Children's art museum collections as Linked Open Data

Konstantinos Kotis* and
Sotiris Angelis

Department of Cultural Technology and Communication,
University of the Aegean,
Lesvos, Mytilene, Greece
Email: kotis@aegean.gr
Email: sotiris@aegean.gr
*Corresponding author

Maria Chondrogianni and
Efstathia Marini

Department of Cultural Technology and Communication,
University of the Aegean,
Lesvos, Mytilene, Greece
and
Greek Children's Art Museum,
Plaka, Athens, Greece
Email: ctm19027@ct.aegean.gr
Email: ctm19017@ct.aegean.gr

Abstract: It has been recently argued that it is rather beneficial to cultural institutions to provide their datasets as Linked Open Data, to achieve cross-referencing, interlinking, and integration with other datasets in the LOD cloud. In this paper, we present the Greek Children's Art Museum (GCAM) linked dataset, along with dataset and vocabulary statistics, as well as lessons learned from the process of transforming the collections to HTML-embedded structured data using the Europeana Data Model and the Schema.org model. The dataset consists of three cultural collections of 121 child artworks (paintings), including detailed descriptions and interlinks to external datasets. In addition to the presentation of GCAM data and the lessons learned from the experimentation of non-ICT experts with LOD paradigm, the paper introduces a new metric for measuring datasets quality in terms of links to and from other datasets.

Keywords: Linked Open Data; HTML-embedded RDF; RDFa; microdata; museum; artwork.

Reference to this paper should be made as follows: Kotis, K., Angelis, S., Chondrogianni, M. and Marini, E. (2021) 'Children's art museum collections as Linked Open Data', *Int. J. Metadata Semantics and Ontologies*, Vol. 15, No. 1, pp.60–70.

Biographical notes: Konstantinos Kotis is currently an Assistant Professor in the Department of Cultural Technology and Communication and in Intelligent Systems Lab at the University of the Aegean, and a Research Associate in the Department of Digital Systems and in AI Lab at the University of Piraeus. His research interests include knowledge/ontology engineering, Semantic Web technologies, semantic data management, the web of things and conversational AI. He has published more than 60 papers in peer-reviewed international journals and conferences, and has served as a reviewer and PC member in several journals and conference events. He has also contributed to several national and European projects in different roles/positions. For more, please visit: http://i-lab.aegean.gr/kotis.

Sotiris Angelis is currently a PhD candidate and a Member of i-Lab in the Department of Cultural Technology and Communication at the University of the Aegean. His research interests focus on Semantic Web, cultural data integration, semantic trajectories and recommendation systems. He received his Master's degree in "Cultural Informatics and Communication" from the same Department, where he studied web, mobile application and game design. He holds a Dipl.-Ing. degree in the Department of Electrical and Computer Engineering from Aristotle University of Thessaloniki. His background includes experience in software engineering, web development, and digital systems.

Maria Chondrogianni is an MSc student in the Master's Program "Cultural Informatics and Communications" in the Department of Cultural Technology and Communication at the University of the Aegean. She holds a Bachelor's degree in Theology-Theological Studies of National and Kapodistrian from University of Athens. Her background includes digitising cultural collections (Harokopeion University-seminar) and planning museum educational programs (National Kapodistrian University of Athens-e-learning program). Currently, she is Volunteering in the Greek Children's Art Museum in Athens, Greece.

Efstathia Marini is an MSc student in the Master's Program "Cultural Informatics and Communications" in the Department of Cultural Technology and Communication at the University of the Aegean. She holds a Bachelor's degree in Greek Cultural Studies from the Hellenic Open University, Greece (2017). Currently, she is working in the Greek Children's Art Museum in Athens, Greece.

1 Introduction

Museums generate a vast amount of cultural data, while documenting artworks, artists, and cultural heritage objects. Recently, there have been numerous attempts to take advantage of Semantic Web technologies to link and open such data to the public as Linked Open Data (LOD) (Angelis and Kotis, 2020; Dijkshoorn et al., 2018; Dragoni et al., 2016; Koho et al., 2020; Skjæveland, 2011; Szekely et al., 2013). Based on the lessons learned from this experience, this paper presents related work conducted by our lab in collaboration with the Greek Children's Art Museum (GCAM) in Athens. More specifically, the work aims at utilising Semantic Web technologies in order to transform the museum of children's artwork collections into linked open data, semantically described as *Child Art*,[1] *Creative Work*,[2] and its subclass *Painting*,[3] and the class of Provided Cultural Heritage Object.[4]

The work that is being presented has been mainly motivated by an experiment conducted within the context of a Master's Thesis project in the Cultural Informatics postgraduate program in our department. More specifically, students of our lab and a curator of GCAM, with no strong Information and Communication Technology (ICT) skills and background, were assigned the task to get familiar with the Linked Data paradigm, learn HTML5 (W3C Recommendation, 2017), RDFa (Hermann et al., 2015) and Microdata (Hickson, 2013) technologies, and be able to develop HTML pages (one for each artwork of the GCAM collections) with embedded RDF structured data. By using two different encodings of RDF i.e., RDFa and Microdata, the researchers' objectives were: a) to investigate how difficult it is for non-experts to understand the LOD paradigm and related processing to generate linked open data (expertise dependency/constraint); b) to estimate how difficult it is for non-experts to use the provided technology i.e., HTML5, RDFa and Microdata (technology dependency/constraint) and c) when given specific models/ vocabularies for the semantic annotation of the provided data i.e., EDM (European Union, 2014) and Schema.org ("Schema.org," 2021), to report how difficult it is to perform this process with the one or the other model (model dependency/constraint).

