[20].


Normal
Basic Auth
Digest Auth
OAuth 1.0
No environment
POST
URL params
Headers (1)
Manage presets
form-data
x-www-form-urlencoded
raw
JSON
Send
Preview
Add to collection
Reset
Body
Cookies (3)
Headers (4)
Pretty
Raw
Preview
JSON
XML

http://zenon.image.ece.ntua.gr/fres/service/ulan

Content-Type: application/json

```
            "date": "01/26/1981",
            "estimation": "after",
        },
        {
            "date": "01/26/1985",
            "estimation": "before",
        },
        ],
        "deathDate": [
        {
            "date": "01/26/1981",
          "estimation": "exact",
        }
        ]}
]
```

STATUS 200 OK TIME 1194 ms

```
[ { "input": { "id": "AA_0", "name": "Leonardo" }, "results": [ { "ULAN": "http://vocab.getty.edu/ulan/500010879", "confidence": "0.75",
"name": "Leonardo da Vinci", "birthYear": "1452", deathYear: "1519" }, { "ULAN": "http://vocab.getty.edu/ulan/500072027", "confidence":
"0.75", "name": "Leonardo da Teramo", "birthYear": "1346", deathYear: "1465" }, { "ULAN": "http://vocab.getty.edu/ulan/500315043",
"confidence": "0.75", "name": "Bruni, Leonardo", "birthYear": "1370", deathYear: "1444" }, { "ULAN":
"http://vocab.getty.edu/ulan/500236640", "confidence": "0.6168996928526715", "name": "Lugli, Leonardo", "birthYear": "1930", deathYear:
"1991" }, { "ULAN": "http://vocab.getty.edu/ulan/500247840", "confidence": "0.6168996928526715", "name": "Mosso, Leonardo", "birthYear":
"1926", deathYear: "2080" }, { "ULAN": "http://vocab.getty.edu/ulan/500027004", "confidence": "0.6035533905932737", "name": "Leonardo,
Jusepe", "birthYear": "1601", deathYear: "1653" }, { "ULAN": "http://vocab.getty.edu/ulan/500061140", "confidence": "0.5807189138830737",
"name": "Chiaiese, Leonardo", "birthYear": "1721", deathYear: "1821" } ] }, { "input": { "id": "AA_1", "name": "Nikolaos Emmanouil Simou" },
"results": [  ] } ]
```

Figure 1: Accessing the ULAN service using postman

---

[20] https://github.com/leafwind/POSTMan-Chrome-Extension

6.2 *Evaluation of the linking process*

The web services described in the previous section were applied to the following datasets:

For ULAN:
1) Local agent authority of Nationalmuseum Sweden: Containing 930 resources.
2) Local agent authority of Philipps-Universitaet Marburg, Bildarchiv Foto Marburg (Germany): Containing 95,756 resources.

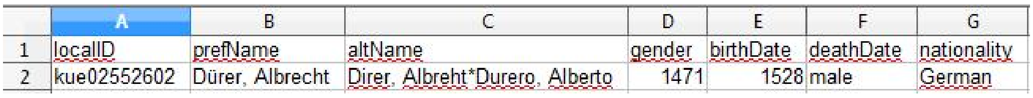
For TGN:
3) Local place authority of Philipps-Universitaet Marburg, Bildarchiv Foto Marburg (Germany): Containing 50,412 resources.

Nationalmuseum Sweden provided through AthenaPlus 73,999 items to Europeana, Bildarchiv Foto Marburg 187,153 items (including their partner institutions). In addition, 765,755 items earlier provided by Bildarchiv Foto Marburg through the Athena project were updated. So from the AthenaPlus content 261,152 items, and in total 1,026,907 items in Europeana are potentially enriched with links to ULAN and TGN, by adding the links from the underlying terminologies to the object descriptions. In case of object descriptions referencing multiple agents or multiple places, this will of course also result in multiple added links.

First, we look at the steps to be followed by the data provider in the workflow:

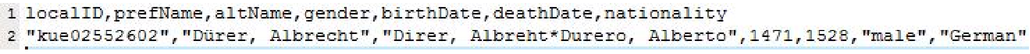
Step 1) The source data from a local agent or place authority has to be provided in a CSV file according to the following pattern:

For ULAN:



	A	B	C	D	E	F	G
1	localID	prefName	altName	gender	birthDate	deathDate	nationality
2	kue02552602	Dürer, Albrecht	Direr, Albreht*Durero, Alberto	1471	1528	male	German

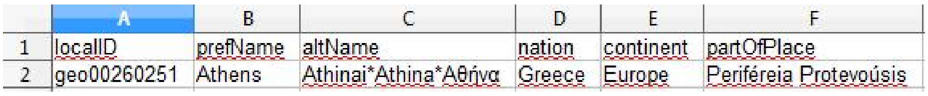
Input file for linking to ULAN as spreadsheet



```
localID,prefName,altName,gender,birthDate,deathDate,nationality
"kue02552602","Dürer, Albrecht","Direr, Albreht*Durero, Alberto",1471,1528,"male","German"
```

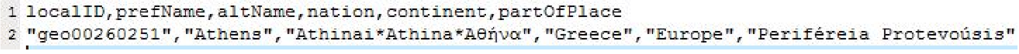
Input file for linking to ULAN as CSV

For TGN:



	A	B	C	D	E	F
1	localID	prefName	altName	nation	continent	partOfPlace
2	geo00260251	Athens	Athinai*Athina*Αθήνα	Greece	Europe	Periféreia Protevoúsis

Input file for linking to TGN as spreadsheet



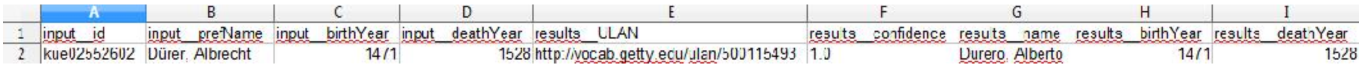
```
localID,prefName,altName,nation,continent,partOfPlace
"geo00260251","Athens","Athinai*Athina*Αθήνα","Greece","Europe","Periféreia Protevoúsis"
```

Input file for linking to TGN as CSV

The input format is for both ULAN and TGN very simple, and most collection management systems will provide export functionalities to satisfy the format.

