Wikidata's linked data for cultural heritage digital resources: an evaluation based on the Europeana Data Model

Nuno Freire
INESC-ID, Portugal
nuno.freire@tecnico.ulisboa.pt

Antoine Isaac
Europeana Foundation, The Netherlands
Vrije Universiteit Amsterdam, The Netherlands
antoine.isaac@europeana.eu

Abstract

Wikidata is an open data source with many potential applications. Our study aims to evaluate the usability of Wikidata as a linked data source for acquiring richer descriptions of digital objects within the context of Europeana, a data aggregator from the cultural heritage domain. Specifically, we aim to crawl and convert Wikidata using the standard approaches and operations developed for the (Semantic) Web of Data, i.e. using technologies like linked data consumption and RDF(S)/OWL ontology expression and reasoning. We also seek to re-use existing "semantic" specifications, such as conversions to and from generic data models like Schema.org and SKOS. We have developed an experimental set-up and accompanying software to test the feasibility of this approach. We conclude that Wikidata's linked data is able to express an interesting level of semantics for cultural heritage, but quality can still be improved and a human operator still must assist linked data applications to interpret Wikidata's RDF.
Keywords: Wikidata; data quality; cultural heritage; Europeana; Schema.org; Semantic Web.

1. Introduction

Europeana aims to facilitate the usage of digitized Cultural Heritage (CH) resources from and about Europe (Niggemann et al, 2018). It seeks to enable users to access content in all European languages via the Europeana Collections portal[1] and allow applications to use cultural (meta)data via its open APIs. Although many European CH institutions are not yet present in Europeana, it already holds metadata from over 3,700 providers, mostly libraries, museums and archives.

Wikidata is an outstanding data source with many potential applications. It provides its data openly in RDF. It claims - and informal explorations confirm it - that it can be a very interesting source for metadata on CH objects[2].

Wikidata uses an elaborate data model, which supports good-quality data but requires quite some effort for one to master it. The latter may be an obstacle for CH data aggregators that are already operating with limited (human) resources, and searching for more efficient ways to perform their data aggregation needs and remain sustainable.

Our general research aim is to evaluate the usability of Wikidata as a data source for descriptions of digital CH objects, in a data acquisition process that we hope to automate as much as possible. In practice, we tested the hypothesis that Wikidata's RDF data can be crawled and converted efficiently by machine-based methods (robots) operating with the data technologies and practices from Semantic Web and Linked Data (LD), in order to provide useful metadata.

Our approach can be structured in terms of data aggregation use cases that extend each other (FIG. 1 left) and represent different levels of semantic detail in data processing and consumption:

- General (Semantic Web) application. This type of applications processes data solely based on the bottom and middle layers of the Semantic Web Stack (Curé & Blin, 2015). It uses general technologies, especially "meta-languages" such as RDF(S), OWL, and LD-style crawling. It requires the use of certain elements from these languages, e.g., the declaration of

[1] https://europeana.eu
[2] See https://www.wikidata.org/wiki/Wikidata:WikiProject_Cultural_heritage

RDF(S)/OWL classes and properties that enable agents to "understand" at a basic level the ontological structure of the dataset that they must crawl and exploit.
- Cross-domain semantic application, processing data with additional requirements on semantics. Such an application relies on general purpose, widely shared ontologies like Schema.org and SKOS, which lay down common entities like *creative works* and *concepts from a controlled vocabulary*. It must either obtain data that already use the ontologies it can consume, or find ontology alignments (also known as crosswalks or mappings) between these ontologies and the specific classes and properties used the data. It thus requires the ontologies used in the data source to be crawlable and indicate ontology alignments.
- Domain-specific semantic application. This type of application depends on even more specific semantics. It provides functionality to meet the purposes of a domain or application that requires a detailed interpretation of the data, conveyed through a dedicated data model.

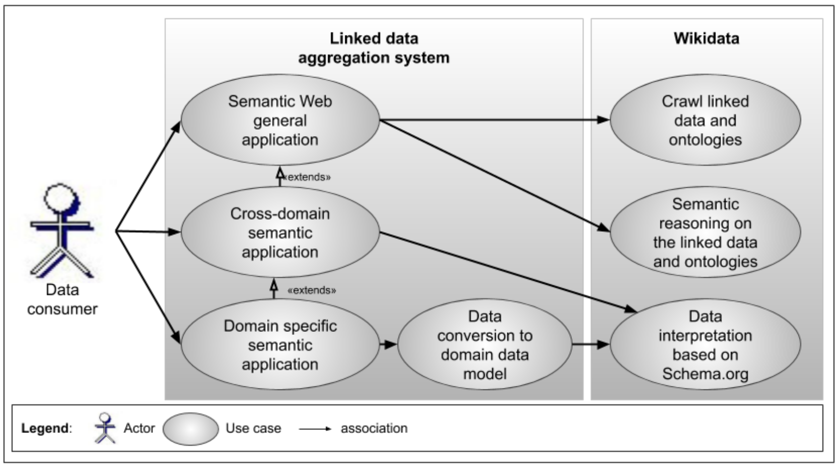

FIG. 1. Use cases of linked data consumption addressed in this study.