Based on this initial motivation, as well as on the later goal to actually deliver a set of GCAM artworks as LOD, the contribution of our work is two-fold: a) shed insight into the lessons learned during the process of transforming the collections to HTML-embedded linked data (in RDFa and Microdata), while using mainly two different models i.e., EDM and Schema.org; and b) contribute to the growth of LOD cloud by interlinking children's art (paintings) work with several external datasets/entities of well-known and widely-used resources (e.g., DBpedia, Wikidata, WikiArt, MoMA). This paper presents the Greek Children's Art Museum[5] dataset, accessible from the museum's server DCAT description webpage (Chondrogianni et al., 2021a), as well as from Zenodo (Chondrogianni et al., 2021b), the Google Dataset Search and the Lod-Cloud (Chondrogianni et al., 2021c).

The structure of the paper is as follows: Sections 2 and 3 present related works on linking data in the cultural domain and the research goals of this paper, respectively. Section 4 describes the artwork collections, the vocabularies used for their semantic annotation, and the datasets selected to link to external resources. Section 5 presents the transformation process (data to linked open data), and Section 6 provides detailed information about the generated dataset. Section 7 presents an example of a semantically annotated artifact as linked data, and the known shortcomings of this work. Finally, Section 8 concludes the paper.

2 Related work

In this section, the paper presents related work on linking data in the cultural domain. More specifically, it reports both on recent efforts that contribute to the extension of LOD cloud and on research approaches followed towards meeting the requirements for 5-stars LOD and FAIR data. The related approaches are also driven by requirements related to the discovery of links between RDF encoded structured data and external datasets.

Dijkshoorn et al. (2018) presented the linked dataset of the Rijksmuseum, along with collection and vocabulary statistics, and the lessons learned from the process of converting the collection into Linked Data. It contains over 350,000 objects, including detailed descriptions and high-quality images released under a public domain license. The dataset uses the EDM vocabulary, which is LOD vocabulary compliant and

its 5-stars LOD. The transformation/conversion of data into RDF is accomplished, when an Extensible Stylesheet Language Transformations (XSLT) file is used via an API running over a database.

Koho et al. (2020) presented the WarSampo Knowledge Graph (KG) and a LOD service for publishing data about World War II, with a focus on Finnish military history. The used metadata schema is an extension of CIDOC-CRM, supplemented by various military history domain ontologies. A repeatable automatic data transformation and linking pipeline has been created to rebuild the whole WarSampo KG from individual source datasets. A query API in SPARQL and other services support the reuse of this dataset (14 million triples in use in nine end-user application views of the WarSampo portal, with over 690,000 end-users since 2015). The dataset is 5-stars LOD and LOD vocabulary compliant.

Dragoni et al. (2016) presented the process of mapping metadata from small size collections with little-known artists to the LOD cloud, thus exposing the acquired knowledge base by using LOD format and making it available to third-party services. A framework is proposed of: a) how to convert data into RDF with the use of a custom ontology that extends the Europeana Data Model and b) how to enrich metadata from content found on the web. The metadata enrichment is then succeeded when external knowledge bases such as DBpedia is linked with the use of Natural Language Processing (NLP) methods so that information is extracted from web pages containing relevant details for the records.

Szekely et al. (2013) described the process and the lessons learned in mapping the metadata of the Smithsonian American Art Museum (SAAM) objects to a custom SAAM ontology, that extends EDM with subclasses and sub-properties to represent attributes unique to SAAM. In this work, the mapping is performed semi-automatically using KARMA integration tool, while the linking is being focused on the interrelation of the museum artists with entities in DBpedia, Getty Union List of Artist Names and the Rijks museum dataset. The paper also discusses the challenges encountered in data preparation, while the researchers were mapping columns to classes, connecting the classes and mapping based on field values.

In the work by Skjæveland (2011), the conversion of the Yellow List of architecturally and culturally valuable buildings in Oslo from Excel spreadsheets into Linked Open Data is presented. The data was mapped in the CIDOC-CRM and converted by XLWrap4 package, a spreadsheet-to-RDF wrapper. The entities of the RDF graph were linked to DBpedia and exposed via a SPARQL endpoint based on a Joseki triple store. A web application that plots the results of SPARQL queries on a Google Map based on the coordinates in the Yellow List was developed to help users to explore the data.

The work that is being presented in this section is only a small fragment of related work on recent developments in metadata enriching, linking and interoperability of data. As reported in the latest editorial of Alemu and Garoufallou (2020), special issue on the future of interlinked, interoperable and scalable metadata, several institutions (e.g., libraries and archives) continuously devise new models and metadata presentation formats, which are made available as RDF and which aim to represent data in granular structures and define unique identification protocols such as URIs. Further related work on this active research topic includes, but not limited to, the work by Gareta (2015), and the work by Faraj and Micsik (2019).

3 Research goals

The related work presented in Section 2 starts with raw cultural data and converts them into RDF and LOD using a number of different conversion/transformation methods and models. Going beyond related work, the goal of the work presented in this paper is two-fold: a) to use a relatively small dataset (the Greek Children's Art Museum dataset) in order to experiment with a non-automated process of transforming artwork collections into RDF structured data for LOD cloud, using different LOD technologies and non-ICT experts and b) to investigate metrics for linked data analysis and propose new ones where needed.

Towards these goals, the paper reports on the lessons learned from the RDFization process of transforming artwork collections into RDF structured data that are embedded in HTML documents, conducted by non-ICT experts, and the latter's experience in linking the generated data to other datasets (linking process). Furthermore, the paper explores the potential impact of new metrics on measuring the linking dynamics of a dataset, going beyond standard measurements such as average, min and max links of resources with external ones. More specifically, the paper presents the quality metric of *D-h-index(in→out)* and *D-h-index(out→in)*.

