Semantic Web 0 (2014) 1
IOS Press

DM2E: A Linked Data Source of Digitised Manuscripts for the Digital Humanities

Konstantin Baierer [a], Evelyn Dröge [a], Kai Eckert [b], Doron Goldfarb [c], Julia Iwanowa [a], Christian Morbidoni [d], Dominique Ritze [b]

[a] *Berlin School of Library and Information Science, HU Berlin, Unter den Linden 6, 10099 Berlin, Germany*
E-mail: first.last@ibi.hu-berlin.de
[b] *Research Group Data and Web Science, University of Mannheim, B6, 26, 68159 Mannheim, Germany*
E-mail: first@informatik.uni-mannheim.de
[c] *Austrian National Library, Josefsplatz 1, 1015 Vienna, Austria*
E-mail: doron.goldfarb@onb.ac.at
[d] *Semedia, Università Politecnica delle Marche, Department of Information Engineering, Ancona, Italy*
E-mail: christian.morbidoni@gmail.com

Abstract. The DM2E dataset is a five-star dataset providing metadata and links for direct access to digitized content from various cultural heritage institutions across Europe. The data model is a true specialization of the Europeana Data Model and reflects specific requirements from the domain of manuscripts and old prints, as well as from developers who want to create applications on top of the data. One such application is a scholarly research platform for the Digital Humanities that was created as part of the DM2E project and can be seen as a reference implementation. The Linked Data API was developed with versioning and provenance from the beginning, leading to new theoretical and practical insights.

Keywords: Linked Data, Dataset, Cultural Heritage, Digital Humanities, Digital Content, Europeana, EDM

1. Introduction

The project "Digitised Manuscripts to Europeana" (DM2E) is an EU funded project with two primary goals:

1. The transformation of various metadata and content formats describing and representing digital cultural heritage objects (CHOs) in the realm of digitized manuscripts from as many providers (cf. Section 2) as possible into the Europeana Data Model (EDM) to get it into Europeana, the European digital library[1].
2. The stable provision of the data as Linked Data and the creation of tools and services to reuse the data in the Digital Humanities, i.e., to support the so-called "Scholarly Primitives" [13]. The basis is the possibility to annotate the data, to link the data, and to share the results as new data.

The Linked Data representation of the metadata as described in this paper can be accessed online[2] and is as part of the LOD cloud also registered on Datahub.[3] DM2E is a five-star Linked Data source adhering to the Linked Data Principles [3], i.e., it uses dereferenceable URIs, provides all metadata in RDF using proper content negotiation together with links to other Linked Data sources. The vocabulary achieves four stars of the Five Stars of Linked Data Vocabulary Use [12], cf. Section 3. The data is not only provided as an end in itself, but forms the basis for a scholarly research plat-

[1] http://www.europeana.eu

[2] http://data.dm2e.eu
[3] http://datahub.io/dataset/dm2e

form allowing scholars to access the underlying content to annotate it and link it to other sources (Section 4). In order to support the scholars in finding relevant content, the RDF data is enriched – contextualized – as part of the ingestion process (Section 5). A specialty is the provision of full provenance of the data and the support of versioning, as described in Section 6. All RDF data is provided under the CC0 public domain dedication.[4] The dedication does neither extend to the digitized content of the manuscripts nor to the original metadata. The rights statements for the described digitized objects are individually assigned by the content providers who have to choose an appropriate statement from the options[5] offered by Europeana and attach this information to each individual item.

2. Sources

One major aspect of the DM2E project was publishing metadata about a number of international high profile collections both as Linked Data and through Europeana. Despite its name, DM2E is not restricted to manuscripts but contains also other historical resources like letters, books, images, journal articles, or archival items. Table 1 shows an overview on the content available as Linked Data, broken down by provider name, metadata source format, collection name and type of content. As can be seen, this dataset is based on the integration of a variety of source metadata formats, reflecting the heterogeneity of the underlying materials, their international character and the flexibility of the DM2E model to effectively represent such diverse content. Accordingly, there is no common workflow for providers to map their data to the DM2E model. Starting with the tools and the documentation provided by DM2E, providers therefore developed their own metadata transformations mainly based on XSLT, although some chose to directly implement export routines into their collection management systems. Initial consistency checks for mapped data revealed that despite the very detailed specification of the DM2E model some providers showed great creativity in individual interpretations of specific model features, most notably regarding the representation of hierarchies. In order to maintain a homegeneous data representation, specific mapping rules have been established for such cases and distributed amongst the providers in form of a recommendation document. In some cases, transformations created by one provider could be reused or adapted for other providers. This especially proved to be effective for the highly standardized library metadata formats such as MARC, METS/MODS and MAB2. The mapping recommendations and the resulting metadata crosswalks are documented on the DM2E Wiki,[6] a more detailed description of the individual transformation workflows is available as project deliverable [4].