Our main motivation lies in the aggregation of metadata for Europeana, which can be seen as a domain-specific (CH) application. In this article, we therefore focus on evaluation for this CH case. The first two cases are still needed for the domain-specific one, though. Especially, we need to rely on Schema.org as an intermediary-level interoperability solution. As Europeana uses the Europeana Data Model (EDM) as its technological solution for data interoperability[3]; we must obtain EDM data from Wikidata. There is however no EDM export from Wikidata, nor an alignment between the Wikidata data model and EDM. We do not foresee being able to maintain such a mapping ourselves, as the Wikidata model is a very extensive and moving target, as Wikidata community members regularly add new elements. On the other hand, there are efforts to connect Wikidata to Schema.org in the Wikidata community. In addition, we have previously proven that Schema.org is suitable for describing CH resources for Europeana and defined mappings between Schema.org and EDM, which we can reuse (Freire et al, 2018c). It seems therefore desirable to use Schema.org as a middle ground to bridge both data spaces.

We have already reported on evaluating Wikidata for tasks from the first two cases (Freire & Isaac, 2019a). However, we repeat some of our observations here, when they provide relevant details or context about how to handle the CH case, or simply when our results have changed, as Wikidata has been updated with more mappings to Schema.org since our earlier study.

We follow, in Section 2, by describing related work on data aggregation based on LD in CH. Section 3 presents the LD crawling set-up for our study. The architecture of our software and

[3] See http://pro.europeana.eu/edm-documentation. EDM is defined collaboratively with all the sectors represented in Europeana. Typically, Europeana data providers rely on ontology alignments to EDM defined by domain experts, who have the necessary know-how to work with the ontologies used in the data.

workflow are explained in Section 4. Section 5 presents the results and our analysis. Section 6 concludes and suggests future work.

2. Related Work

There is a variety of LD research topics related to our work. Scalability is one of the most often addressed ones, with many facets such as indexing, federated querying, and aggregation. The reuse of published LD by third parties has revealed data quality to be a challenge as well, both at the semantic and syntactic levels (Rietveld, 2016; Radulovic et al, 2018; Beek et al, 2017). Significant work has been done to facilitate reuse by aggregation and cleaning (Beek et al, 2016; Fernández et al, 2017). Reasoning on LD is also an active topic - for a comprehensive analysis and description of techniques see Hogan (2014). Finally, generic technical solutions have been proposed for enabling aggregation of LD, for example Rietveld et al. (2015).

The data quality of Wikidata has been extensively evaluated in Farber et al. (2018). This evaluation addresses aspects that do not overlap with our work, however. While we focus on resources for the domain of CH, Farber et al. evaluated Wikidata as general knowledge graph, comparing it with other knowledge graphs. Also related with the evaluation of Wikidata is the work from Ringler & Paulheim (2017), which compares knowledge graphs too, but focusing on their coverage and overlap.

Use of LD in CH has been the focus of much research. Yet the literature addresses mainly data publication (Simou et al, 2017; Hyvönen, 2012; Jones & Seikel, 2016) and not much is devoted to addressing how data aggregation can be performed on top of already published data, let alone to propose standardized procedures for it.

The work of the Dutch Digital Heritage Network[4] (NDE) and the Research and Education Space project[5] (RES) is more similar to ours. NDE is a (still ongoing) national program aiming to increase the social value of the collections of all Dutch libraries, archives and museums. Its initial proposal uses a specific API to enable data providers to centrally register the LD URIs of their resources (Meijer & de Valk, 2017). Being based on an adhoc API, NDE does not yet provide a solution purely based on LD, though. The Research and Education Space project (finished in 2017) has successfully aggregated a considerable number of LD resources from CH collections. The aggregated dataset can be accessed online[6], but the project has not published any evaluation of its aggregation procedures and results.

The work presented in this article and in Freire & Isaac (2019a) is part of research on improving the efficiency and sustainability of data aggregation within the Europeana Network[7]. After identifying LD and Schema.org as promising Internet technology solutions for our aggregation case (Freire et al, 2018a, Freire et al, 2018b), we evaluated the Schema.org metadata from two specific American CH collections, with the help of a Schema.org-EDM mapping (Freire et al, 2018c). We have then begun to define and test (with NDE and the Dutch national library) a workflow and accompanying specifications to aggregate metadata from Europeana's partners using LD and Schema.org (Freire et al, 2019b; Freire et al, 2018d).

