Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022), pages 6280–6289
Marseille, 20-25 June 2022

A Large Interlinked Knowledge Graph of the Italian Cultural Heritage

Stefano Faralli, Andrea Lenzi, Paola Velardi
Department of Computer Science University of Rome Sapienza
Rome, Italy
{faralli, lenzi, velardi}@di.uniroma1.it

Abstract
Knowledge is the lifeblood for a plethora of applications such as search, recommender systems and natural language understanding. Thanks to the efforts in the fields of Semantic Web and Linked Open Data a growing number of interlinked knowledge bases are supporting the development of advanced knowledge-based applications. Unfortunately, for a large number of domain-specific applications, these knowledge bases are unavailable. In this paper, we present a resource consisting of a large knowledge graph linking the Italian cultural heritage entities (defined in the ArCo ontology) with the concepts defined on well-known knowledge bases (i.e., DBpedia and the Getty GVP ontology). We describe the methodologies adopted for the semi-automatic resource creation and provide an in-depth analysis of the resulting interlinked graph.

Keywords: Cultural Heritage, Linked Open Data, Knowledge Graph, Ontology, ArCo, GVP, DBpedia

1. Introduction

Numerous efforts have been made in recent years to create knowledge bases (Abu-Salih, 2021). Humans and machines require a formal definition of knowledge to enhance a variety of activities in support of applications such as automated reasoning, cataloging, documentation, discovery, promotion, and recommendation of large collections of entities and facts.

Specifically, all the applications involving Machine Learning, or in general, Artificial Intelligence algorithms, perform better under the presence of large collections of data (Feigenbaum, 1984). Unfortunately, the creation of formal representations of knowledge is both a complex task, requiring the participation of experts, and a time-consuming process.

To address this problem, the Semantic Web and Linked Open Data (LOD) communities have provided mechanisms for the definition of formal ontologies and the interconnection of entities through knowledge bases (Gandon, 2018), thus helping both humans and machines during the creation and engineering of knowledge bases. LOD data have been created in many domains, but in many others, they are still largely unavailable. For example, despite the great interest in cultural heritage data (Mountantonakis and Tzitzikas, 2019), interlinking efforts are limited to relatively small resources, such as entities from libraries, archeological sites and museums (Pontes et al., 2020; Kyriaki-Manessi et al., 2018).

In this paper, we describe the adopted methodologies and the resulting edition of a very large resource obtained within the ArCo, GVP and DBpedia Linking Initiative (AGDLI) (Faralli et al., 2021). AGDLI is a research activity within the project *SMARTOUR: intelligent platforms for tourism*, funded by the Italian Ministry of University and Research, aimed at linking the *ArCo*'s catalogue of Italian cultural entities (Carriero et al., 2019) to the well-known *Getty Vocabulary Program* (GVP) (Harpring, 2010) and DBpedia (Auer et al., 2007) ontologies, with the main goal of providing a semantically rich and structured representation of the Italian cultural heritage for knowledge-based applications. The main objective of our linking initiative is putting together the qualities of two large resources, one of which, ArCO, provides rich, but mostly unstructured, descriptions of Italian cultural heritage *entities*, while the other, GVP, is a highly structured lexical taxonomy of cultural heritage *concepts*.

Our resource is aimed at supporting and enhancing various activities, such as, for example, cataloging activities in museums, or suggesting personalized tours based on the artistic and historical interests of users (Binucci et al., 2017). Furthermore, the resource provides the basis for an on-going graph completion task, since, given the semi-automatic nature of the adopted linking methodologies and the incompleteness of the linked resources, the resulting knowledge graph is expected to be both error-prone and incomplete.

The rest of this paper is organized as follows: in Section 2, we describe the resources we are interlinking; in Section 3, we provide details about the investigated semi-supervised methodologies involved for our linking purposes; in Section 4, we provide an in-depth analysis of the resulting resource; in Section 5 we briefly describe the related work; and finally, in Section 6, we discuss future research directions and plans.

2. Linked Resources

The resource described in this article is the result of a research activity that aims to study semi-supervised methodologies to improve the semantic definitions of the Italian cultural heritage, to be used in various knowledge-based applications related to tourism, such as recommender systems (Zhang et al., 2021) and semantically-enriched augmented reality tools for point of interests discovery (Ruta et al., 2014).

Table 1: Example extracted from the properties defined in ArCo for a 17th century painting, located in the city of Rome, Italy. As shown, property values are mostly unstructured textual strings (*literals*).

property	value
uri	`https://w3id.org/arco/resource/HistoricOrArtisticProperty/1200252386`
type	"dipinto"
subject	"paesaggio (dipinto) by Salvator Rosa (scuola) (sec. XVII, seconda metà) "
date	"1650-1699"
latitude	41.90745
longitude	12.498603
city	`https://w3id.org/arco/resource/City/roma`
city label	"ROMA"
author	`https://w3id.org/arco/resource/Agent/aea151c6ed45d80e78eb79b4ec150aca`
author label	"Salvator Rosa, Scuola"
author date	"1615/ 1673"

Table 2: Example excerpt of the properties provided in our resource, in the form of links to external concepts, for the cultural heritage entity described in Table 1.

property	value	
dateFrom	1650	
dateTo	1699	
authorDateFrom	1615	
authorDateTo	1673	
AAT	`http://vocab.getty.edu/aat/300265015`	["dipinti"]
AAT	`http://vocab.getty.edu/aat/300033618`	["paintings (visual works)"]
AAT	`http://vocab.getty.edu/aat/300015636`	["landscapes (representations)"]
AAT	`http://vocab.getty.edu/aat/300054709`	["landscaping"@en]
AAT	`http://vocab.getty.edu/aat/300386959`	["scenery (landscape)"]
TGN	`http://vocab.getty.edu/tgn/7032914`	["Via Clodia"]
DBpedia	`https://dbpedia.org/resource/Rome`	

	entities	classes	max height
ArCo	650K	14	4
ULAN	307K	32	4
	leaf concepts	**concepts**	**max height**
AAT	44.8K	10.6K	17
	leaf geo-entities	**geo-entities**	**max height**
TGN	2.5M	10.5k	10

Table 3: Number of entities, classes and max height of taxonomic structures in *ArCo* (`https://w3id.org/arco/ontology/arco`), *AAT*, *TGN* and *ULAN*. Note that *AAT* is a taxonomy of concepts.