4 Collections, vocabularies and datasets

Although a large number of artworks are continuously created in the museum, due to limited resources (and to the Covid-19 pandemic), the GCAM's artwork selected for this experimental work concerns 121 paintings organised into three collections: a) a collection of paintings inspired by artists; b) a collection of paintings inspired by the environment (nature) and c) a collection of paintings inspired by Greek and international cultural heritage. All 121 artworks have been created by children in the premises of the museum and digitised in *.jpg* format.

The metadata for semantically annotating and structuring the data of the 121 artworks were taken from the well-known and widely used models/vocabularies, that is, European Data Model (EDM) and Dublin Core (DC) as it is significantly reused by EDM, Schema.org, and Friend-of-a-Friend (FOAF). Although other related vocabularies are available for semantically annotating museum data (e.g., LIDO (COM-CIDOC LIDO Working Group, 2020), Linked Art (Community, 2020), this choice is made by the extensive use of EDM and Schema.org in several similar cultural projects and cultural aggregators (e.g., Europeana and SearchCulture.gr cultural data aggregators).

The namespaces of the metadata and the used prefixes are the following:

rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
dc: <http://purl.org/dc/terms/>
foaf: <http://xmlns.com/foaf/0.1/>
schema: <http://schema.org/>
edm: <http://www.europeana.eu/schemas/edm/>

To link the data of the 121 artworks to other resources, several well-known and widely used external LOD datasets were used were used:

- DBpedia: http://dbpedia.org/resource/
- Wikidata: http://www.wikidata.org/wiki/
- WikiArt: https://www.wikiart.org/
- Wikimedia: https://commons.wikimedia.org/wiki/
- MoMA: https://www.moma.org/collection/works/

5 The transformation process

The process of transforming artworks into semantically annotated digital artworks, embedding structured data to HTML webpages, was performed manually, while Resource Description Framework (RDF) and a text editor (Notepad++) were being used. Depending on the RDF serialisation/encoding used (RDFa or Microdata), participants were consulting the chosen model/vocabulary for the appropriate metadata to use for every property that GCAM curators proposed to describe their data. To validate the encodings during the process, a number of tools/environments were used, such as the RDFa Play environment,[6] RDFa and Microdata validators, converters and distillers (e.g., W3C RDF validator,[7] EasyRDF,[8] Ruby RDF distiller,[9] Google structure data testing tool[10]).

The mapping of the data (corresponding to the 121 artworks) to the metadata selected from the used models/ vocabularies is presented in the following sections. As stated above, driven by the goal of experimentation with these models and non-ICT experts, participants' selection of metadata was based on the extensive study of the models and their correspondence to specific museum data i.e., data semantically related to *child art*, *artwork*, *creative work* and *paintings*.

5.1 Schema.org and FOAF

The models of Schema.org and FOAF were used for the semantic description of 73 (out of 121) artifacts, classified as (of type) paintings (*schema:Painting*). Table 1 outlines the related metadata with a short description.

5.2 EDM, DC and FOAF

The models of EDM (with the extensively reused DC) and FOAF were used for 48 (out of 121) artifacts, classified as (of type) provided cultural heritage objects (*edm:providedCHO*). Table 2 outlines the related metadata with a short description.

As presented in Tables 1 and 2, the selected metadata of Schema.org model were mapped (one-to-one) to the related metadata of EDM/DC model. Hence, *schema:creator* corresponds to *dc:creator*, *schema:about* corresponds to *dc:abstract*, *educationalUse* corresponds to *dc:educationLevel*, and so on. Such a design choice was made to preserve the same quantity (number) of metadata for each subset and collection, as well as the same (or similar) quality (semantics).

Table 1 Schema.org selected metadata

rdf:type	*schema:Painting (Thing → CreativeWork → Painting)*
creator	The name of the creator (child's name).
about	General description of the artwork, composed by the type of artwork (painting), the creator's name (the child's name), the creator's age, when the artifact was created, the name of the original artwork it was inspired by, and the framework it was created within.
educationalUse	The purpose of a work within the context of education (educational framework).
dateCreated	The year that the painting is created.
material	The material used for the creation of the artifact.
copyrightHolder	The person/organisation that holds the legal copyright of the artifact.
maintainer	The person/organisation that distributes and manages the publication of the artefact; e.g., Guernica is exhibited in the Museo Reina Sofia.
isBasedOn	A painting from which this artifact is derived; e.g., "Las Meninas" (Picasso) is based on the painting "Las Meninas" by Diego Velázquez.
citation	A citation to another artist.
locationCreated	The location that the artifact is created.
genre	The genre of the artefact; e.g., Guernica is an example of Cubism.
image	The digital image file of the painting (URL).
foaf:page	The HTML file with the embedded RDF representation of the painting (URL).
foaf:depiction	Additional to schema:image.

Table 2 EDM/DC selected metadata

rdf:type	*edm:providedCHO*
dc:creator	The name of the creator (child's name).
dc:abstract	General description of the artwork, composed by the type of artwork (painting), the creator's name (the child's name), the creator's age, when the artifact was created, the name of the original artwork it was inspired by, and the framework it was created within.
dc:educationLevel	An educational or training context for which the artifact is intended.
edm:year	The year that the object is created.
dc:medium	The material used for the creation of the artifact.
dc:rights	Information about rights held in and over the artifact.
edm:provider	The person/organisation that distributes and manages the publication of the artifact.
edm:isSimilarTo	A painting to which this artifact is similar; e.g., "Las Meninas" (Picasso) is similar to the painting "Las Meninas" by Diego Velázquez.
edm:isRelatedTo	The artist that the artifact is related to.
edm:currentLocation	The current location of the artifact.
dc:type	The type of the artefact; e.g., Guernica is an example of Cubism.
dc:image	The digital image file of the painting.
edm:landingPage	The HTML file with the embedded RDF representation of the painting (URL).
foaf:depiction	Additional to dc:image.