3. Experimental setup

Our main objective is to study the usability of Wikidata for the Europeana case, therefore, we focus our evaluation on Wikidata resources that correspond to Europeana (digital) CH resources. FIG. 2 illustrates the setup for our experiment, featuring the data sources, APIs, software components, datasets, dataflows and manual tasks involved. Our main steps are:

[4] https://github.com/netwerk-digitaal-erfgoed/

[5] https://bbcarchdev.github.io/res/

[6] https://bbcarchdev.github.io/res/collections

[7] The Europeana Network is a community of over 2,000 experts with the shared mission to expand and improve access to Europe's digital CH, in their own organizations and projects and/or by contributing to shape Europeana's services. See https://pro.europeana.eu/network-association

1. Identification and harvesting of CH objects described in both Europeana and Wikidata.
   a. Relevant Wikidata items are identified with Wikidata's SPARQL query API, selecting items with statements that use the property *Europeana ID*[8].
   b. We collect the data for the URIs in this subset of Wikidata, using our software for LD crawling. We identified 77,103 Wikidata items with a *Europeana ID*, however, we found that 65,305 of them had invalid or obsolete values, and we removed them[9]. This results in a dataset of 11,798 Wikidata items.
   c. The *Europeana ID* values are transformed to Europeana LD URIs, by adding the same prefix to all of them (Europeana uses a uniform URI structure documented in the *formatter URL* attribute of the *Europeana ID* property).

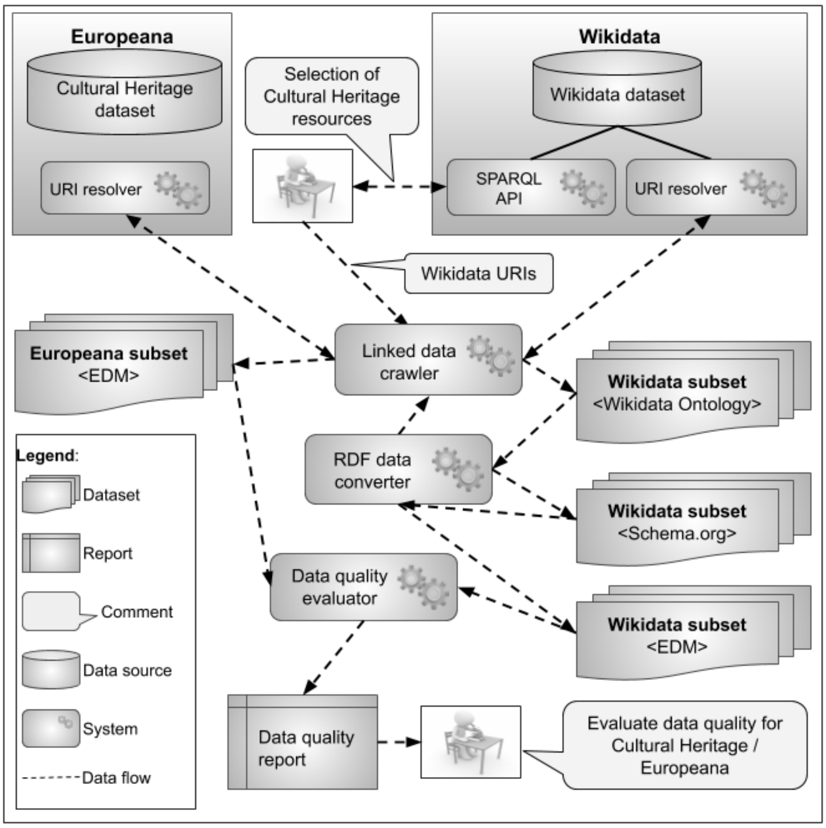

FIG. 2. The experimental setup

2. Conversion of the Wikidata sample to Schema.org using our RDF converter software. This step uses the alignments to Schema.org that exist in the RDF for Wikidata classes and properties[10]. When an alignment to Schema.org cannot be found for a Wikidata class or property, we crawl up the hierarchy of Wikidata's ontology, looking for an equivalence to Schema.org for a more generic Wikidata entity. As soon as an such an equivalence is found, the interpreter assumes that by inheritance the original Wikidata entity can also be mapped to the Schema.org entity found. At this stage we evaluate

[8] https://www.wikidata.org/wiki/Property:P727

[9] We found out that most of the items in our subset derive from the matching work of one (prominent) Wikidata GLAM community member, Martin Dammers. After he loaded links to Europeana in Wikidata, Europeana refreshed two of the biggest datasets concerned, which resulted in changing the Europeana URIs for the objects they contain, an unfortunate case of 'link rot'. Further details can be seen on the Wikidata discussion page https://www.wikidata.org/wiki/Property_talk:P727.

[10] We use the linked data version of the Wikidata Ontology, i.e., for https://www.wikidata.org/entity/Q3305213, we fetch the description available as linked data at this URI, using content negotiation.