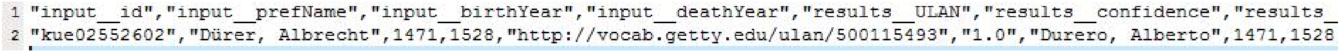
Step 2) The result files delivered by the web services converted from JSON to CSV will look like

For ULAN:



	A	B	C	D	E	F	G	H	I
1	input__id	input__prefName	input__birthYear	input__deathYear	results__ULAN	results__confidence	results__name	results__birthYear	results__deathYear
2	kue02552602	Dürer, Albrecht	1471	1528	http://vocab.getty.edu/ulan/500115493	1.0	Durero, Alberto	1471	1528

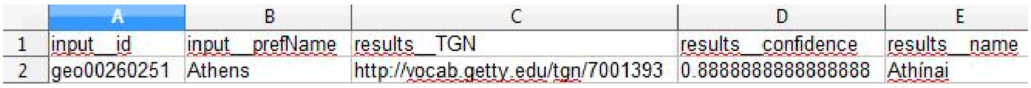
Result file for linking to ULAN as spreadsheet



```
"input__id","input__prefName","input__birthYear","input__deathYear","results__ULAN","results__confidence","results_
"kue02552602","Dürer, Albrecht",1471,1528,"http://vocab.getty.edu/ulan/500115493","1.0","Durero, Alberto",1471,1528
```

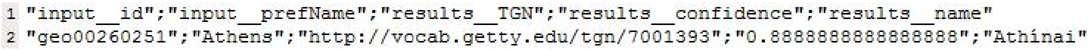
Result file for linking to ULAN as CSV

For TGN:



	A	B	C	D	E
1	input__id	input__prefName	results__TGN	results__confidence	results__name
2	geo00260251	Athens	http://vocab.getty.edu/tgn/7001393	0.8888888888888888	Athínai

Result file for linking to TGN as spreadsheet



```
"input__id";"input__prefName";"results__TGN";"results__confidence";"results__name"
"geo00260251";"Athens";"http://vocab.getty.edu/tgn/7001393";"0.8888888888888888";"Athínai"
```

Result file for linking to TGN as CSV

Step 3) The data provider will need to evaluate the linking results and decide if and to which extent he wishes to feed them back into the local authority records, with the ultimate goal to further enrich the object descriptions which reference the authority records.

In the next section we present the evaluation and observations from the two data providers participating in the linking.

6.2.1 *Linking to ULAN vocabulary*

Local agent authority of Nationalmuseum Sweden

For Nationalmuseum Sweden results of the linking process can be summarized as follows – out of the 930 resources included in the matching, for 615, i.e. 66.1 % of the resources one or more ULAN resources were returned as matching candidate, here is the overview of the evaluation:

Confidence c	Number of resources		Of those manually checked		Of those are correct		Of those are incorrect/ need more research	
c = 1.0	212	22,8 %	20	9,4 %	20	100 %	0	0 %
0.9 < c < 1.0	218	23,4 %	23	10,6 %	21	91,3 %	2 / 0	8,7 %
0.8 < c <= 0.9	52	5,6 %	9	17,3 %	9	100 %	0	0 %
0.7 < c <= 0.8	69	7,4 %	9	13,0 %	7	77,8 %	0 / 2	22,2 %
0.6 < c <= 0.7	31	3,3 %	n/a	n/a	n/a	n/a	n/a	n/a
0.5 < c <= 0.6	33	3,5 %	n/a	n/a	n/a	n/a	n/a	n/a
No match	315	33,9 %	n/a	n/a	n/a	n/a	n/a	n/a
Total	**930**	**100 %**						

Overview of matching results for Nationalmuseum Sweden.

Most of the matches with a confidence value <0.7 return more than one hit which will need additional evaluation. Of the 615 hits 212 had a duplicate hit, of which 12 were manually tested: 5 were incorrect and 7 correct. Of the correct hits it was always the one with the higher confidence which was right.

In total more than 12 % of the resources having matches were manually checked and the Nationalmuseum Sweden ends up with the conclusion that they will feed back into their records automatically only those matches with confidence value =1.0 while performing manual checking for all other resources. In this opportunity they would also amend their local records with information from ULAN since they suspect that low confidence values may also result from wrong or missing data in their source records.

6.2.1.2 Local agent authority for people related to art and architecture, Bildarchiv Foto Marburg

For Bildarchiv Foto Marburg quantitative results of the linking process can be summarized as follows – out of the 95,756 resources included in the matching, for 40,539, i.e. 42,3 % of the resources one or more ULAN resources were returned as matching candidate. Among those with matches 77,4 % had one match and 22,6 % had between two and five matches in ULAN.