To this end, we applied information extraction techniques to link the entities defined in *ArCo*[1] (Carriero et al., 2019) with the concepts belonging to the *Getty Vocabulary Program*[2] (*GVP*) (Harpring, 2010) and *DBpedia* (Auer et al., 2007)[3] ontologies.

ArCo is an ontology network for representing the Italian cultural heritage, which defines more than 650K georeferenced cultural entities. Table 1 shows and example of cultural heritage definition. As shown, semantic properties (such as type and subject) are valued with textual strings and are not linked with concepts of external ontologies. Even dates and city labels do not follow a precise syntax, as detailed later. Furthermore, the ArCo entities are linked to an ontology that defines only few high-level classes of cultural heritage.

The *GVP* is a lexical ontology composed by three resources: the *Art & Architecture Thesaurus®* (*AAT*), the *Getty Thesaurus of Geographic Names®* (*TGN*), and the *Union List of Artist Names®* (*ULAN*).

Table 3 shows that, while *ArCO*[4] provides rich, although mostly unstructured, information on cultural *entities*, the *AAT* is characterized by a highly structured conceptual taxonomy, with only a few entities connected to them. It provides semantic definitions for concepts, useful for supporting and enhancing activities such as cataloging, documenting and for retrieving information related to art, architecture, and other material culture. The *TGN* is a taxonomy of geographical entities, related by inclusion (*subAreaOf*) relations. Finally, the *ULAN* is a list of artists, ordered in alphabetical classes (e.g., artists whose name begins by A, B . . .). We do not provide statistics on *DBpedia* since it is a very popular and well-known resource: interested

[1] `http://wit.istc.cnr.it/arco/?lang=en`.

[2] `https://www.getty.edu/research/tools/vocabularies/`.

[3] `https://www.dbpedia.org/`.

[4] ArCo Module Namespace, `https://w3id.org/arco/ontology/arco`

readers can refer to (Auer et al., 2007) and other publications. By targeting both the *GVP* and *DBpedia* ontologies, we aim to generate, with high coverage, conceptual links for *ArCo* entities and their properties (as shown in Table 2).

3. Approach

As previously summarized, the purpose of our work is to connect entities of the *ArCo* catalogue primarily with the concepts of the *GVP* taxonomies, which are specific of the art and architecture domain, and in addition, to *DBpedia*, which may provide additional information concerning places and authors. The task is complex since, first, *ArCo* entities are described by textual strings, as previously noted. Secondly, *GVP* concepts are mostly, although non exclusively, in English. Translations are provided for a number of languages, including Italian, but they are extremely sparse.

In Figure 1, we depict the workflow of the steps performed to generate the resource. Block 1 retrieves ArCo *rdf* triples by submitting *SPARQL* queries (Pérez et al., 2006) to a dedicated endpoint[5]. In block 2, property values are retrieved for each entity, for subsequent processing by dedicated information extraction modules (described in detail in the next sections). In block 3, *AAT concept labels* are semi-automatically translated in Italian, and *entity linking* (block 4) is performed to link *ArCo type* and *subject* textual descriptions to the appropriate *AAT* concepts, e.g.: "*dipinti*" → "painting (visual works)", "paesaggio" → "landscapes (representations)".

Block 7 aligns *city label*s with *DBpedia* concepts describing the related places (e.g., "ROMA" → `https://dbpedia.org/resource/Rome`). Block 8 process unstructured date expressions to identify intervals "1650-1699"→ $(1650, 1699)$, as better detailed in Section 3.3. Finally, blocks 5 and 6 align the coordinates (latitude and longitude) of the *ArCo* entities with concepts of the *TGN* taxonomy of geographic names, $(41.90745, 12.498603)$ → `http://vocab.getty.edu/tgn/7032914` (via Clodia, an old Roman road)[6].

3.1. Processing of Types and Subjects

The *type* and *subject* properties describe in natural language each cultural heritage. The *type* is typically a word, e.g., "dipinto" ("painting") or a multi-word, for example: "lapide commemorativa ai caduti" which translates "commemorative stone to the fallen", while the *subject* is a free sentence describing the entity. For example, the subject value in Table 1 translates as follows: "landscape (painting) by Salvator Rosa (school) (second half of XVII cent.)". The task of block 4 in Figure 1 is to extract from textual strings one or more entities matching the concepts of *AAT*, to provide appropriate linking. The over 55k concepts of *AAT* have been translated in Italian with the support of *Google Translation AI*[7] with some post-editing to fix errors. Next, the *type* and *subject* strings are lemmatized (to normalize gender and grammatical number of words), and converted into *bag of words* counting the occurrences of *AAT* concepts labels in the strings. Table 2 shows five matching *ATT* concepts retrieved for the entity of Table 1.

Although the *AAT* is a domain specific ontology, there are a few concept labels for which different meanings are provided, some of which may be incorrect (a discussion on the resulting error rate is reported in Section 4.2). We note that entity disambiguation is still an on-going activity within the *AGDLI* project[8].