Additionally, *foaf:depiction* and *foaf:page* metadata were used from FOAF vocabulary. The first *(foaf:depiction)* was used as an alternative to *schema:image* and *dc:image,* respectively, and *foaf:page* was used as correspondence to *edm:landingPage*.

6 The dataset

The following information is provided as a general description of the dataset. The list is derived from the dataset description guidelines of SWJ[11] and from the latest published work of data quality vocabulary (DQV) (Albertoni and Isaac, 2020).

- *Name*: Greek Children's Art Museum dataset.
- *URLs*:
  - https://www.childrensartmuseum.gr/ LD-project/(on museum's server)
  - https://zenodo.org/record/4425440 (on Zenodo)
  - https://lod-cloud.net/dataset/ greekchildrensartmuseum (on lod-cloud.net)
  - http://ct-linkdata.aegean.gr/gcam/sparql (SPARQL endpoint)
  - http://ct-linkdata.aegean.gr/query/gcam (query GUI)
- *Version date/number*: 01-01-2021, v.1.
- *Licensing*: CC BY-NC 4.0.
- *Availability*: data is present, obtainable as Turtle, and RDF/XML, and ready for use (SPARQL endpoint available using Fuseki2 Server[12]).
- *Completeness*: 100%, i.e., all artworks (121 of 121) of the GCAM provided collections.
- *OpenData5Star*: 5 stars.
- *Topic coverage*: Child Art, Paintings, Artwork.
- *Source for the data*: Greek Children's Art Museum collections (121 artworks).
- *Metadata description*: DCAT,[13] DCAT-Zenodo[14] and Schema.org (dataset description metadata).

In Table 3, the structure of the GCAM dataset is presented. The dataset is organised in four subsets, based on: a) the two main models (EDM/DC and Schema.org) used to semantically annotate the data and b) the two encodings (RDFa and Microdata) of the structured data used to embedding it in the HTML documents. Each of the subsets is related to one of the three collections (two subsets represent data from the same collection, that is, the collection inspired by cultural heritage). Each subset uses external datasets to link its resources (total of seven datasets). Subset-1 uses all seven datasets for linkage because of the rich information available about those artworks (i.e., artworks inspired by other artworks created by famous artists). Based on this organisation, the number of 121 artworks (objects) is almost evenly distributed per subset (model/encoding).

Table 3 GCAM dataset structure: four subsets, two models (Schema.org, EDM)

	Subset-1	*Subset-2*	*Subset-3*	*Subset-4*	*Data set*
	RDFa / Schema.org	*RDFa / EDM*	*Microdata / Schema.org*	*Microdata / EDM*	
Vocabulary	Schema.org	EDM	Schema.org	EDM	2
Encoding	RDFa	RDFa	Microdata	Microdata	2
Collection	inspired-by-artists	inspired-by-cultural-heritage	inspired-by-the environment	inspired-by-cultural-heritage	3
Linked datasets	DBpedia, Wikidata, Wikiart*, Wikimedia*, Moma*, SearchCulture*, Benaki*	DBpedia, Wikidata.	DBpedia, Wikidata.	DBpedia, Wikidata.	7
Objects Num.	37	25	36	23	121
External Links Num.	259	119	187	109	674
Average num. of external links per object (links / object)	7	4.76	5.19	4.73	5.64
Max. number of external links	9	8	9	7	9
Min. number of ext. links	7	5	5	5	5
D-h-index (in→out)	8	6	7	6	8

Note: * These datasets are only used to link resources as values of unique properties to subset-1 i.e., *isBasedOn* and *citation* properties.

As far as links to external resources are concerned, the total number of 674 links results in an average of 5.64 links per object. Although such an average number is not suitable for evaluating linking coverage (not possible to represent 0.64 as link), it is only used here as an indication of the percentage of linking coverage. However, a more important and interesting observation is to count the minimum and maximum numbers of external links per object, which in this case are 5 and 9, respectively. This number concerns the linking of the resources (objects) to external ones (in → out direction), but not the other way around.

The last row of Table 3 presents a new metric introduced in this paper, namely, the *D-h-index* metric. It is well-known in academia that *h-index* is an author-level metric that measures both the productivity and citation impact of the publications of a scientist or scholar. The *h* number is defined as the largest number *h* such that *h* articles have at least *h* citations, e.g., an h-index of 10 means that 10 articles (of a particular author) have each one earned at least 10 citations (maybe more). Similarly, this paper introduces a new metric for datasets, namely, the *D-h-index* metric, to measure the largest number *h* of objects (resources) such that *h* objects have at least *h* links to other objects (external to the dataset). To distinguish our metric from the metric of the opposite direction (that measures the links of external resources to the resources of a dataset), as defined in the 5-stars rating system by Janowicz et al. (2014), we define two distinct dataset-related *h-index* metrics, that is, *D-h-index(in→out)* and *D-h-index(out→in)*. The *D-h-index(in→out)* of our dataset is 8, where eight objects in our dataset have each at least eight links to external resources. Such a metric goes beyond standard metrics of dataset quality in terms of linkage. Although other research efforts on well-known datasets, such as Wikidata, measure the number of external links to own resources, *D-h-index(out→in)* metric can further reveal the quality of those links as an overall linkage dynamics estimation, that is, the dynamics of the dataset, in respect to other datasets, to have a major impact in the LOD cloud.