Matches	Number of input resources	
1	31,357	32,75 %
2	3,642	3,80 %
3	1,385	1,45 %
4	708	0,74 %
5	3,447	3,60 %
No match	55,217	57,66 %
Total	**95,756**	**100 %**

Distribution of number of ULAN matches for Bildarchiv Foto Marburg

For the resources with one match in ULAN an analysis of 419 matching results was performed with the following distribution:

Confidence c	Number of resources		Manually checked		Of those are correct		Of those are incorrect		Need more research	
c = 1.0	10,227	32,61 %	68	0,66 %	68	100 %	0	0 %	0	0 %
0.9 < c < 1.0	9,203	29,35 %	153	1,66 %	148	96,7 %	0	0 %	5	3,3 %
0.8 < c <= 0.9	5,648	18,01 %	94	1,66 %	87	92,6 %	0	0 %	7	7,4 %
0.7 < c <= 0.8	2,222	7,09 %	37	1,67 %	11	29,7 %	8	21,6 %	18	48,6 %
0.6 < c <= 0.7	1,880	6,00 %	31	1,65 %	3	9,7 %	17	54,8 %	11	35,5 %
0.5 < c <= 0.6	2,177	6,94 %	36	1,65 %	2	5,6 %	28	77,8 %	6	16,7 %
Total	**31,357**	100 %	419	**1,34 %**						

Overview of matching results for Bildarchiv Foto Marburg for resources with one ULAN match

One can observe a remarkably clear shift in the validity of results between confidence > 0.8 and < 0.8. Since none of the manually checked match with confidence 0.8 < c < 1.0 was incorrect, and only a minor number of matches remain ambiguous, for the 25,075 resources in this group the local authority records are automatically enriched with the ULAN match.

For the resources with more than one match in ULAN again an analysis was performed with the following approach: Each of the matches has been validated separately, and it was checked if the correct match was the one with the highest confidence value. The assumption of choosing from multiple matches the one with the highest confidence value proved true for all cases that were manually checked.

Number of matches					Match with highest confidence					
Confidence c	Number of resources		Manually checked		correct		incorrect		Need more research	
Two matches										
0.8 < c <= 1.0	2,018	55,4 %	26	1,3 %	25	96,2 %	0	0 %	1	3,8 %
0.5 < c <= 0.8	1,624	44,6 %	22	1,3 %	0	0 %	14	63,6 %	8	36,4 %
Total	3,642	100 %								
Three matches										
0.8 < c <= 1.0	531	38,3 %	7	1,3 %	5	71,4 %	1	14,3 %	1	14,3 %
0.5 < c <= 0.8	854	61,7 %	12	1,4 %	0	0 %	10	83,3 %	2	16,7 %
Total	1,385	100 %								
Four matches										
0.8 < c <= 1.0	231	32,6 %	3	1,3 %	2	66,7 %	0	0 %	1	33,3 %
0.5 < c <= 0.8	477	67,4 %	6	1,3 %	0	0 %	2	33,3 %	4	66,7 %
Total	708	100 %								
Five matches										
0.8 < c <= 1.0	608	17,6 %	8	1,3 %	3	37,5 %	2	25,0 %	3	37,5 %
0.5 < c <= 0.8	2,839	82,4 %	36	1,3 %	0	0 %	25	69,4 %	11	30,6 %
Total	3,447	100 %								

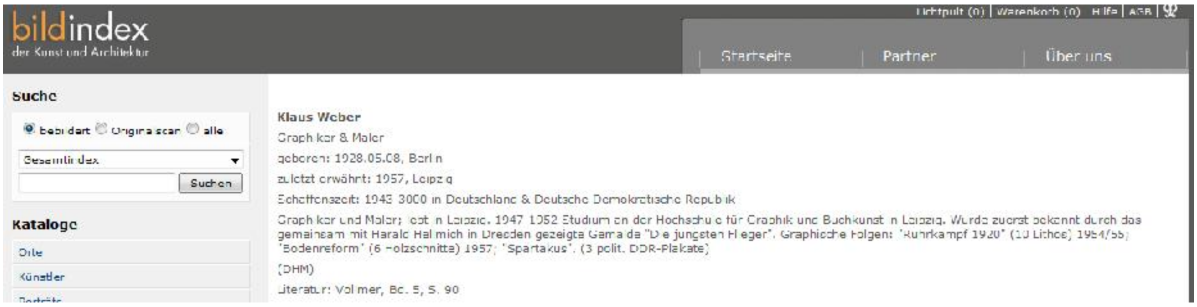
Overview of results for Bildarchiv Foto Marburg for resources with two to five ULAN matches

Again, in the case of two matches the ones with confidence 0.8 < c < 1.0 were always either correct or maybe ambiguous, but never definitely incorrect so the additional 2,018 resources from this group are also enriched automatically in the local authority.

Both persons working on the manual checking reported independently from each other that in fact the source records for the groups with confidence < 0.8 were lacking sufficient contextual information for actually identifying the person. E.g. there are typically no actual life dates known, and these are only estimated from the activity dates. The less contextual information is available, the more matches will be returned by the service, with lower confidence values – as one can see from the group of resources with five matches of which only 17,6 % returned a confidence 0.8 < c < 1.0.

In such cases indeed any link established would remain ambiguous because it is not determinable if it is the same person, or more additional context would be needed, e.g. if the cultural heritage objects referencing this person are of the same type of object which matches the profession(s) of the person.