3. Data Model

The DM2E model is an application profile of EDM, i.e., an application-specific specialization for the representation of manuscripts and similar historical content like old prints, posters, books and old journals [6]. EDM itself is very generic to represent various kinds of resources in Europeana provided by museums, libraries, archives and galleries all over Europe. It is again based on top-level ontologies like OAI-ORE, Dublin Core and SKOS. Core classes are *edm:ProvidedCHO* for the described cultural heritage object (CHO), *ore:Aggregation* for the metadata record provided for the described CHO and *edm:WebResource* for views of the described CHO, such as images. CHOs can be further qualified by links to contextual resources being instances of *edm:Agent*, *edm:TimeSpan*, *edm:Place*, or *skos:Concept*. The DM2E model adds mostly subclasses and -properties for the domain of manuscripts. Compared to the EDM data that is available via the Europeana LOD pilot [11], the specialized DM2E data forms a smaller, complementary dataset including RDF statements about more than 2.5 million object pages at a very detailed level by the end of the project in January 2015.

The namespace URI for the DM2E model schema is `http://onto.dm2e.eu/schemas/dm2e/`, the URI for instances is `http://data.dm2e.eu/data/`. The latest model version is 1.2 rc 2, the latest stable release 1.1 [7]. The full documentation of the model as well as mapping recommendations can be accessed via the DM2E Wiki.

The DM2E model was created and refined an an iterative, agile process taking into account several

[4] `http://creativecommons.org/publicdomain/zero/1.0/`

[5] `http://pro.europeana.eu/available-rights-statements`

[6] `http://wiki.dm2e.eu`

Table 1
DM2E Data Sources

Provider	Format	Collection	Type*
Berlin Brandenburg Academy of Sciences	TEI	Deutsches Textarchiv	B
University of Bergen	TEI	Wittgenstein Archive Bergen	M
Bulgarian Academy of Sciences	TEI	Codex Suprasliensis	M
Humboldt University Berlin	TEI	Polytechnisches Journal	J
ERC AdG EUROCORR	TEI	The European Correspondence to Jacob Burckhardt	L
University Library JCS Frankfurt am Main	METS/MODS	Medieval Manuscript Collection	M
		Hebrew Manuscript Collection	M
Georg Eckert Institute for Textbook Research	METS/MODS	GEI-Digital	B
Brandeis University Library via EAJC**	METS/MODS	Spanish Civil War Posters	I
Center for Jewish History via EAJC	MARC	YIVO Institute for Jewish Research Collection	B/A
		Leo Baeck Institute Collection	M/B/J/A
National Library of Israel	MARC	Hebrew & various language Manuscripts	M
		Hebrew, Yiddish & various language Books	B
		Archival Material	A
Berlin State Library	EAD	Personal Papers Collection of Adelbert von Chamisso	M/L
		Personal Papers Collection of Gerhart Hauptmann	M/L
		Publisher Archives of Gebauer & Schwetschke	M/L
		Western Manuscripts	M
American Jewish Joint Distribution Committee (JDC) via EAJC	EAD	Records of the New York Office of the JDC, 1914-18	F
Austrian National Library	MAB2	Austrian Books Online	B/J
		Codices	M
Max Planck Institute for the History of Science	Openmind	Islamic Scientific Manuscripts Initiative (ISMI)	M
	index.meta	MPIWG Digital Rare Book Library	M/B
		The manuscripts of Thomas Harriot	M
Petőfi Literary Museum	DC	A Tett Magazine	J

* M: Manuscripts / L: Letters / B: Books / I: Images / J: Journal Articles / A: Archival Items / F: Archival File
** European Association for Jewish Culture

mapping workshops, constant feedback by the data providers and application developers. Whenever possible, established vocabularies have been reused, precisely: BIBO, CIDOC-CRM, FABIO, PRO, rdaGr2, VIVO and VoID. DM2E-specific usage guidelines for each reused element is provided via *dm2e:scopeNote*. On the five stars scale for LOD vocabulary use proposed by [12], the model gets four stars, as it is dereferencable and machine-readable, linked to other vocabularies, has metadata about it but is not (yet) linked to by other vocabularies.

As of September 30, 2014, the DM2E model contains 71 additional properties and 25 additional classes. The DM2E dataset currently includes descriptions for 2,489,872 cultural heritage objects (*edm:ProvidedCHO* + *ore:Aggregation*), 2,323,774 of which representing single annotatable pages. Regarding contextual resources, 143,395 objects of type *skos:Concept* are available, 30,740 of type *edm:TimeSpan*, 14,999 of type *edm:Agent*, 1,737 of type *foaf:Organization*, 78,265 are typed as *foaf:Person* and 21,775 as *edm:Place*. The class quantity was summed over unique counts per collection.