6.1 Analysis of vocabularies and dataset usage

In Table 4, the number of links that the resources of GCAM dataset have to external datasets is presented. According to this Table, DBpedia resources have the largest linkage coverage (489 links) to GCAM resources, Wikidata the second largest coverage (151 links). WikiArt and MoMA datasets are linked with 17 and two links, respectively.

There are also several other external datasets that contribute one link to each GCAM dataset (a total of 15 links). These are mainly datasets available from Greek museums (e.g., The Benaki Museum of Greek Culture) and cultural data aggregators (SearchCulture.gr).

Table 4 Number of links to external datasets

Dataset	*Num. of links*
DBpedia	489
Wikidata	151
WikiArt	17
MoMA	2
Others (Wikimedia Commons, The Benaki Museum of Greek Culture, SearchCulture.org, etc.) with 1 link each	15
Total	674

In Table 5, the number of external links per model and the link-facilitator property of models is presented. This analysis is

valuable because it tracks down which metadata contribute more to the overall linkage coverage of the dataset. For instance, *schema:material* property contributes 153 links, and *schema:genre* 146 links, respectively. Overall, Schema.org contributes 446 links, whereas EDM/DC 228, i.e., almost the half of links that Schema.org contributes. Of course, this is also justified both by the nature of Subset-1 and the number of objects that Subset-1 and Subset-3 have (in contrast to the other two EDM/DC-based ones).

Table 5 Number of external links per model and link-facilitation properties

Property	*Schema (Subset-1, Subset-3)*	*EDM/DC (Subset-2, Subset-4)*	*Total*
schema:material e.g., *dbpedia:Colored_pencil*	153	0	153
dc:medium e.g., *dbpedia:Tempera*	0	84	84
schema:genre e.g., *dbpedia:Child_art*	146	0	146
dc:type e.g., *dbpedia:Child_art*	0	96	96
schema:locationCreated e.g., *dbpedia:Athens*	73	0	73
edm:currentLocation e.g., *dbpedia:Athens*	0	48	48
schema:citation e.g., *dbpedia:Henri_Matisse*	37	0	37
schema:isBasedOn e.g., *dbpedia:Guernica_(Picasso)*	37	0	37
Total	446	228	674

Perhaps, the most indicative of the contribution that the selected metadata/properties have in the linkage coverage of the GCAM dataset is the data presented in Table 6. As presented, some properties have more than one value as a resource of an external dataset. For instance, *schema:material* contributes five links to external resources (this is the maximum number), since some artworks have been created with five distinct materials (all available as resources in DBpedia; e.g., *dbpedia:Colored_pencil*, *dbpedia:Tempera*). This case is observed in both models, with their two correspondent properties, that is, a) *schema:material* and *dc:medium* and b) *schema:genre* and *dc:type*.

Table 6 Properties with more than 1 values as external resources

Property	*Max num. of ext. links /object*
schema:material e.g., *dbpedia:Colored_pencil*	5
dc:medium e.g., *dbpedia:Tempera*	4
schema:genre e.g., *dbpedia:Child_art*	2
dc:type e.g., *dbpedia:Child_art*	2

6.2. Data quality (5-stars system)

Apart from presenting GCAM dataset, this paper also introduces a metric for measuring the quality of LD dataset in terms of links to and from other datasets, namely, the *D-h-index* metric. As already stated, to distinguish our index metric from the two directions of links (3-stars and 5-stars of the 5-stars rating system by Janowicz et al. (2014)), we define two distinct dataset-related metrics; that is, *D-h-index(in→out)* and *D-h-index(out→in)*. However, in contrast to this 5-stars system of linked data vocabulary, we measure all kinds of links generated between the resources of the examined dataset (internal URIs) and external resources; that is, vocabulary or data resources (external URIs).

Although the presented and published GCAM dataset is LOD 5-stars compliant, according to Tim Berners-Lee's system (Berners-Lee, 2006), that is, *data on the Web, **machine-readable data, ***non-proprietary format, ****RDF standards, and *****Linked RDF, we have additionally examined the quality of GCAM dataset against the "5-stars of Linked Data vocabulary use" system (Janowicz et al., 2014). The quality of vocabulary (re)use is at least at the 4-stars level, since its metadata schema (defined 100% by reusing well-known vocabularies such as EDM, Schema.org, DC, FOAF) is dereferenceable by humans (1 star) and machines (2 stars), it is linked to other vocabularies (3 stars) and it is annotated using DCAT and Schema.org (dataset description metadata) vocabulary (4 stars). Furthermore, it can be conjectured that the 5th star is present, since the vocabulary-level links between external vocabularies and the vocabularies used in GCAM's metadata schema are examined. For instance, DBpedia ontology[15] (dbo prefix) defines class Painting (*dbo:Painting*) as an equivalent class (*owl:equivalentClass*) of the Schema.org Painting class (*schema:Painting*) reused in GCAM's metadata schema as the type of artwork objects/ resources.

Regarding inferencing, RDFS basic axioms are present; for instance, an object (e.g., painting01) of type *schema:Painting* is also classified (via inferencing) as a *schema:CreativeWork* (subsumption axiom), as nicely presented in OpenLink Data Explorer and explored with URIBurner browser plugin in Chrome (see Figure 3).

7 Example and known shortcomings

7.1 Example

Figures 1, 2, and 3 present an example of a painting artwork created in Athens (a DBpedia resource entry[16]) by a child named "Χαρά Τζόλα" (the child's name in Greek), when she was 10 years old (in 2013). The artwork was based on the work of the famous Greek artist Nikos Hadjikyriakos-Ghikas (a Wikidata resource entry[17]). The painting was created with tempera (a DBpedia resource entry[18]). This painting is based on the original painting "hosted" by the Greek cultural aggregator SearchCulture.gr.[19] The web page of this artwork is accessible at https://www.childrensartmuseum.gr/LD-project/rdfa/

inspired-by-artists/painting01.html. A Turtle representation of the structured data for this example artwork is accessible at https://www.childrensartmuseum.gr/LD-project/example.ttl.