We finally add two examples which highlight why simple string matching of names will not return sufficiently reliable results even in the case of such a focused and domain-specific target as ULAN is.


bildindex
der Kunst und Architektur
Lichtpult (0) | Warenkorb (0) Hilfe | AGB
Startseite
Partner
Über uns
Suche
● bebildert ○ Originalscan ○ alle
Gesamtindex
Suchen
Kataloge
Orte
Künstler
Porträts

Klaus Weber

Graphiker & Maler

geboren: 1928.05.08, Berlin

zuletzt erwähnt: 1957, Leipzig

Schaffenszeit: 1943-3000 in Deutschland & Deutsche Demokratische Republik

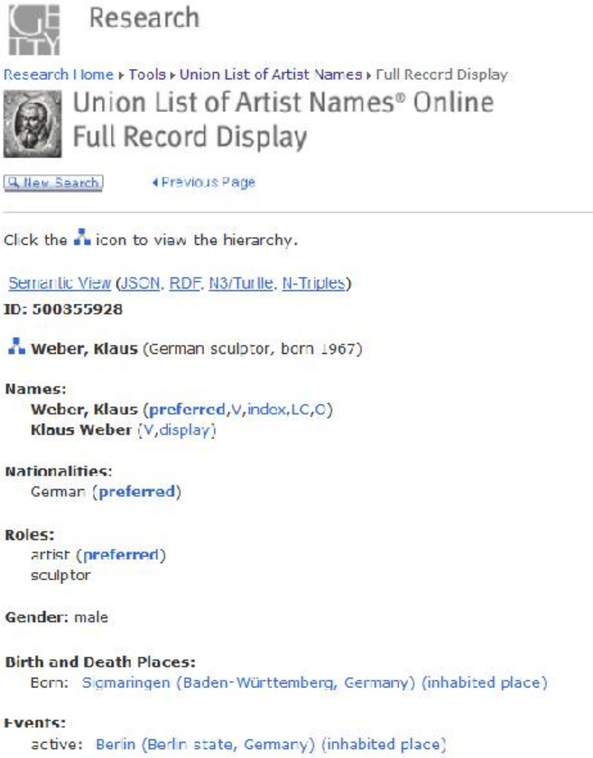
Graphiker und Maler; lebt in Leipzig. 1947-1952 Studium an der Hochschule für Graphik und Buchkunst in Leipzig. Wurde zuerst bekannt durch das gemeinsam mit Harald Hellmich in Dresden gezeigte Gemälde "Die jungsten Flieger". Graphische Folgen: 'Ruhrkampf 1920" (10 Lithos) 1954/55; 'Bodenreform' (6 Holzschnitte) 1957; 'Spartakus'. (3 polit. DDR-Plakate)

(DHM)

Literatur: Vollmer, Bd. 5, S. 90

Bildarchiv Foto Marburg local resource "Weber, Klaus"[21]

The name "Weber, Klaus" matches in both resources exactly with the preferred name, and the name is unique in both resources of the source and the target – however, the resources describe definitely different persons: The one in the Bildarchiv Foto Marburg resource is born 1928 in Berlin (Germany) and lastly mentioned in 1957 in Leipzig, while the one in the ULAN resource is born 1967 in Sigmaringen (Baden-Württemberg, Germany).


Research
Research Home ▸ Tools ▸ Union List of Artist Names ▸ Full Record Display
Union List of Artist Names® Online
Full Record Display
New Search
◂ Previous Page

Click the icon to view the hierarchy.

Semantic View (JSON, RDF, N3/Turtle, N-Triples)

ID: 500355928

Weber, Klaus (German sculptor, born 1967)

Names:
Weber, Klaus (**preferred**,V,index,LC,O)
Klaus Weber (V,display)

Nationalities:
German (**preferred**)

Roles:
artist (**preferred**)
sculptor

Gender: male

Birth and Death Places:
Born: Sigmaringen (Baden-Württemberg, Germany) (inhabited place)

Events:
active: Berlin (Berlin state, Germany) (inhabited place)

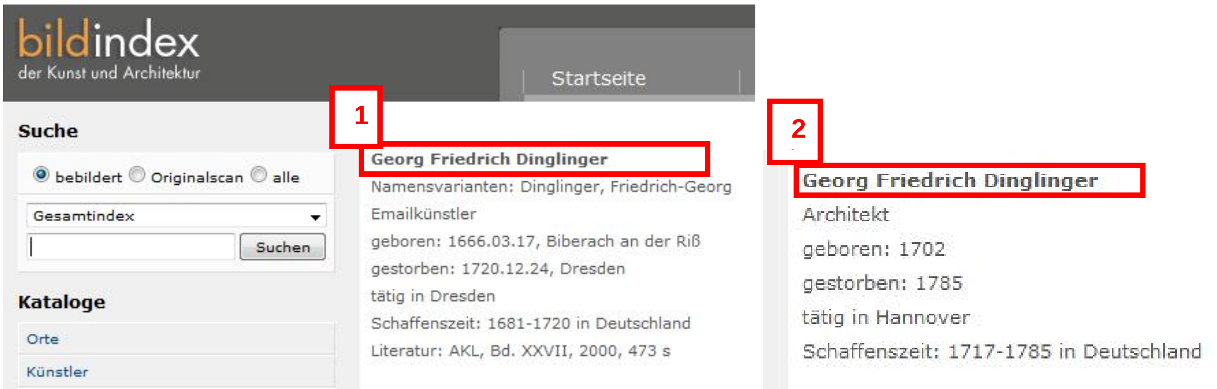
ULAN resource "Weber, Klaus"[22]

The following example is even more challenging: Beside the fact of exact string matching of names with a unique hit in the target, it turns out that there are two resources in the source authority describing different persons.