3.2. City labels and geographic coordinates

The literals corresponding to *city labels* describe the Italian city (or geographic area) where a cultural heritage is located. With reference to the excerpt example in Table 1, the cultural heritage is located in "Roma"[9]. City labels are used in block 7 of Figure 1 "Entity Alignment with DBpedia places", to obtain an alignment between `https://w3id.org/arco/resource/City/roma` and `https://dbpedia.org/page/Rome`[10]. We performed entity alignment by first ranking all the *DBpedia* places by the edit distance with *ArCo city label*[11]. Next, guided by the automatic ranking, we perform the selection, with some manual post-editing.

The reason for matching geographic names with *DBpedia* rather than with *TGN* is that in *TGN* location names are in English (while *DBPedia* provides multi-language labels) and furthermore, smaller Italian municipalities are not included in the *TGN*. Rather, the *TGN* is used in a more reliable way by matching the geographic coordinates (*latitude* and *longitude*) provided in *ArCo* with those in *TGN*. City or location labels are next retrieved through proximity and reverse geocoding in block 6 of Figure 1 ("Entity alignment by proximity"). Note that, in this way, we may obtain more than one geo-reference for a given entity. Subsequent processing verifies that retrieved locations are compatible, that is, either coincide, or are included one into the other (through the relation SubAreaOf), for example: *"Palazzo Poli"* → *"Roma"*).

[5] https://dati.beniculturali.it/sparql

[6] We note that, unfortunately, *ULAN* mostly includes modern artists, and there is a very limited overlapping with *ArCo* artists. Therefore, at the moment, we are seeking better resources to enrich the ArCo information on artists.

[7] `https://cloud.google.com/translate`

[8] the on-going task of knowledge graph completion includes both automated pruning and enrichment of the created resource.

[9] The Italian capital city.

[10] *ArCo* already defines an alignment for some of the cities.

[11] Note that in *ArCo* all entities are linked to some city, which is the closest municipality.

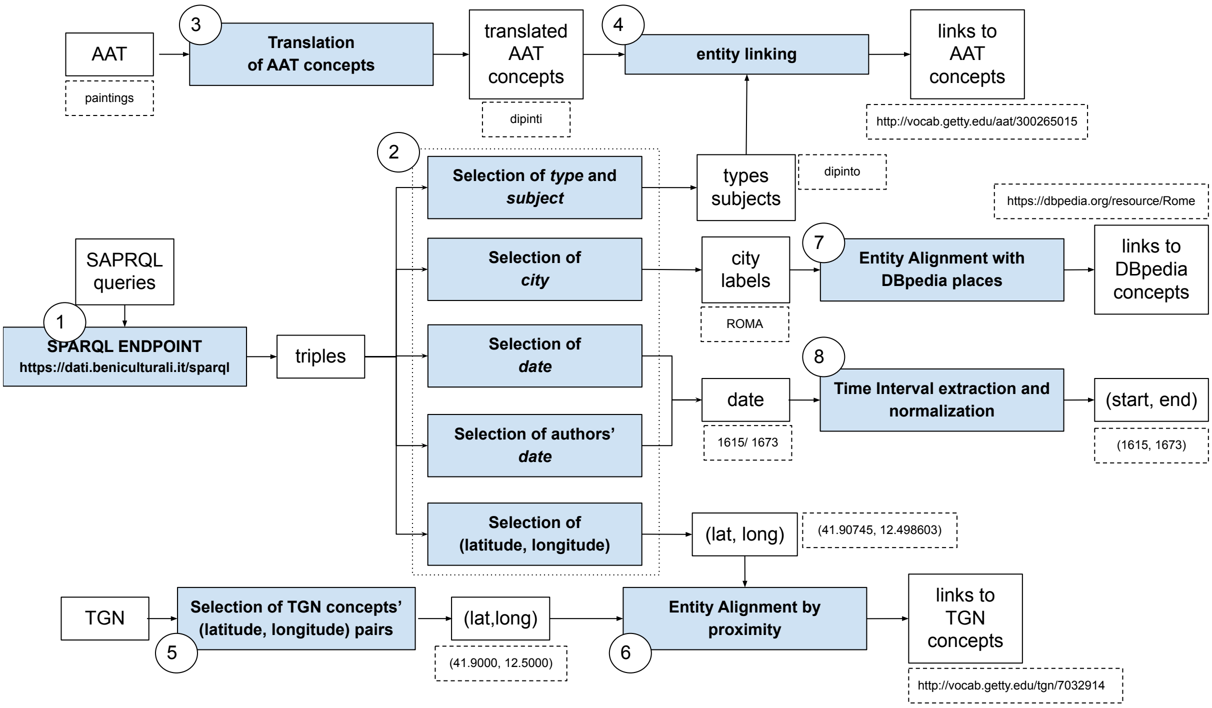

Figure 1: Workflow of the tasks performed to create the presented resource.