Figure 1 depicts *painting01* object of type *schema:Painting* (subclass of *schema:CreativeWork*) in RDFa Play visualisation environment: a) Encoding: RDFa; b) Model: Schema.org; c) Collection: inspired-by-artists. Figure 2 depicts the same object in Turtle serialisation. Figure 3 depicts the same object in the Linked Data URIBurner browser, using the OpenLink Data Explorer plugin in Chrome browser.

Figure 1 RDFa/Schema.org, inspired-by-artists, painting01 object of type schema:Painting (subclass of schema:CreativeWork) in RDFa Play visualisation app

Figure 2 RDFa/Schema.org, inspired-by-artists, painting01 object of type schema:Painting (subclass of schema:CreativeWork) in Turtle

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix schema: <http://schema.org/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@base <https://www.childrensartmuseum.gr/LD-project/>.

<rdfa/inspired-by-artists/painting01> a schema:Painting;
    schema:about """Πίνακας ζωγραφικής από τη Τζόλα Χαρά, 10 ετών, με αφορμή το έργο του Νίκου Χατζηκυριάκου
    Δημιουργήθηκε στο Μ.Ε.Π.Τ. στο πλαίσιο των 8μηνών εικαστικών εργαστηρίων."""@el;
    schema:citation <https://www.wikidata.org/wiki/Q3341835>;
    schema:copyrightHolder <https://www.childrensartmuseum.gr/>;
    schema:creator "Τζόλα Χαρά"@el;
    schema:dateCreated "2013"^^xsd:gYear;
    schema:educationalUse "8μηνά εικαστικά εργαστήρια Μ.Ε.Π.Τ."@el;
    schema:genre <https://www.wikidata.org/wiki/Q1430886?fbclid=IwAR2WZHzM9t39u3R-mjtHc0mGgQAg7g9lceUgCIm_P7
    <http://dbpedia.org/resource/Child_art>;
    schema:image <https://www.childrensartmuseum.gr/images/photos/545405584525d2effed8328.84669433.jpg>;
    schema:isBasedOn <https://www.searchculture.gr/aggregator/edm/National_Gallery/000047-%3A437663>;
    schema:locationCreated <http://dbpedia.org/resource/Athens>;
    schema:maintainer <https://www.childrensartmuseum.gr/>;
    schema:material <http://dbpedia.org/resource/Tempera>;
    foaf:depiction <https://www.childrensartmuseum.gr/images/photos/545405584525d2effed8328.84669433.jpg>;
    foaf:page <rdfa/inspired-by-artists/painting01.html> .
```

Figure 3 RDFa/Schema.org, inspired-by-artists, painting01 object of type schema:Painting (subclass of schema:CreativeWork) in Linked Data URIBurner browser (OpenLink Data explorer plugin in chrome)

Property	Value
Type	▪ CreativeWork
sameAs	▪ https://www.childrensartmuseum.gr/LD-p...pired-by-artists/painting01.html#this
Title	▪ Έργο τέχνης της Τζόλα Χαράς με αφορμή το έργο του Νίκου Χατζηκυριάκου Γκίκα
described by	▪ https://www.childrensartmuseum.gr/LD-project/rdfa/inspired-by-artists/painting01.html
container of	▪ Tempera ▪ Child art ▪ Athens ▪ https://www.childrensartmuseum.gr/ ▪ https://www.childrensartmuseum.gr/LD-p...a/inspired-by-artists/painting01.html »more»
content	▪ proxy:entity/https/www.childrensartmus...ed-by-artists/painting01.html#content
links to	▪ Tempera ▪ Child art ▪ Athens ▪ https://www.childrensartmuseum.gr/ ▪ https://www.childrensartmuseum.gr/LD-p...a/inspired-by-artists/painting01.html »more»
mentions	▪ [Diffbot] 10 ▪ [Diffbot] URI ▪ [Diffbot] 2013 ▪ [Diffbot] 8
encodingFormat	▪ url
schema:accessmode	▪ visual
stylesheet	▪ https://www.childrensartmuseum.gr/LD-p...fa/inspired-by-artists/pageformat.css
keywords	▪ Τζόλα Χαρά, Πίνακας Ζωγραφικής, Παιδική Τέχνη, Νίκος Χατζηκυριάκος-Γκίκας

Examples of GCAM linked data can be retrieved via a custom query GUI at http://ct-linkdata.aegean.gr/query/gcam using Fuseki2 Server. For instance, an example SPARQL query as the one provided below retrieves children's artwork that was created with *tempera* as material/medium:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX dc: <http://purl.org/dc/terms/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX schema: <http://schema.org/>
PREFIX edm: <http://www.europeana.eu/schemas/edm/>
PREFIX dbr: <http://dbpedia.org/resource/>
SELECT ?object
WHERE {
 {?object dc:medium dbr:Tempera}
 UNION
 {?object schema:material dbr:Tempera}
}
```

The example query retrieves five objects (paintings): four objects from RDFa encoded artworks (of the collections inspired by artists and cultural heritage) and one from the Microdata encoded artworks (of the collection inspired by the environment).

7.2 Known shortcomings

In this work, specific models/vocabularies were selected for the semantic annotation of the available data. Authors acknowledge the existence of other related models/schemas that could be used to model museum LOD, such as Linked Art[20] and LIDO,[21] which rely on CIDOC-CRM. A critical evaluation of all these models to identify the most suitable for GCAM (or other) museum datasets that need to be transformed to LOD is out of the scope of this paper, but remains a future goal of our work.