Wikidata for the generic Semantic Web and the cross-domain use cases, reporting about the RDF data obtained, the difficulties encountered while crawling Wikidata and automatically interpreting its RDF statements as well as the workarounds we had to implement in our (standards-based) software in order to palliate them. See Freire & Isaac (2019a) for more details.
3. Conversion of the resulting Schema.org data to EDM, using the RDF converter and the alignment between Schema.org and EDM that is specified in Freire et al. (2018c).
4. Collection of the corresponding subset of Europeana, also using our LD crawler, based on the set of Europeana URIs.
5. Comparison and data quality measurement of the two subsets from Wikidata and Europeana in EDM RDF (resulting from conversion or direct harvesting). We apply a software component that implements a metric for the completeness of the data in EDM and generates a report that supports our evaluation of the CH use case.

4. Architecture of the system for linked data aggregation

To perform our study, we have applied and improved experimental software that we develop for Europeana's research on aggregation of CH data, as an extra iteration over our earlier work (Freire et al, 2018c). This software includes components for data crawling, reasoning on ontologies, data conversion and data analysis. The architecture of the system supporting the execution of our experiment is illustrated in FIG. 3. It is composed of 5 subsystems, supported by data repositories and (algorithms implementing) relevant specifications[11]:

- Workflow engine. This system allows a human operator to coordinate and monitor the execution of the workflow for the experiment.
- Linked data crawler. This HTTP crawler implements the specifications and best practices of LD, such as content negotiation. It also checks instruction in robots.txt files.
- Linked data interpreter. This system performs semantic reasoning, allowing to interpret data models according to their use of data modeling constructs and axioms from RDF(S), OWL and SKOS. It exploits *resource types (esp. classes and properties)*, *type equivalence, class hierarchy and inheritance,* and *property hierarchy and inheritance.* It uses the definition of the data to populate a repository of statements (i.e. a triple store) to align data models. This repository may also be populated with statements added by an operator.
- RDF data converter. This system transforms an RDF graph into another RDF graph by executing data conversion operations, based on the specifications of alignments between the models and vocabularies used in these graphs. These specifications are represented in a machine-actionable form that allow the converter to create detailed reports about the conversion of a dataset, for example for listing properties that are used in the original RDF graph but not mapped. These reports assist data experts in refining mappings and in identifying and fixing of data quality issues (missing data, invalid data types, etc.) (Freire et al, 2018c).
- Data quality evaluator. This system continues on our research on evaluation of data quality in CH (Király et al, 2018). Based on plugins that implement particular metrics or validations, it generates reports for aggregators to evaluate the fitness of a dataset for specific purposes. In this study, we implemented a first plugin for measuring data completeness using a measure discussed within the Europeana Data Quality Committee[12] that identifies most important EDM data elements for providing information and functionality to end-users in the Europeana portal. This metric checks in an EDM record for the existence of 16 properties or groups of related properties (where the occurrence of a group is fulfilled by the occurrence of just one of its members). It outputs a score in the [0,1] range. We developed a second plugin for EDM valida-

[11] The software is developed within the data aggregation experimental software that supports several of our research activities in the context of CH data. Its development is open source and made available at https://github.com/nfreire/data-aggregation-lab

[12] https://pro.europeana.eu/project/data-quality-committee

tion, applying the regular procedure performed by Europeana based on Schematron validation rules[13]. We simplified the Europeana validation for our evaluation, as it requires administrative data that is defined by its data providers and therefore cannot be found in any Wikidata record. We removed three validation rules: the requirements for the name of the (original) data provider (*edm:dataProvider*) and the name of the metadata aggregator (*edm:provider*); and the requirement for rights metadata using one of the 14 rights statements currently supported by Europeana (*edm:rights*)[14]. It shall be clear that this simplification is meant to be specific to this experiment, where we do not seek to replicate a full data ingestion run into Europeana and already know that the data source does not include these data elements[15]. A production-level data acquisition process for Europeana shall include a way to source these missing elements.