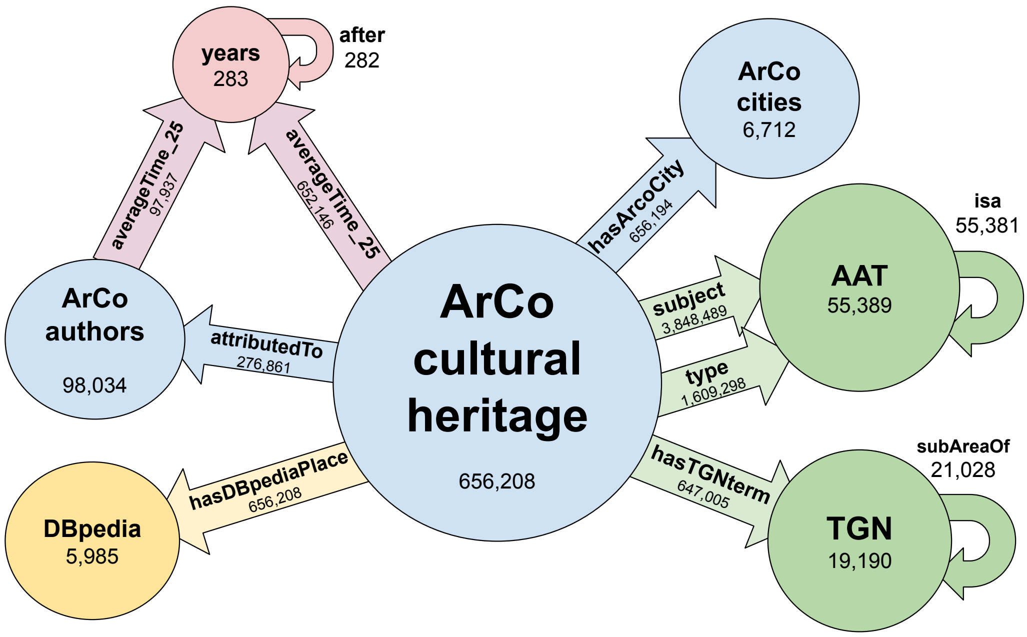

Figure 2: A high-level representation of the resulting interlinked graph.

3.3. Dates

As shown in the example of Table 1, dates are literals defining the date of the cultural heritage "1650-1699" and, when the cultural heritage has been attributed to an author, to the life span or the activity time interval of the author "1615/ 1673". Dates can be expressed in a variety of ways, like: "ca 1800- ca 1900" ("ca" is an abbreviation of "about"), or "XIX - primo quarto" ("first quarter of XIX cent."). In this step (block 8 in Figure 1), to extract machine readable pairs of the form *(start, end)* and a single date *(year)*, we defined a collection of patterns and rules to automatically extract the *start* and the *end* of a time interval. Table 4 shows a summary of such patterns and their coverage. Finally, to provide a simplified time representations, for each pair *(start,end)* we also computed the following:

- the average $\frac{start+end}{2}$;
- the approximation into quarters of century (e.g.,

Table 4: Top 20 patterns with example matches and their coverage.

Pattern	Example	Coverage
Y-Y	'1901-1972'	83.39%
Y ac-Y ac	'625 ac-550 ac'	4.22%
M/Y or Y/M	'1/1875' or '1920/3'	3.09%
D/M/Y-D/M/Y	'5/6/1805-1/9/1890'	2.81%
Y-	'1732-'	1.18%
C	'XX'	1.17%
D-D/M/Y	'10-25/08/2001'	0.62%
Y	'1650'	0.45%
Y/	'1315/'	0.36%
D/M/Y-Y	'23/11/1426-1430'	0.33%
C-C	'XVI-XVII'	0.25%
-D/M/Y	'-24/03/1290'	0.22%
D/M/Y-	'28/01/1135-'	0.21%
M Y-	'06 1468-'	0.20%
C ac-C ac	'XII ac-XI ac'	0.18%
M/Y-M/Y	'10/1890-01/1950'	0.17%
-Y	'-1922'	0.16%
Y ac-Y	'35 ac-46'	0.14%
M-M/Y	'10-12/1968'	0.13%
C ac-C	'I ac-I'	0.08%

1915 is approximated with 1900, and 1927 with 1925), the *start* value and the *end* value.

The reason for these additional representations is to link entities to a temporal timeline (through the *after* relation) and to support future graph completion algorithms, as mentioned in the introduction. In fact, the near contemporaneity of two artworks can help to infer characteristics of that of the two less richly described.

4. Resource

This section describes the current status of the generated resource.

4.1. Statistics

After the application of the methodologies described in Section 3, we obtained a large interlinked graph. As shown in Figure 2, $656k$ entities of *ArCo* have been linked through typed relationships to *AAT*, *TGN* and *DBPedia* concepts. More precisely, with reference to Figure 2 and Tables 5 and 6:

1. *ArCo* entities are linked through almost 4 millions relationships extracted from the *subject* field, and 1.6 million relationships extracted from the *type* to over $55k$ concepts of the *AAT* taxonomy, which are in turn related by *is-a* relations (as shown by the self-loop in the upper right of Figure 2);

2. Both *ArCo* entities and *ArCo* authors (over $98k$) are related to a temporal taxonomy of 283 years (ordered by the *after* relation);

Table 5: Statistics on the relations (edges) in the resulting interlinked graph.

relations	#
subject	3,843,489
type	1,609,298
hasDBpediaPlace	656,208
hasArCoCity	656,194
averageTime_25_ch	652,146
hasTGNterm	647,000
attributedTo	276,816
averageTime_25_agent	97,937
isA	55,381
subAreaOf	21,028
after	282
total	**8,515,779**

Table 6: Statistics on the nodes in the resulting interlinked graph.

nodes	#
cultural heritage	656,208
authors	98,034
AAT	55,389
TGN	19,190
ArCo cities	6,712
DBpedia places	5,985
years	283
total	**841,801**

3. *ArCo* entities are further geo-referenced with $656k$ *hasArcoCity* relations to $6.7k$ municipalities, with $656k$ *hasDBpediaPlace* relations to $5.9k$ locations in *DBPedia* and with 647k *hasTGNterm* relations to 19.190 TGN areas, taxonomically ordered according to *subAreaOf*. ArCO cities, TGN areas and DBPedia are in turn linked to each other (not shown for the purpose of clarity).