Although a small set of metadata was selected to structure the data, this selection was driven by the availability of data in already digital form (provided by museum curators). We could further extend the models by using additional metadata, for instance, to represent knowledge related to the dataset and its distinct subsets that it comprises and the different collections that have been provided as input to the LOD processing. Linked Art, LIDO and Schema.org provide such related metadata for collections and sets (e.g., https://linked.art/ns/terms/Set, https://linked.art/model/collection/ or https://schema.org/Collection). As far as our experimental GCAM dataset (three collections, four subsets) is concerned, such design patterns would be highly suitable. This is also left for future work.

Another shortcoming of our dataset is that a data quality vocabulary (DQV) (Albertoni and Isaac, 2020) description is missing. This will not be provided in the future work of this research line. Having said that, in previous sections of this paper we have already provided quality information about the dataset following the semantics of DQV, that is, availability, completeness, and open data 5-stars compliance.

To the best of our knowledge, the information we have about the actual use of the dataset so far is limited to the information we took from Zenodo (135 unique views, seven unique downloads, as recorded at the paper's submission date). However, we strongly support the reuse of our dataset by a permanent SPARQL endpoint we maintain at http://ct-linkdata.aegean.gr/gcam/sparql.

8 Conclusions

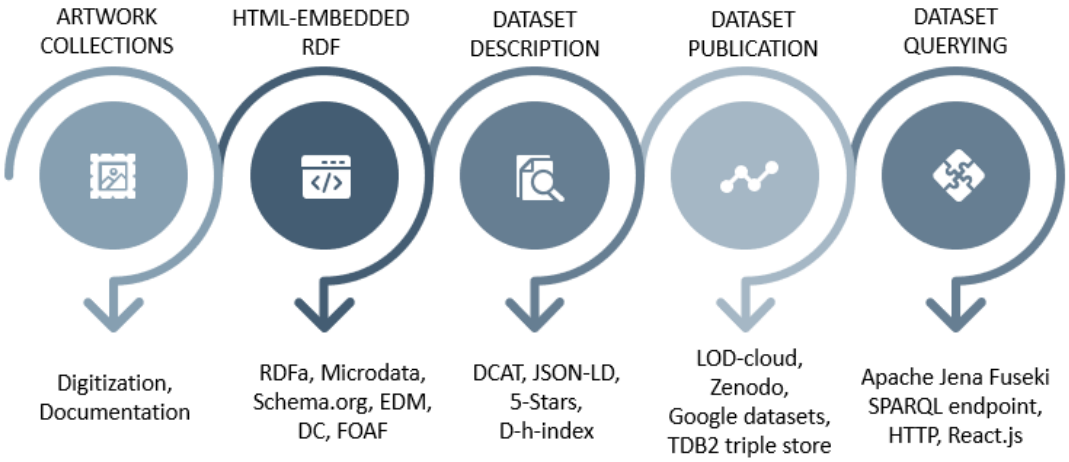
The paper presents the Greek Children's Art Museum collections as linked open data; a dataset created in the context of an academic experiment related to the use of Semantic Web technologies by non-ICT experts. The structured data is represented in the RDF data model and encoded as RDFa and Microdata in HTML5 webpages, using well-known and widely used vocabularies (for their semantic annotation) and external datasets (to link to external resources). A specific processing pipeline of the overall workflow is depicted in Figure 4: from the digitisation and documentation of the art collections to the publication and querying of the created RDF dataset. For each step of the pipeline, a set of specific technologies and actions has been proposed based on the experience gained from GCAM dataset experimentation.

Based on qualitative observations during the experiment, it can be concluded that for non-ICT experts: a) to understand the LOD paradigm and related processing for generating linked open data (expertise dependency/constraint) is not a "mission impossible"; b) to use HTML5, RDFa, and Microdata (technology dependency/constraint) is hard, if applied with text editors such as Notepad++, in contrast to semi-automated semantic annotation tools such as KARMA and c) it makes no difference, in terms of difficulty in use, the specific models/vocabularies provided for the semantic annotation of the provided data (model dependency/constraint).

In addition to contributing to the extension of LOD cloud, this paper has introduced a new data quality metric to measure the linkage dynamics of RDF datasets, namely, the *D-h-index* metric. Although the paper reports on its impact on the generated GCAM dataset, only one direction of linkage (in → out) is possible to be computed at this time. Furthermore, since it is a new metric, it would be rather interesting to further investigate how this metric (in both directions of linking) will influence future efforts in LOD cloud, that is, if future efforts of linked open data creation will adapt this metric in favour of data quality.

Future work on this line of our research will include: a) the critical evaluation of culture-related models/vocabularies so that their strengths and limitations for modelling museum data can be identified; b) the integration of metadata related to the semantic description of collections and sets/subsets of a museum so that their distinct features and purpose of existence can be captured; c) the integration of data quality metadata that are provided by DQV vocabulary and d) the extension of the dataset in two directions, that is, i) adding more collections from GCAM and ii) discovering more links to external resources by utilising the results of other lines of our recent research related to automated links discovery.

Figure 4 Overall process: from digitisation and documentation to data publication and querying



References

Albertoni, R. and Isaac, A. (2020) 'Introducing the data quality vocabulary (DQV)', *Semantic Web*, Vol. 12, No. 1, pp.81–97. Doi: 10.3233/SW-200382.

Alemu, G. and Garoufallou, E. (2020) 'The future of interlinked, interoperable and scalable metadata', *International Journal of Metadata, Semantics and Ontologies*, Vol. 14, No. 2, pp.81–87. Doi: 10.1504/IJMSO.2020.108340.