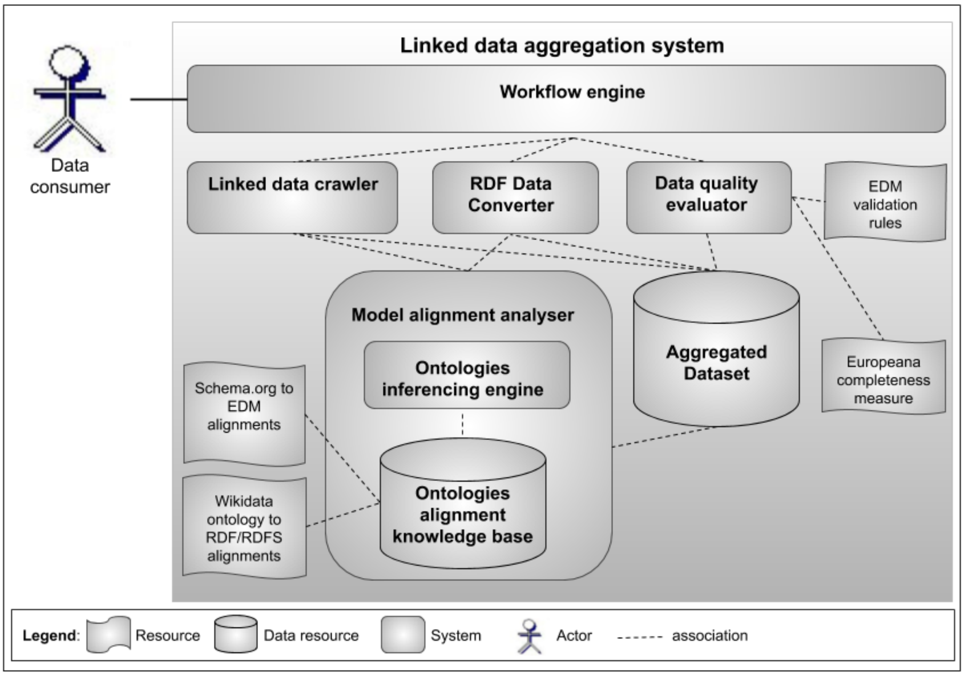

FIG. 3. High level system architecture of the linked data aggregation system

5. Results and Analysis

We present here the results of the study, as per our last check on 27 July 2019. We observed that Wikidata's RDF presents some difficulties for our cases. First, Wikidata uses properties from its own (meta-level) ontology instead of equivalent properties from the data modelling languages RDF(S), OWL or SKOS[16]. Especially, most typing statements use the specific Wikidata property *wdt:P31* and the standard *rdf:type*[17] is not used in semantically meaningful statements. Besides, the RDF data defining these Wikidata properties is not accessible in a standard way: Wikidata's URIs are not always resolvable. For example, *wdt:P31* is referred in the data as

[13] See the part on validation at https://pro.europeana.eu/resources/standardization-tools/edm-documentation

[14] See https://pro.europeana.eu/page/available-rights-statements

[15] This may change, however. Especially, Wikidata has recently minted a property to refer to the rights statements required by Europeana (https://rightsstatements.org/en/2019/05/rightsstatements-in-wikidata.html).

[16] See (Erxleben et al, 2014) for details on Wikidata's expression in RDF.

[17] For readability purposes, we abbreviate namespaces as follows: rdf for http://www.w3.org/1999/02/22-rdf-syntax-ns#; wdt for http://www.wikidata.org/prop/direct/; edm for http://www.europeana.eu/schemas/edm/; dc for http://purl.org/dc/elements/1.1/; dct for http://purl.org/dc/terms/.

http://www.wikidata.org/prop/direct/P31, which is not resolvable. One needs to find and resolve *http://www.wikidata.org/entity/P31* instead.

Such unresolvable namespaces hamper automated data interpretation. Without human intervention, it is impossible for a general application to use the data when it requires to identify the classes used or to interpret class and property hierarchies following the standard Semantic Web frameworks. We thus manually added the essential equivalence statements in our knowledge base. In fact, most of the equivalences we need (6 out of 8) are present in Wikidata, but they are expressed using meta-level predicates from the Wikidata namespace. Additional details of this work can be consulted in (Freire & Isaac, 2019a).

This issue also hinders our second use case. Wikidata's RDF uses a very limited number of machine-oriented "semantic" properties (i.e., those that do not have string values, which includes the properties aligning Wikidata with other data models as well as the ones used to create richly interlinked instance-level metadata) from standard namespaces; "semantic" properties mostly come from the Wikidata namespace. Most of the standard properties used, as reported in Freire & Isaac (2019a), are simple labels. This could be palliated by exploiting alignments between Wikidata's own properties and more standard ones like Schema.org. But agents cannot easily get these alignments in Wikidata as they are (1) not accessible via (LD style) resolving of the properties they align; (2) expressed using Wikidata's own metamodeling properties. The work above allows to bypass the metamodeling issue. And it was possible for us to continue the study by adapting the code of the LD interpreter to convert the URIs to the ones that Wikidata is able to resolve.

That adaptation is arguably very specific to one data source, but it allowed us to acquire Schema.org data from the use equivalence relations stated in the RDF of the classes and properties in Wikidata, exploiting the ontology hierarchy as presented in Section 3.

The listing of the individual ontology alignments to Schema.org that we found in Wikidata is available online[18]. Additional details and statistics that we computed for these alignments may be consulted in (Freire & Isaac, 2019a). In general, we found alignments for around 27.4% of the ontology elements used in the subset of Wikidata we use. We found 102 distinct classes in use in the dataset, 57 (55.9%) of which had alignments to Schema.org. We found 266 properties in use in the dataset, 44 (16.5%) of which had alignments to Schema.org. We consider these results to be indicative that many applications can use the structured data, and they allowed us to proceed to the evaluation of our third use case and test the semantic value of the resulting data for CH.