Table 7 focuses on the semantic links, by summarizing the statistics of the links to the *AAT* conceptual taxonomy of cultural heritage. The average number of linked concepts extracted from each *subject* field is 5.86, while from the *type* filed, we extracted an average of 2.45 concepts.
Indeed, the extensive linking to the *AAT* cultural heritage taxonomy is one of the main results of the *AGDLI* initiative, which paves the way to the application of knowledge graph completion algorithms for further refinement, pruning and extension of the released resource. Moreover, our resource provides rich temporal and spatial information, which is also taxonomically structured through the *after* and *subAreaOf* relations. To the best of our knowledge, although the problem of spatio-temporal linking has been addressed in some previous work (e.g., (Santipantakis et al., 2019) among others), this is the first large-scale LOD effort to create a resource that exploits spatio-temporal tax-

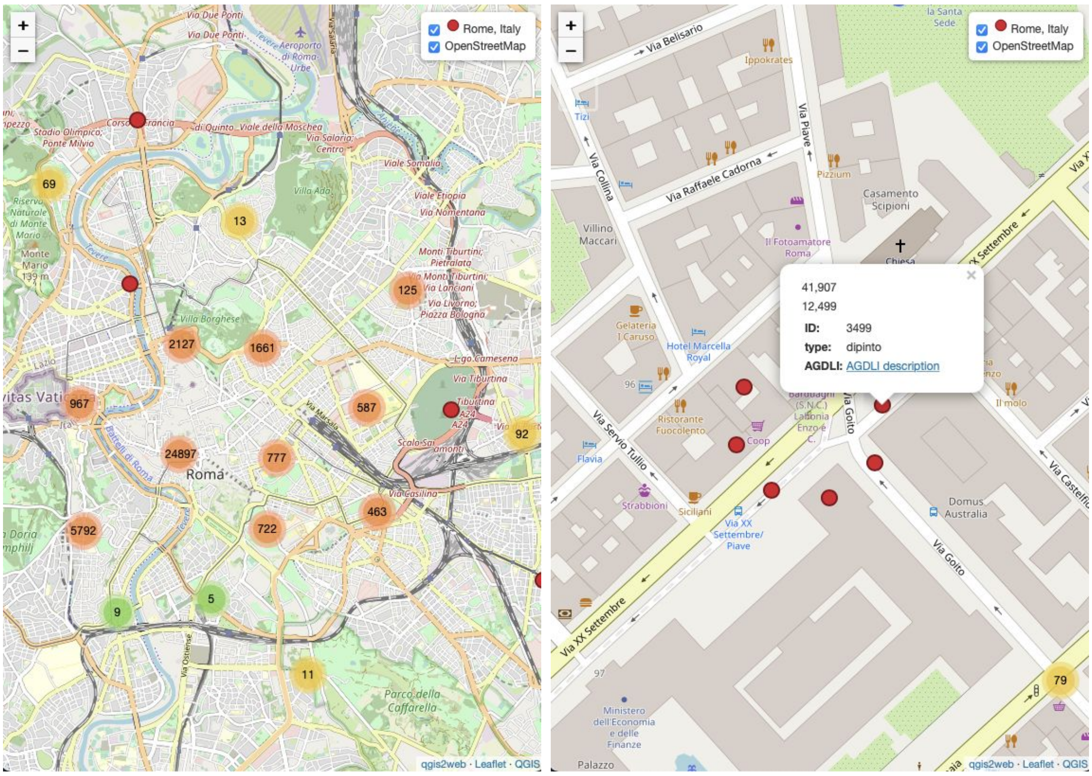

Figure 3: Interactive map (left) of the 38K *ArCo* cultural heritage entities located inside the circular area of radius 5 km (3.17 mile), around the Rome center (Latitude: 41.902782, Longitude: 12.496366) and detailed view (right) of the cultural heritage entity described in Table 1. When the user selects an entity, a pop-up provides a link to all the information listed in Tables 1 and 2 .

Table 7: Statistics about the links to *AAT* concepts.

	"subject" relation	***"type" relation***
mean	5.86	2.45
std	3.76	1.54
min	0	0
median	5	2
max	51	27

Table 8: Estimated error rate of generated links.

relations	**number tested**	**error rate**
subject	1,000	0.21
type	1,000	0.09
hasDBpediaPlace	500	0.00
hasTGNterm	500	0.07
averageTime_25_ch	500	0.00
averageTime_25_agent	500	0.00

onomies, with the objective of facilitating the detection of similarities among entities.

4.2. Resource quality

Although we assume that the on-going graph completion activity, supported by machine learning algorithms, will improve the overall quality of the resource, by both adding new links and removing wrong ones, we manually estimated over a number of randomly selected samples the error rate of the extracted links. The results are summarized in Table 8. Each sample link has been analyzed and evaluated by two human annotators[12]. Specifically, we extracted a random sample of 1,000 links, for each set of relations to candidate *AAT* entities extracted from the *type* and *subject* properties, and a random sample of 500 relations for the other set of links pointing to *DBpedia* places, *TGN* entities and averaged timestamps (i.e., *years*). By manual inspection, we observed that errors are mostly due to semantic ambiguity (entities related to wrong concepts due to ambiguous terms in the *type* and *subject* field) for example: the *AAT* concept `http://vocab.getty.edu/aat/300379382` "kids (goats)" has been linked to the many occurrences of the name "bambino" which means "child".

4.3. Availability

The resource is available[13] under Creative Commons Attribution 4.0 International (CC BY 4.0)[14].

[12] A third annotator has been involved to solve the few cases of disagreement.

[13] `https://sites.google.com/uniroma1.it/agdli/`.

[14] `https://creativecommons.org/licenses/by/4.0/deed.en`.