---

[21] http://www.bildindex.de/dokumente/html/kue03191094

[22] http://vocab.getty.edu/page/ulan/500355928

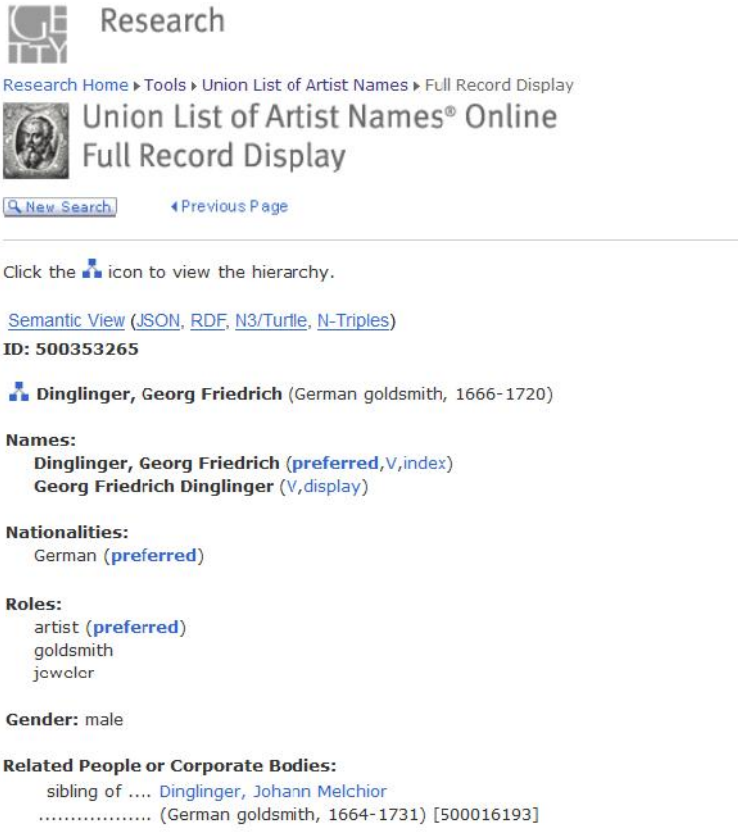

bildindex
der Kunst und Architektur
Startseite
Suche
● bebildert ○ Originalscan ○ alle
Gesamtindex
Suchen
Kataloge
Orte
Künstler
1
2

Georg Friedrich Dinglinger
Namensvarianten: Dinglinger, Friedrich-Georg
Emailkünstler
geboren: 1666.03.17, Biberach an der Riß
gestorben: 1720.12.24, Dresden
tätig in Dresden
Schaffenszeit: 1681-1720 in Deutschland
Literatur: AKL, Bd. XXVII, 2000, 473 s

Georg Friedrich Dinglinger
Architekt
geboren: 1702
gestorben: 1785
tätig in Hannover
Schaffenszeit: 1717-1785 in Deutschland

Bildarchiv Foto Marburg local resources
1) "Dinglinger, Georg Friedrich", born 1666[23]
2) "Dinglinger, Georg Friedrich", born 1702[24]

It is therefore not only necessary to take into account multiple matches for one source resource, but also to check if one match in the target has multiple source resources.


Research
Research Home ▸ Tools ▸ Union List of Artist Names ▸ Full Record Display
Union List of Artist Names® Online
Full Record Display
New Search
◂Previous Page

Click the icon to view the hierarchy.

Semantic View (JSON, RDF, N3/Turtle, N-Triples)
ID: 500353265

Dinglinger, Georg Friedrich (German goldsmith, 1666-1720)

Names:
Dinglinger, Georg Friedrich (**preferred**,V,index)
Georg Friedrich Dinglinger (V,display)

Nationalities:
German (**preferred**)

Roles:
artist (**preferred**)
goldsmith
jeweler

Gender: male

Related People or Corporate Bodies:
sibling of .... Dinglinger, Johann Melchior
.... (German goldsmith, 1664-1731) [500016193]

ULAN resource "Dinglinger, Georg Friedrich"[25]

In conclusion, the approach taken in the web service for linking to ULAN proves to be of high validity as decreasing confidence values as well as the number – one or multiple – of matches returned clearly correlate with imprecise or incomplete data in either the source or the target record(s). ULAN is a domain-specific agents authority well-fitting with the scope of the local source authorities, so the underlying information with name, birth and death date conforming in a source and the target record turns out to provide enough context to ensure valid matching results.

---

[23] http://www.bildindex.de/dokumente/html/kue20083100
[24] http://www.bildindex.de/dokumente/html/kue20560103
[25] http://vocab.getty.edu/page/ulan/500353265

6.2.2 *Linking to TGN vocabulary*

The quantitative results of the linking process to TGN can be summarized as follows – out of 50,412 resources from Bildarchiv Foto Marburg included in the matching, for 40,768, i.e. 80,9 % of the resources one or more TGN resources were returned as matching candidate. Among those with matches 54,6 % had one match and 45,4 % had two or more matches in TGN – for many resources the service returned quite a lot of matches, in the case of Berlin, Germany the number was 42 matches.