Angelis, S. and Kotis, K. (2021) Generating and exploiting semantically enriched, integrated, linked and open museum data', in Garoufallou, E. and Ovalle-Perandones, M.A. (Eds): *Metadata and Semantic Research. MTSR 2020. Communications in Computer and Information Science*, Vol. 1355. Springer, Cham. https://doi.org/10.1007/978-3-030-71903-6_34

Berners-Lee, T. (2006) *Linked Data – Design Issues*. Available online at: https://www.w3.org/DesignIssues/LinkedData.html

Chondrogianni, M., Marini, E., Angelis, S. and Kotis, K. (2021a) *DCAT for Greek children's art dataset*. Available online at: https://www.childrensartmuseum.gr/LD-project/ (accessed on 9 January 2021).

Chondrogianni, M., Marini, E., Angelis, S. and Kotis, K. (2021b) *Greek Children Art Museum dataset*, Zenodo. Doi: 10.5281/zenodo.4425440.

Chondrogianni, M., Marini, E., Angelis, S. and Kotis, K. (2021c) *LOD Cloud*. Available online at: https://lod-cloud.net/dataset/greekchildrensartmuseum (accessed on 9 January 2021).

COM-CIDOC LIDO Working Group (2020) *LIDO v1.1 Public Beta-Lightweight Information Describing Objects*. Available online at: https://gitlab.gwdg.de/lido/development.

Community, L.A. (2020) *Linked Art: Objects*. Available online at: https://linked.art/model/ (accessed on 8 February 2021).

Dijkshoorn, C., Jongma, L., Aroyo, L., Van Ossenbruggen, J., Schreiber, G., Ter Weele, W. and Wielemaker, J. (2018) 'The Rijksmuseum collection as linked data', *Semantic Web*, Vol. *9*, No. 2, pp.221–230. Doi: 10.3233/SW-170257.

Dragoni, M., Cabrio, E., Tonelli, S. and Villata, S. (2016) 'Enriching a small artwork collection through semantic linking', in Sack, H., Blomqvist, E., D'Aquin, M., Ghidini, C., Ponzetto, S.P. and Lange, C. (Eds): *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer International Publishing, Cham, pp.724–740. Doi: 10.1007/978-3-319-34129-3_44.

European Union (2014) *Definition of the Europeana Data Model v5.2.5*.

Faraj, G. and Micsik, A. (2019) 'Enriching wikidata with cultural heritage data from the COURAGE project', *Communications in Computer and Information Science*, Springer, pp.407–418. Doi: 10.1007/978-3-030-36599-8_37.

Gareta, C. (2015) 'ALIADA, an open source tool for automatic publication of linked data from libraries and museums', in Garoufallou, E., Hartley, R.J. and Gaitanou, P. (Eds): *Metadata and Semantics Research*, Springer International Publishing, Cham, pp.453–456.

Hermann, I., Adida, B., Sporny, M. and Birbeck, M. (2015) *RDFa 1.1 Primer – Third Edition*. Available online at: https://www.w3.org/TR/rdfa-primer/ (accessed on 5 January 2021).

Hickson, I. (2013) *HTML Microdata*, W3C Working Group Note 29, pp.1–29. Available online at: https://www.w3.org/TR/microdata/

Janowicz, K., Hitzler, P., Adams, B., Kolas, D. and Vardeman, C. (2014) 'Five stars of linked data vocabulary use', *Semantic Web*, *5*(3), 173–176. Doi: 10.3233/SW-140135

Koho, M., Ikkala, E., Leskinen, P., Tamper, M., Tuominen, J. and Hyvönen, E. (2020) 'WarSampo knowledge graph: Finland in the second world war as linked open data', *Semantic Web*, pp.1–14. Doi: 10.3233/sw-200392.

Schema.org (2021) *Schema.org*. Available online at: https://schema.org/ (accessed on 5 January 2021).

Skjæveland, M.G. (2011) *From Spreadsheets to 5-star Linked Data in the Cultural Heritage Domain : A Case Study of the Yellow List-Audun Stolpe*.

Szekely, P., Knoblock, C. A., Yang, F., Zhu, X., Fink, E. E., Allen, R. and Goodlander, G. (2013) 'Connecting the smithsonian american art museum to the linked data cloud', *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, pp.593–607. Doi: 10.1007/978-3-642-38288-8-40.

W3C Recommendation (2017) *HTML 5.2*. Available online at: https://www.w3.org/TR/html52/ (accessed on 5 January 2021).

Websites

1 http://dbpedia.org/page/Child_art

2 https://schema.org/CreativeWork

3 https://schema.org/Painting

4 http://www.europeana.eu/schemas/edm/ProvidedCHO

5 http://www.childrensartmuseum.gr

6 http://rdfa.info/play/

7 https://www.w3.org/RDF/Validator/

8 https://www.easyrdf.org/converter

9 http://rdf.greggkellogg.net/

10 https://search.google.com/structured-data/testing-tool

11 http://www.semantic-web-journal.net/authors#types

12 https://jena.apache.org/documentation/fuseki2/index. html

13 https://www.childrensartmuseum.gr/LD-project/dcat.ttl

14 https://www.childrensartmuseum.gr/LD-project/dcat_zenodo.rdf

15 http://dbpedia.org/ontology/

16 http://dbpedia.org/resource/Athens

17 https://www.wikidata.org/wiki/Q3341835

18 http://dbpedia.org/resource/Tempera

19 https://www.searchculture.gr/aggregator/edm/National_Gallery/000047-%3A437663

20 https://linked.art/

21 http://cidoc.mini.icom.museum/working-groups/lido/lido-technical/specification/