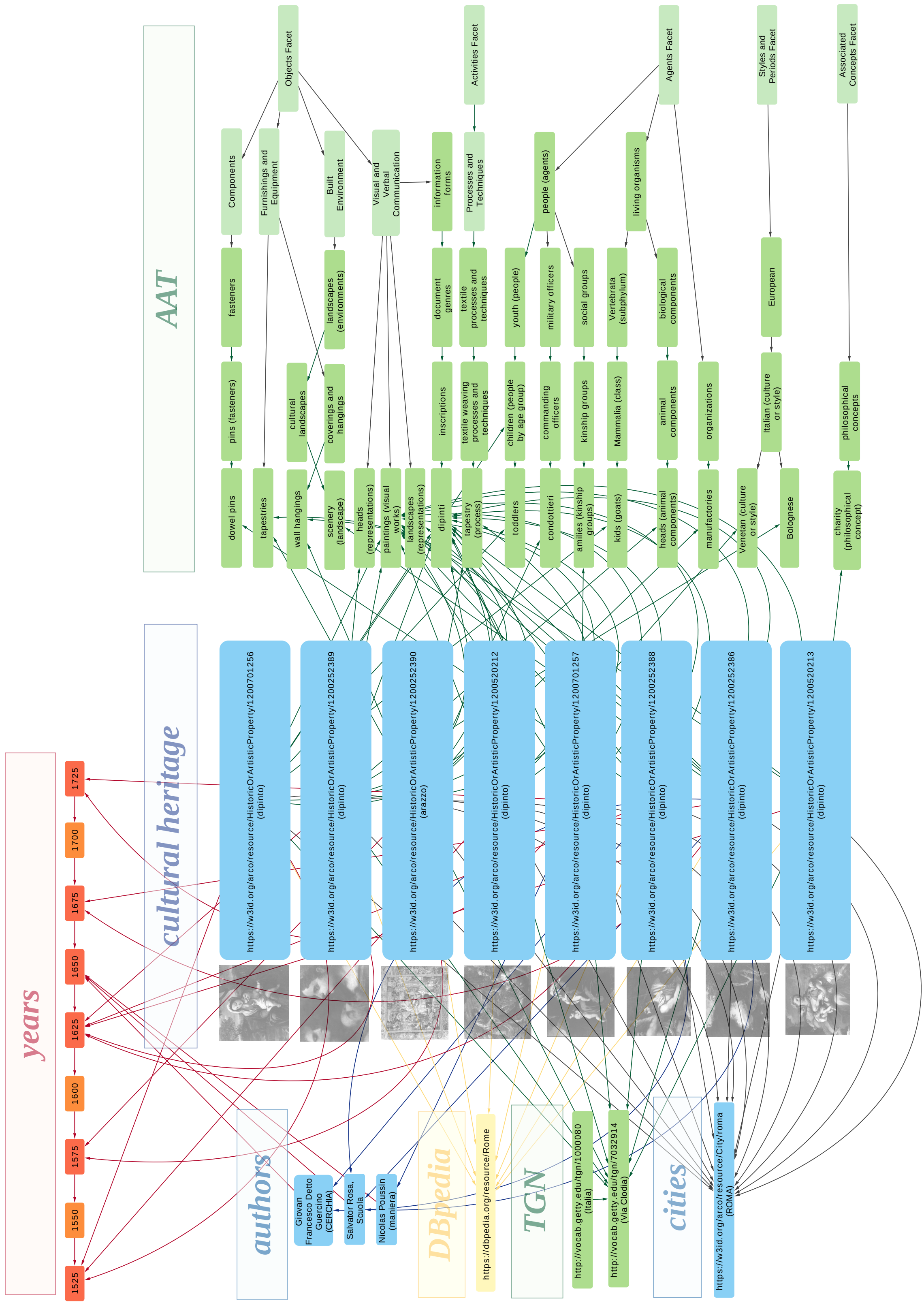

Figure 4: Portion of our interlinked graph describing the eight cultural heritage entities focused in Figure 3 (right). Some of the links to *AAT* and *TGN* are omitted, and the corresponding taxonomies are simplified for readability.

The current version of the graph is released in the form of a *tab separated values* (*.tsv*) file consisting of three columns (i.e., *subject*, *predicate* and *object*). Hence, our resource is ready to be processed with stateof-the-art tools such as, for instance, PyKEEN[15] "a python package designed to train and evaluate knowledge graph embedding models (incorporating multimodal information)" (Ali et al., 2021). The graph can be visualized through interactive maps, as the one in Figure 3 (left)[16] showing the $38K$ *ArCo* entities[17] found in the historical center of Roma, the capital city of Italy. In the same Figure (right) we focus on the interactive map of eight cultural heritage entities in Roma, and in Figure 4 we show the portion of our interlinked graph[18]. describing the same eight entities and the generated interlinks (i.e., to *years* and to the *DBpedia*, *AAT* and *TGN* ontologies).

5. Related Work

Knowledge graphs (KGs) are machine readable representations of knowledge, providing a formal definition for entities and relationships (Hogan et al., 2021; Ji et al., 2021). KGSs are more and more important, since their availability enables the development of a variety of applications in the fields (among others) of big data (Zou, 2020) and smart systems (Nguyen et al., 2020). Unfortunately, it is well known (Feigenbaum, 1984) that the creation and maintenance of KGs is a highly expensive and time consuming process involving experts and competences across different area of knowledge. To partially cope with creation costs, Linked Open Data offers principles, guidelines and tools for the interlinking of KGs, thus promoting the reuse of existing dictionaries (Saha and Mandal, 2021; Zeng et al., 2021) rather than, for instance, learning new dictionaries from scratch (Khadir et al., 2021).

The maintenance of KGs and specifically the activity of populating them with new entities, is a process that often relies on the processing of existing open datasets (Meherhera et al., 2020) by means of use case/domain specific transformation methodologies (Atemezing et al., 2012; Hallo et al., 2014; De Meester et al., 2017) or information extraction techniques (Fernàndez-Cañellas et al., 2020; Martínez-Rodríguez et al., 2020).