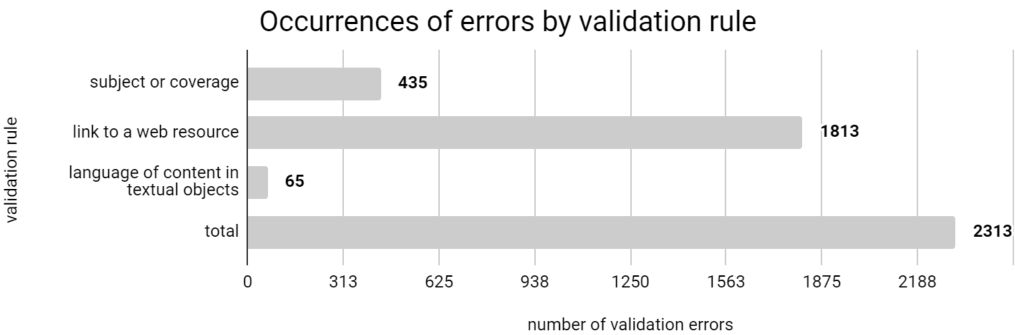

FIG. 4. EDM validation rules not fulfilled by the data from Wikidata (after conversion from Schema.org to EDM)

After the conversion of the Wikidata subset into EDM, we have found 2313 validation rule failures. These concern 3 individual validation rules, as shown in FIG. 4[19]: 1813 records did not

[18] https://github.com/nfreire/data-aggregation-lab/blob/master/data-aggregation-casestudies/documentation/wikidata/SchemaOrg-ontology-alignments-listing.md

[19] For readability purposes we abbreviated the rule descriptions. Formally, they correspond to the Schematron rules 'dc:subject or dc:type or dc:coverage or dct:temporal or dct:spatial', '(edm:isShownAt and exists(edm:isShownAt/@rdf:resource)) or (edm:isShownBy and exists(edm:isShownBy/@rdf:resource))' and 'not(edm:type='TEXT') or (edm:type='TEXT' and exists(dc:language))'.

contain a link to a digital object; 435 records did not contain any data about the subject or the coverage of the CH object; and 65 records about textual CH objects (books, newspapers, etc.) did not state the language of the text. These validation errors occurred in 1906 individual records from Wikidata, therefore, 83.7% of the dataset validated successfully. 12.8% of the records failed in just one of the rules, 3.5% failed in 2 rules, and 1 record failed all three rules.

We then compared the two datasets we have harvested, calculating the metadata completeness metric of each record. The results obtained for the Wikidata dataset were clearly inferior to those for the existing data in Europeana, as shown in FIG. 5. The average completeness scores were 0.74 and 0.51 in the Europeana and Wikidata datasets, respectively.

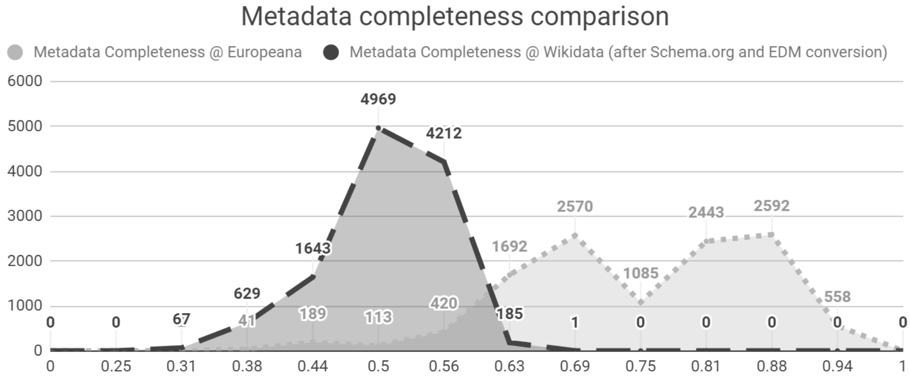

FIG. 5. Results of the comparison between the data at Europeana and the data from Wikidata (after conversion from Schema.org to EDM), using a metadata completeness measure designed within the Europeana Data Quality Committee.

In our opinion, these are not negative results, however. Many Wikidata items have metadata completeness scores that are higher than Europeana's lowest-quality records: the average completeness score in the Wikidata dataset was higher than the completeness score of 2.94% of the records in the Europeana dataset. Furthermore, Europeana runs in its data ingestion workflow several tasks of data normalization and enrichment, and also organizes data curation activities with its partners, resulting in higher quality in its aggregated dataset. Finally, our current completeness measure fails to capture all of Wikidata's potential. As it is merely based on the presence or absence of fields for an object, it does not consider the "semantic richness" of a field's value, i.e., when this value is a Wikidata URI for another item (a person, a concept), which could bring more useful data, e.g., multilingual labels that are key for providing a better Europeana service. In fact, some of the abovementioned enrichment at Europeana relies on automatic linking to third-party semantic resources, including Wikidata, which are collected in a knowledge graph (Europeana's "Entity Collection") (Charles et al, 2018). It would be interesting to see whether the data available for objects present in Wikidata can complement - or just confirm - existing Europeana enrichments.