Matches	Number of input resources	
1	21,765	43,17 %
2	4,846	9,61 %
3 or more	12,750	25,29 %
No match	11,051	21,92 %
Total	**50,412**	**100 %**

Distribution of number of TGN matches for Bildarchiv Foto Marburg

For the resources with one match in TGN an analysis of 291 matching results was performed with the following distribution:

Confidence c	Number of resources		Manually checked		Of those are correct		Of those are incorrect		Need more research	
c = 1.0	11,933	54,83 %	120	1,0 %	115	95,8 %	0	0 %	5	4,2 %
0.9 < c < 1.0	6,096	28,01 %	61	1,0 %	56	91,8 %	0	0 %	5	8,2 %
0.8 < c <= 0.9	1,809	8,31 %	53	2,9 %	43	81,1 %	8	15,1 %	2	3,8 %
0.7 < c <= 0.8	599	2,75 %	18	3,0 %	13	72,2 %	4	22,2 %	1	5,6 %
0.6 < c <= 0.7	847	3,89 %	25	3,0 %	19	76,0 %	5	20,0 %	1	4,0 %
0.5 < c <= 0.6	481	2,21 %	14	2,9 %	9	64,3 %	4	28,6 %	1	7,1 %
Total	**21,765**	**100 %**	**291**	**1,3 %**						

Overview of matching results for Bildarchiv Foto Marburg for resources with one TGN match

Looking at the unclear matches with confidence 0.9 < c <= 1.0 it appears that in 8 of the 10 cases the actual geographic area is correct and the difference is on the administrative level, i.e. the place type, so that the match is very likely to be correct. In one case the source data was not up-to-date. It is therefore concluded that possible mismatches on this level are accepted, and the local authority records are automatically enriched for the 18,029 resources with confidence 0.9 < c <= 1.0, without further checking.

However, when analyzing the results for resources with more than one TGN match it turns out that we cannot take the same approach as with ULAN results, i.e. using the match with the highest confidence value. The TGN service very often returns the exactly same confidence values for different matches, even multiple matches with all having confidence 1.0 are found. Therefore no automatic processing of the results as they stand is possible.

We have identified two areas in which further refinement of the matching service could be investigated:

a) The fuzzy string searching may not only be applied for the preferred and alternative names of the actual resource, but also for the names of the contextual data, i.e. continent, nation, part of place information.

b) The place type should be added as contextual information from the source data if available. In many cases multiple or incorrect matches apparently could be avoided with a comparison of the type of the place in the source and the target. For example, in the case of Berlin, Germany the distinction between Berlin as inhabited place and the German state Berlin could be unambiguously distinguished this way although all other context information is the same for both places. However, this approach would require a preceding mapping of place types occurring in the source and in the target. In TGN place types are controlled concepts from the AAT.

It goes without saying that georeferences will also be used if available from the source data. However, the lack of georeferences in the local place authority is typically one of the primary motivations for a data provider to link to a published source.

An even more promising strategy would probably be to proceed in iterations through the hierarchy, e.g:

1) Match continent level of source resources – use results for step 2)
2) Match nation level of source resources – use results for step 3)
3) Match state level, and so on.

Such an approach would of course require more preliminary analysis of the source data.

6.3 Providing linking results to Europeana

Finally, after having performed the steps as described in the previous section:

Step 1) Provide source data from a local agent or place authority in a CSV file for the web service;

Step 2) Run the web service and provide result files in JSON or CSV format to provider;

Step 3) Evaluate the linking results and decide if and to which extent, based on the confidence values, the local authority records will be enriched with the links, either automatically or by intellectual double-check;

the following steps 4) to 6) need to be taken by the provider for sending the results of the linking process to Europeana:

Step 4) Make sure that the export functionality for object descriptions as provided by the collection management system does include in the exported metadata records not only internal references to the local authority records, but also the enriched links; if this is not the case update the export functionality accordingly.

Step 5) Make sure that the mapping of the exported metadata records to the target schema LIDO or EDM does also include the enriched links, i.e. links are carried on to the LIDO and eventually the EDM records; if this is not the case update the mapping accordingly.

Step 6) Provide an update of all object descriptions delivered to Europeana on basis of the enriched local authority records and the refined mapping in order to include the embedded links into the metadata published by Europeana.

In addition, all future publications of metadata in Europeana of the same provider will automatically have included the enriched links as they continue to be stored in the local collection management system.

In this way, Europeana benefits in the most efficient and sustainable way from the linking results, i.e. by harvesting records that have been annotated by the semi-automatic way we described that guarantees good results.

6.4 Linking to other external sources

While the linking to external sources through the web services presented in the previous sections was the focus of our work in this deliverable, we finally want to highlight on another, so to say basic strategy which was recommended in the conclusions of D4.2:

> "*In general, it can be expected that establishing links will be the more reliable the closer the linking process is tied to the source data in terms of its semantics. It is therefore a highly recommended strategy to include links, e.g. URIs for resources in LOD datasets, already in the actual metadata production phase. So partners may use the list of LOD source candidates presented in this deliverable as an inspiration for their own metadata production process and figure out opportunities to include such links from the outset.*"[26]

To support this approach, a number of target sources like the Partage Plus vocabularies for object type, material and technique, or the Europeana Photography vocabulary for subject information were integrated into the MINT instance of AthenaPlus. This way data providers had the opportunity to link their content to these targets during the mapping process.

As there are so many different sources that have potentially been used as targets, and so many content providers involved, no detailed analysis of the content delivered to Europeana is available regarding the number of links to external resources which were included this way. However, it can be stated that the approach combined with the recommended strategy of enriching underlying terminologies proves to be very efficient, as we can see in the example of Philipps-Universitaet Marburg – Bildarchiv Foto Marburg.