In the domain of cultural heritage, as surveyed in (Nishanbaev et al., 2019), a growing number of repositories are created by national organizations. The availability of linked cultural heritage data opens to a plethora of applications ranging from historical landscape studies (van Lanen et al., 2022) to authoring virtual exhibitions (Monaco et al., 2022), among others. In this domain, existing efforts share a number of technical challenges, for instance: datasets are designed with different models, schemas, or formats; and the same entities occur in different catalogs with different URIs. To cope with the above challenges, researchers are investigating methodologies and developing tools of linked data integration, as surveyed in (Mountantonakis and Tzitzikas, 2019). However, despite the large number of existing works in this field, linking efforts in the cultural heritage have been limited to relatively small resources, such as entities from libraries, archeological sites and museums (Pontes et al., 2020; Kyriaki-Manessi et al., 2018). To the best of our knowledge, AGDLI is the largest interlinked resource made available to date, both for the number and variety of interlinked entities and for the number of concepts.

6. Conclusions and Future work

In this paper, we presented a resource aimed at providing interlinks between the *ArCo* cultural heritage and the concepts and entities in *GVP* and *DBpedia* ontologies. To the best of our knowledge, this is one of the largest efforts in the domain of cultural heritage entity linking, resulting in a knowledge graph of one of the richest cultural heritage in the world (the Italian one), linking 656k entities to over 55k concepts of a highly structured taxonomy of art and architecture. In addition, entities are enriched with space-time links ordered according to inclusion and sequentiality criteria. Finally, upon a first manual analysis of its quality, we estimated an error rate less than 1% for temporal and geographical links, and around 20% for conceptual links, mostly caused by semantic ambiguity of concept labels.

In our linking initiative, we have planned to perform additional research activities to address the limitations of our current work. Regarding the disambiguation of candidate entities (see Section 3.1), we started the experimentation with entity disambiguation algorithms. To this end we are considering leveraging the results of recent investigations such as (Tedeschi et al., 2021) (analyzing the benefits in combining *named entity recognition* and *entity linking*) or (Yin et al., 2019) (leveraging deep models to cope with semantic ambiguity). Additional planned research activities are aimed at studying effective graph completion techniques (see the review (Chen et al., 2020)) for heterogeneous knowledge graphs such as the one we presented in this paper.

Acknowledgements

This work was carried out within the research project "SMARTOUR: intelligent platform for tourism" funded by the Italian Ministry of University and Research with the Regional Development Fund of European Union (PON Research and Competitiveness 2007-2013).

[15] `https://github.com/pykeen/pykeen`

[16] Link to the interactive map: `https://www.stefanofaralli.it/maps/smartour/rome/index.html#14/41.9059/12.4927`

[17] Link to the list of $38K$ cultural heritage entities: `https://www.stefanofaralli.it/maps/smartour/rome/html/arco/casestudy.html`

[18] Link to an interactive view of the graph `https://www.stefanofaralli.it/maps/smartour/graph8/`

7. Bibliographical References

Abu-Salih, B. (2021). Domain-specific knowledge graphs: A survey. *Journal of Network and Computer Applications*, 185:103076.

Ali, M., Berrendorf, M., Hoyt, C. T., Vermue, L., Sharifzadeh, S., Tresp, V., and Lehmann, J. (2021). PyKEEN 1.0: A Python Library for Training and Evaluating Knowledge Graph Embeddings. *Journal of Machine Learning Research*, 22(82):1–6.

Atemezing, G., Corcho, O., Garijo, D., Mora, J., Poveda-Villalón, M., Rozas-Larraondo, P., Vila-Suero, D., and Villazon Terrazas, B. (2012). Transforming meteorological data into linked data. *Semantic Web journal, Special Issue on Linked Dataset descriptions, 2012. IOS Press, ISSN: 1570-0844*, 01.

Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., and Ives, Z. (2007). Dbpedia: A nucleus for a web of open data. In *The Semantic Web*, pages 722–735, Berlin, Heidelberg. Springer Berlin Heidelberg.

Binucci, C., De Luca, F., Di Giacomo, E., Liotta, G., and Montecchiani, F. (2017). Designing the content analyzer of a travel recommender system. *Expert Systems with Applications*, 87:199–208.

Carriero, V. A., Gangemi, A., Mancinelli, M. L., Marinucci, L., Nuzzolese, A. G., Presutti, V., and Veninata, C. (2019). Arco: The italian cultural heritage knowledge graph. In *The Semantic Web - ISWC 2019 - 18th International Semantic Web Conference, Auckland, New Zealand, October 26-30, 2019, Proceedings, Part II*, volume 11779 of *Lecture Notes in Computer Science*, pages 36–52. Springer.

Chen, Z., Wang, Y., Zhao, B., Cheng, J., Zhao, X., and Duan, Z. (2020). Knowledge graph completion: A review. *IEEE Access*, 8:192435–192456.

De Meester, B., Maroy, W., Dimou, A., Verborgh, R., and Mannens, E. (2017). Declarative data transformations for Linked Data generation: The case of DBpedia. In Eva Blomqvist, et al., editors, *Proceedings of the 14th ESWC*, volume 10250 of *Lecture Notes in Computer Science*, pages 33–48. Springer, May.

Faralli, S., Lenzi, A., and Velardi, P. (2021). AGDLI: arco, GVP and dbpedia linking initiative. In Oshani Seneviratne, et al., editors, *Proceedings of the ISWC 2021 Posters, Demos and Industry Tracks: From Novel Ideas to Industrial Practice co-located with 20th International Semantic Web Conference (ISWC 2021), Virtual Conference, October 24-28, 2021*, volume 2980 of *CEUR Workshop Proceedings*. CEUR-WS.org.

Feigenbaum, E. A. (1984). Knowledge engineering. *Annals of the New York Academy of Sciences*, 426(1):91–107.