6. Conclusion and future work

The general conclusions from this study are that Wikidata's LD is able to express an interesting level of data semantics for CH, but currently a human operator must assist LD applications to interpret Wikidata's RDF. This requires training on Wikidata's data model and its expression in RDF. Technical problems also exist in resolving some Wikidata URIs, which need to be addressed by human intervention by adapting generic tools to Wikidata's specific practices.

In more detail, we argue that Wikidata is only a few steps away from appropriate machine interpretability, as it contains enough alignments to RDF(S), OWL, SKOS and Schema.org. Unfortunately, the semantics of the Wikidata's RDF output are somehow *locked-in* by the usage of predicates from Wikidata's own (meta-)ontology, making them uninterpretable for data crawlers based on general Semantic Web data processing approaches.

Since during our evaluation the data has undergone a conversion to Schema.org, we also argue that our study provides evidence that cross-domain applications can obtain enough semantics from Wikidata for their purposes.

Regarding the completeness of data, the Wikidata dataset scores clearly lower than the existing metadata in Europeana. Nevertheless, a reasonable subset of Wikidata was able to express an interesting level of semantics for CH. And in real-world cases, quality can still be improved, as in Europeana's own data ingestion process. In any case, Wikidata could be conceived as an interesting complement to Europeana's existing metadata.

One may argue that our tests focus too much on Europeana and are biased against third-party metadata. We claim that our quality criteria are not too Europeana-specific, though. Our measure is based on the presence of fields that are relevant for many applications (and present in many datasets) in the CH domain - this is intrinsic to the community-based design process of EDM. The validation procedures test the presence of fields according to rules that have been vetted by our community, via the Europeana Data Quality Committee[20]. And for our experiment, we have lowered the emphasis on Europeana by removing the three validation rules that cannot be fulfilled by any Wikidata record. We will however remain open to using third party (including Wikidata-originated) quality tests, should they be relevant to CH. We are aware of some efforts, but none of them are designed for CH data. In the meantime, we will refine our completeness measure by incorporating recent progress of the Europeana Data Quality Committee. The next version of the measure will enable us to better reflect the presence of contextual entities and multilingual data in Wikidata[21].

In parallel, we will further scrutinize whether the dual conversion process (from Wikidata to Schema.org and then from Schema.org to EDM) has unfairly reduced Wikidata's quality, by checking whether some properties rich in semantics (i.e., linking an item to others) have been left out. We will focus especially on the first step, as we are rather confident that our Schema.org-EDM mapping includes all the Schema.org properties relevant for Europeana[22].

In addition, we will seek to conduct larger-scale tests: the subset of Wikidata we have exploited is relatively small, especially after discarding the items with dead Europeana links. As explained, a large part of the subset derives from the one-time work of one Wikidata community member. We claim this does not necessarily make our dataset biased. We have measured that it includes objects from 70 datasets in Europeana and 201 CH collections or institutions in Wikidata, which is fairly diverse. Still, we could check that these different components are represented in a balanced way. Furthermore, the Wikidata-Europeana alignment that was once made could be run again, possibly using different criteria for matching Wikidata items with Europeana objects, and hopefully resulting in more matches. Finally, we plan to perform studies on LD published by data providers from the Europeana network, also focusing on the aspects of machine interoperability and data interpretability.

Acknowledgements

This work was partly supported by Portuguese national funds through Fundação para a Ciência e a Tecnologia (FCT) with reference UID/CEC/50021/2019 and by the European Commission under contract number 30-CE-0885387/00-80. We thank the DCMI reviewers for their comments, which helped to improve the article, and our colleagues Valentine Charles, Hugo Manguinhas and Liam Wyatt for their help during the design of the various elements of our study.

[20] For additional details see the parts on mandatory metadata in *Data quality committee - 2016 report*, http://pro.europeana.eu/files/Europeana_Professional/EuropeanaTech/EuropeanaTech_WG/DataQualityCommittee/DataQualityCommittee-2016Report.pdf

[21] For more hints about these directions, see the most recent release of the Europeana Publishing Framework at https://pro.europeana.eu/post/publishing-framework

[22] We shall note in fact that if the mapping from Wikidata to Schema.org misses relevant data elements, it can also be seen as a Wikidata quality issue, as these mappings are sourced in the descriptions of Wikidata properties.

References

Beek W., L. Rietveld, F. Ilievski, S. Schlobach. (2017). LOD Lab: Scalable Linked Data Processing. In: Reasoning Web: Logical Foundation of Knowledge Graph Construction and Query Answering. Reasoning Web 2016. Lecture Notes in Computer Science, vol 9885. Springer, Cham.

Beek, W., L. Rietveld, S. Schlobach, F. van Harmelen. (2016). LOD Laundromat: Why the Semantic Web Needs Centralization (Even If We Don't Like It). In: IEEE Internet Computing, 20(2). IEEE.

Charles, V., H. Manguinhas, A. Isaac, N. Freire and S. Gordea (2018). Designing a Multilingual Knowledge Graph as a Service for Cultural Heritage – Some Challenges and Solutions. Proceedings of the International Conference on Dublin Core and Metadata Applications, 2018.