	Number of object descriptions	**Partage Plus vocabulary links[27]**	**AAT links**
Bildarchiv Foto Marburg	59,245	76,102	22,415
Germanisches Nationalmuseum	24,800	30,557	54,391
Gleimhaus Halberstadt	5,043	6,612	9,619
Herzog August Bibliothek Wolfenbüttel	26,466	28,653	45,785
LWL-Museum für Kunst und Kultur	18,497	24,567	40,076
Staatliche Graphische Samml. München	9,760	12,485	20,346
Staatsbibliothek zu Berlin	22,900	26,101	48,882
Universitätsbibliothek Leipzig	14,350	20,288	32,765
Veste Coburg	6,092	18,290	30,020
Bildarchiv Foto Marburg – Update of content delivered through Athena	765,755	1,001,954	299,447
Total	**952,908**	**1,245,609**	**603,746**

Figure: Overview of links to external sources in Europeana by Bildarchiv Foto Marburg

After manually enriching their local vocabulary for object types, material, and techniques with links to the Partage Plus vocabularies and to the Getty Thesaurus of Art and Architecture, Bildarchiv Foto Marburg delivered with its new content in AthenaPlus and with the update of the content previously delivered to Europeana through the Athena project the above number of links in its object descriptions. Each object description, in addition to other, by everyday routine included external links, e.g. to the ICONCLASS vocabulary, now contains on average 1.94 links to these two external sources.

---

[26] AthenaPlus D4.2 Review on Linked Open Data sources, p.52, http://www.athenaplus.eu/getFile.php?id=190
[27] Including exact matches (skos:exactMatch) to Getty AAT.

7 CONCLUSIONS

In this deliverable we presented the approach taken by the AthenaPlus project to encourage activities and support partners in the provision of semantically richer metadata to the cultural heritage community, and particularly Europeana, by linking the metadata to external data sources.

Taking into account considerations from both Europeana's side, as expressed by their taskforces on a Multilingual and Semantic Enrichment Strategy and on Metadata Quality, and from the data providers' side, we identified the following important factors for a successful, high-quality and sustainable enrichment of metadata with links to external sources:

4) The enrichment should be adopted as early as possible in the process of metadata production.
5) The enrichment should be based not only on string matching mechanisms, but rather exploit further contextual information.
6) The enrichment workflow should allow the data providers to validate and use the enrichment results according to their own specific criteria.

The core strategy we therefore followed was to support the enrichment of underlying local terminologies with links to the external sources, with the ultimate goal to feed these links into the metadata provided, both updating existing metadata and enriching any newly produced metadata.

To this end, we developed a web service for linking to two Getty vocabularies, the Union List for Artist Names for agents, and the Getty Thesaurus for Geographic names for places. These were chosen from the list of possible targets as provided through the preceding Review on Linked Open Data Sources, based on the following considerations:

- They are published with an Open License.
- They support RDF and SPARQL 1.1.
- Many partners aim at linking to the Getty vocabularies anyway as they are domain-specific and most reliable sources in terms of quality and sustainability.

The evaluation of the web services suggests that the linking process to ULAN is of high validity, though only using the life dates of an agent as additional context information to the preferred and alternative names. Decreasing confidence values as well as the number – one or multiple – of matches returned by the service clearly correlate with imprecise or incomplete data in either the source or the target record(s). For the linking process to TGN, however, further refinements are suggested as the context information as currently exploited – continent, nation, and other places the actual place in question is part of – is not sufficiently ensuring unique and valid results.

In general, the strategy of enriching underlying terminologies in the local metadata production environment with links to external sources proves to be very efficient for increasing the number of links and thereby the quality of the metadata on cultural heritage objects eventually published as Linked Open Data.

Also Europeana benefits in the most efficient and sustainable way from the linking results, i.e. by harvesting records that have been annotated by the semi-automatic way here described that guarantees good results.

APPENDIX 1: REFERENCES

Dangerfield, MC. et. al (2015): Report and Recommendations from the Task Force on Metadata Quality. http://pro.europeana.eu/files/Europeana_Professional/Publications/Metadata QualityReport.pdf (Last access October 31, 2015)

Isaac, A., Manguinhas, H., Stiller, J., & Charles, V. (eds.) (2015): EuropeanaTech Task Force on Evaluation and Enrichment: Final report. http://pro.europeana.eu/files/Europeana_Professional/EuropeanaTech/EuropeanaTech_taskforces/Enrichment_Evaluation//FinalReport_EnrichmentEvaluation_102015.pdf (Last access October 31, 2015)

Köhler, W., Stein, R. (2013), Review on Linked Open Data Sources. AthenaPlus Deliverable D4.2. http://www.athenaplus.eu/getFile.php?id=190 (Last access October 31, 2015)

Stiller, J., Isaac, A. & Petras, V. (eds.) (2014): EuropeanaTech Task Force on a Multilingual and Semantic Enrichment Strategy: Final report. http://pro.europeana.eu/files/Europeana_Professional/EuropeanaTech/EuropeanaTech_taskforces/MultilingualSemanticEnrichment//Multilingual%20Semantic%20Enrichment%20report.pdf (Last access October 31, 2015)

APPENDIX 2: TERMS AND ABBREVIATIONS

CSV	Comma Separated Values
DC	Dublin Core
EDM	Europeana Data Model
JSON	JavaScript Object Notation
LIDO	Lightweight Information Describing Objects
RDF	Resource Description Framework
SPARQL	SPARQL Protocol and RDF Query Language
TGN	Getty Thesaurus of Geographic Names
ULAN	Union List of Artist Names
XML	Extensible Markup Language
XSD	XML Schema