Fernàndez-Cañellas, D., Marco Rimmek, J., Espadaler, J., Garolera, B., Barja, A., Codina, M., Sastre, M., Giro-i Nieto, X., Riveiro, J. C., and Bou-Balust, E. (2020). Enhancing online knowledge graph population with semantic knowledge. In Jeff Z. Pan, et al., editors, *The Semantic Web – ISWC 2020*, pages 183–200, Cham. Springer International Publishing.

Gandon, F. (2018). A survey of the first 20 years of research on semantic web and linked data. *Ingénierie des Systèmes d Inf.*, 23(3-4):11–38.

Hallo, M., Luján-Mora, S., and Trujillo, J. (2014). Transforming library catalogs into linked data. In *Proceedings of the 7th International Conference of Education, Research and Innovation (ICERI 2014)*, pages 1845–1853, Seville (Spain).

Harpring, P. (2010). Development of the Getty Vocabularies: AAT, TGN, ULAN, and CONA. *Art Documentation: Journal of the Art Libraries Society of North America*, 29(1):67–72.

Hogan, A., Blomqvist, E., Cochez, M., d'Amato, C., Melo, G. d., Gutierrez, C., Kirrane, S., Gayo, J. E. L., Navigli, R., Neumaier, S., et al. (2021). Knowledge graphs. *Synthesis Lectures on Data, Semantics, and Knowledge*, 12(2):1–257.

Ji, S., Pan, S., Cambria, E., Marttinen, P., and Yu, P. S. (2021). A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–21.

Khadir, A. C., Aliane, H., and Guessoum, A. (2021). Ontology learning: Grand tour and challenges. *Computer Science Review*, 39:100339.

Kyriaki-Manessi, D., Vassilakaki, E., Triantafyllou, I., Giannakopoulos, G., and Antoniou, A., (2018). *The Use of Ontologies for Creating Semantic Links Between Cultural Artifacts and Their Digital Resources*, pages 541–545. Springer International Publishing, Cham.

Martínez-Rodríguez, J.-L., Hogan, A., and López-Arévalo, I. (2020). Information extraction meets the semantic web: A survey. *Semantic Web*, 11(2):255–335.

Meherhera, A., Mekideche, I., Zemmouchi-Ghomari, L., and Ghomari, A. R. (2020). A Survey of Current Approaches for Transforming Open Data to Linked Data. In *4th Edition of the National Study Days on Research on Computer Sciences, JERI'2020*, Saida, Algeria, June.

Monaco, D., Pellegrino, M. A., Scarano, V., and Vicidomini, L. (2022). Linked open data in authoring virtual exhibitions. *Journal of Cultural Heritage*, 53:127–142.

Mountantonakis, M. and Tzitzikas, Y. (2019). Large-scale semantic integration of linked data: A survey. *ACM Comput. Surv.*, 52(5), sep.

Nguyen, H. L., Vu, D. T., and Jung, J. J. (2020). Knowledge graph fusion for smart systems: A survey. *Information Fusion*, 61:56–70.

Nishanbaev, I., Champion, E., and McMeekin, D. A. (2019). A survey of geospatial semantic web for cultural heritage. *Heritage*, 2(2):1471–1498.

Pérez, J., Arenas, M., and Gutierrez, C. (2006). Semantics and complexity of sparql. In Isabel Cruz, et al., editors, *The Semantic Web - ISWC 2006*, pages 30–43, Berlin, Heidelberg. Springer Berlin Heidelberg.

Pontes, E. L., Cabrera-Diego, L. A., Moreno, J. G., Boros, E., Pontes, E. L., Hamdi, A., Sidère, N., Coustaty, M., and Doucet, A. (2020). Entity Linking for Historical Documents: Challenges and Solutions. In *22nd International Conference on Asia-Pacific Digital Libraries, ICADL 2020*, volume 12504 of *Lecture Notes in Computer Science*, pages 215–231. Springer, November.

Ruta, M., Scioscia, F., De Filippis, D., Ieva, S., Binetti, M., and Di Sciascio, E. (2014). A semantic-enhanced augmented reality tool for openstreetmap poi discovery. *Transportation Research Procedia*, 3:479–488. 17th Meeting of the EURO Working Group on Transportation, EWGT2014, 2-4 July 2014, Sevilla, Spain.

Saha, S. and Mandal, S. (2021). Application of tools to support linked open data. *Library Hi Tech News*, 38(6):21–24, Jan.

Santipantakis, G. M., Glenis, A., Doulkeridis, C., Vlachou, A., and Vouros, G. A. (2019). Stld: Towards a spatio-temporal link discovery framework. In *Proceedings of the International Workshop on Semantic Big Data*, SBD '19, New York, NY, USA. Association for Computing Machinery.

Tedeschi, S., Conia, S., Cecconi, F., and Navigli, R. (2021). Named entity recognition for entity linking: What works and what's next. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 2584–2596.

van Lanen, R. J., van Beek, R., and Kosian, M. C. (2022). A different view on (world) heritage. the need for multi-perspective data analyses in historical landscape studies: The example of schokland (nl). *Journal of Cultural Heritage*, 53:190–205.

Yin, X., Huang, Y., Zhou, B., Li, A., Lan, L., and Jia, Y. (2019). Deep entity linking via eliminating semantic ambiguity with bert. *IEEE Access*, 7:169434–169445.

Zeng, K., Li, C., Hou, L., Li, J., and Feng, L. (2021). A comprehensive survey of entity alignment for knowledge graphs. *AI Open*, 2:1–13.

Zhang, Q., Lu, J., and Jin, Y. (2021). Artificial intelligence in recommender systems. *Complex & Intelligent Systems*, 7(1):439–457, Feb.

Zou, X. (2020). A survey on application of knowledge graph. *Journal of Physics: Conference Series*, 1487(1):012016, mar.

8. Language Resource References