Curé, O., G. Blin. (eds). (2015). Chapter Three - RDF and the Semantic Web Stack. In RDF Database Systems, pp 41-80. Morgan Kaufmann.

Erxleben, F., M. Günther, M. Krötzsch, J. Mendez, D. Vrandečić. (2014). Introducing Wikidata to the Linked Data Web. In: The Semantic Web – ISWC 2014, pp. 50–65. Lecture Notes in Computer Science, vol 8796. Springer, Cham.

Färber, M., Bartscherer, F., Menne, C., and Rettinger, A. (2018). Linked data quality of DBpedia, Freebase, OpenCyc, Wikidata, and YAGO. Semantic Web, 9(1), 77-129. doi:10.3233/SW-170275

Fernández J.D., W. Beek, M.A. Martínez-Prieto, M. Arias. (2017). LOD-a-lot. In: The Semantic Web – ISWC 2017. Lecture Notes in Computer Science, vol 10588. Springer, Cham.

Freire, N., H. Manguinhas, A. Isaac, G. Robson, J. B. Howard. (2018a). Web technologies: a survey of their applicability to metadata aggregation in cultural heritage. In: Information Services & Use Journal, 37(4), Expanding Perspectives on Open Science: Communities, Cultures and Diversity in Concepts and Practices.

Freire, N., G. Robson, J.B. Howard, H. Manguinhas, A. Isaac. (2018b). Cultural heritage metadata aggregation using web technologies: IIIF, Sitemaps and Schema.org. In International Journal on Digital Libraries, 20(2). Springer.

Freire, N., V. Charles A. Isaac. (2018c). Evaluation of Schema.org for Aggregation of Cultural Heritage Metadata. In: The Semantic Web. ESWC 2018. Lecture Notes in Computer Science, vol 10843. Springer, Cham.

Freire, N., E. Meijers, R. Voorburg, A. Isaac. (2018d). Aggregation of cultural heritage datasets through the Web of Data. In: SEMANTiCS 2018 – 14th International Conference on Semantic Systems Conference, Elsevier B.V..

Freire, N., A. Isaac. (2019a). Technical usability of Wikidata's linked data: evaluation of machine interoperability and data interpretability. In: W. Abramowicz, A. Paschke (eds.). Lecture Notes in Business Information Processing. Springer, Cham.

Freire, N., E. Meijers, R. Voorburg, R. Cornelissen, A. Isaac, S. de Valk. (2019b). Aggregation of Linked Data: A case study in the cultural heritage domain. In: Information, 10(8), MPDI. doi:10.3390/info10080252

Hogan, A. (2014). Reasoning Techniques for the Web of Data. In: Studies on the Semantic Web, 19. IOS Press.

Hyvönen, E. (2012). Publishing and Using Cultural Heritage Linked Data on the Semantic Web. In: Synthesis Lectures on the Semantic Web: Theory and Technology. Morgan & Claypool.

Jones, E., M. Seikel (eds). (2016). Linked Data for Cultural Heritage. Facet Publishing.

Király, P., J. Stiller, V. Charles, V. Bailer, N. Freire. (2018). Evaluating Data Quality in Europeana: Metrics for Multilinguality. In: Metadata and Semantic Research. MTSR 2018. Communications in Computer and Information Science, vol 846. Springer, Cham.

Meijers, E., S. de Valk. (2017). A distributed network of heritage information. White paper.

Niggemann, E., J. Cousins, M. Sanderhoff. (2018). Europeana Business Plan 2018 'Democratizing culture'. Europeana Foundation.

Radulovic, F., N. Mihindukulasooriya, R. García-Castro, A. Gomez-Pérez. (2018). A comprehensive quality model for Linked Data. In: Semantic Web, 9(1). IOS Press.

Rietveld L., R. Verborgh, W. Beek. M. Vander Sande, S. Schlobach. (2015). Linked Data-as-a-Service: The Semantic Web Redeployed. In: The Semantic Web. Latest Advances and New Domains. ESWC 2015. Lecture Notes in Computer Science, vol 9088. Springer, Cham.

Rietveld, L.. (2016). Publishing and Consuming Linked Data: Optimizing for the Unknown. In: Studies on the Semantic Web, vol. 21. IOS Press.

Ringler, D., and Paulheim, H. (2017). One Knowledge Graph to Rule Them All? Analyzing the Differences Between DBpedia, YAGO, Wikidata & co. In: KI 2017: Advances in Artificial Intelligence. Lecture Notes in Computer Science, vol 10505. Springer, Cham. doi:10.1007/978-3-319-67190-1_33

Simou, N., A. Chortaras, G. Stamou, S. Kollias. (2017). Enriching and Publishing Cultural Heritage as Linked Open Data. In: Mixed Reality and Gamification for Cultural Heritage pp 201-223. Springer, Cham.