APPENDIX 3: ULAN Input JSON

```
[{
    "gender": "Male",
    "names": [{
        "fullName": "Leonardo",
        "lang": "en"
      }
    ]
  },
  {
    "gender": "Male",
    "names": [{
        "fullName": "Nikolaos Emmanouil Simou",
        "lang": "en"
      } ,
      {
        "fullName": "Νίκος Σίμου",
        "lang": "el"
      },
      {
        "fullName": "Νίκολαος Ε. Σίμου",
        "lang": "el"
      }
    ],
    "birthDate": [{
        "date": "01/26/1981",
        "estimation": "after",
    },
    {
        "date": "01/26/1985",
        "estimation": "before",
    },
    ],
    "birthDate": [
    {
        "date": "01/26/1981",
      "estimation": "exact",
    }
    ]}
]
```

APPENDIX 4: TGN Input JSON

```
[
 {
  "prefName": {
   "name": "Athens",
   "lang": "en"
  },
  "altName": [
   {
    "name": "Αθήνα",
    "lang": "el"
   },
   {
    "name": "Athene",
    "lang": "en"
   },
  ],
  "nation":[{
    "name": "Greece",
    "lang": "en"
  },
  {
    "name": "Hellas",
    "lang": "en"
  },
  {
    "name": "Ελλάδα",
    "lang": "en"
  }
  ],
  "continent":[{
    "name": "Europe",
    "lang": "en"
  },
  {
    "name": "Ευρώπη",
    "lang": "en"
  }
  ],
  "partOfPlace":[{
    "name": "Attica",
    "lang": "en"
  },
  {
    "name": "Αττική",
    "lang": "en"
  }
  ]
 },
{
  "prefName": {
   "name": "Athens",
   "lang": "en"
  },
  "altName": [
   {
    "name": "Αθήνα",
    "lang": "el"
   },
   {
    "name": "Athene",
    "lang": "en"
   },
  ],
  "nation":[{
    "name": "Greece",
    "lang": "en"
  },
  {
    "name": "Hellas",
    "lang": "en"
  },
  {
    "name": "Ελλάδα",
    "lang": "en"
  }
  ],
  "continent":[{
    "name": "Europe",
    "lang": "en"
  },
  {
    "name": "Ευρώπη",
    "lang": "en"
  }
  ],
  "partOfPlace":[{
    "name": "Attica",
    "lang": "en"
  },
  {
    "name": "Αττική",
    "lang": "el"
  },
  {
    "name": "Στερεά Ελλάδα",
    "lang": "el"
  }
  ]
 }

]
```

APPENDIX 5: ULAN Output JSON

```
[
  {
    "input": {
      "id": "AA_0",
      "name": "Leonardo"
    },
    "results": [
      {
        "ULAN": "http://vocab.getty.edu/ulan/500010879",
        "confidence": "0.75",
        "name": "Leonardo da Vinci",
        "birthYear": "1452",
        deathYear: "1519"
      },
      {
        "ULAN": "http://vocab.getty.edu/ulan/500072027",
        "confidence": "0.75",
        "name": "Leonardo da Teramo",
        "birthYear": "1346",
        deathYear: "1465"
      },
      {
        "ULAN": "http://vocab.getty.edu/ulan/500315043",
        "confidence": "0.75",
        "name": "Bruni, Leonardo",
        "birthYear": "1370",
        deathYear: "1444"
      },
      {
        "ULAN": "http://vocab.getty.edu/ulan/500236640",
        "confidence": "0.6168996928526715",
        "name": "Lugli, Leonardo",
        "birthYear": "1930",
        deathYear: "1991"
      },
      {
        "ULAN": "http://vocab.getty.edu/ulan/500247840",
        "confidence": "0.6168996928526715",
        "name": "Mosso, Leonardo",
        "birthYear": "1926",
        deathYear: "2080"
      },
      {
        "ULAN": "http://vocab.getty.edu/ulan/500027004",
        "confidence": "0.6035533905932737",
        "name": "Leonardo, Jusepe",
        "birthYear": "1601",
        deathYear: "1653"
      },
      {
        "ULAN": "http://vocab.getty.edu/ulan/500061140",
        "confidence": "0.5807189138830737",
        "name": "Chiaiese, Leonardo",
        "birthYear": "1721",
        deathYear: "1821"
      }
    ]
  },
  {
    "input": {
      "id": "AA_1",
      "name": "Nikolaos Emmanouil Simou"
    },
    "results": [

    ]
  }
]
```

APPENDIX 6: TGN Output JSON

```
{
  "input": {
    "id": "geo00000007",
    "name": "Syrakus (griechisch)"
  },
  "results": [
    {
      "TGN": "http://vocab.getty.edu/tgn/1001753",
      "name": "Siracusa",
      "confidence": 0.7337297168305051
    },
    {
      "TGN": "http://vocab.getty.edu/tgn/7003794",
      "name": "Syracuse",
      "confidence": 0.7337297168305051
    },
    {
      "TGN": "http://vocab.getty.edu/tgn/7039097",
      "name": "Stazione Siracusa",
      "confidence": 0.5555555555555556
    },
    {
      "TGN": "http://vocab.getty.edu/tgn/7039096",
      "name": "Stazione Siracusa Marittima",
      "confidence": 0.43611853819547153
    },
    {
      "TGN": "http://vocab.getty.edu/tgn/7056784",
      "name": "Belvedete",
      "confidence": 0.41942488165887604
    }
  ]
}
```