The example model excerpt shown in Figure 1 illustrates how DM2E data is built-up. Bold resources are added in the DM2E model, others are part of the underlying EDM. In the example, a manuscript by the philosopher Ludwig Wittgenstein is described. The class *ore:Aggregation* is used for information about the metadata record like who has created and mapped the metadata and where is the object shown on the Web whereas *edm:ProvidedCHO* is about the physical object that is described. The DM2E model allows that CHOs have multiple hierarchical layers. This CHO here has two layers: the manuscript as a whole and paragraphs within the manuscript. The type of the CHO is given via *dc:type* in *edm:ProvidedCHO*. The property *dm2e:displayLevel* in *ore:Aggregation* is used to indicate whether the described hierarchical level should be shown in a result list of a search interface or not. If a user searches for example for Wittgenstein, he may not want to have every paragraph of a Wittgenstein manuscript in his search results. Agents are divided into organizations and persons and can be further described. The property *dc:creator*, which is part of EDM, is specialized in DM2E so that we can state that Wittgenstein was the author of the manuscript.

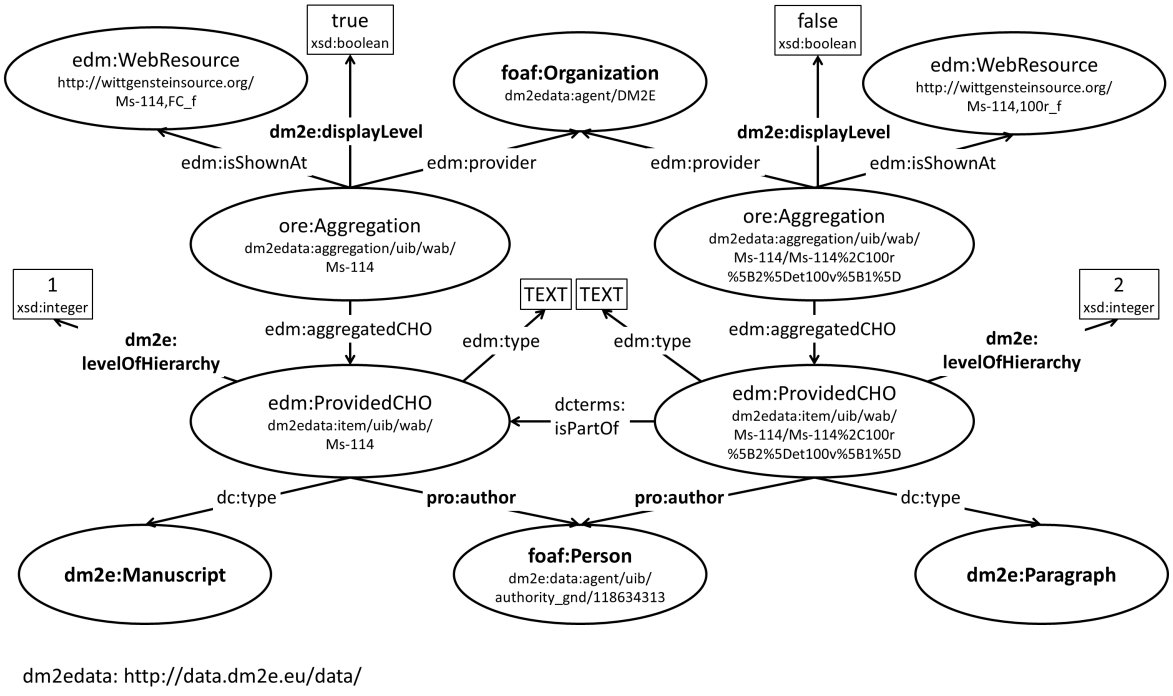

Fig. 1. Example of the DM2E model in use: representation snippet of a Wittgenstein manuscript.

dm2e:levelOfHierarchy "1" says that the manuscript is the highest hierarchical level of this object.

The DM2E model can be used to model manuscripts in more detail but fits exactly the needs of the DM2E providers. Therefore, it may still have to be extended for other institutions that use the model in future. The broader the usage of the domain specific model, the better the chance to achieve cross-provider access to datasets with higher quality. A recent evaluation of the mappings [2] has shown that mapping institutions use the model quite differently and that about a third of all properties and more than a half of all classes of the model are never used. To avoid having an unnecessarily complex model and to get more homogeneous data, the amount of classes and properties in the model will be reduced in its next version.

4. Application

Two applications consuming the DM2E Linked Data have been implemented in the DM2E project to provide the scholarly research platform. The first one is a faceted browser that allows scholars to make sense of the DM2E collections and navigate them along several dimensions, iteratively restricting the search results by language, author, publishing institutions, and other metadata fields. Facets are derived from a SOLR[7] index populated by running queries against the DM2E SPARQL endpoint. In the prototype implementation[8] such queries focus on RDF properties relevant for the end-user and use the *dm2e:displayLevel* property to exclude hierarchical levels that should not show up in search results. As we currently do not provide a public SPARQL endpoint, the RESTful SOLR search API is also an important building block to search the DM2E data programmatically.[9]

Each resource shown in the faceted browser can be opened in its provider's own digital library (following the EDM property *edm:isShownAt*) or in the DM2E semantic annotation environment. The latter constitutes the second web application built on top of the data and is based on Pundit and Feed, two software components developed as part of the DM2E project. Pundit[10] is an annotation tool that allows users to enrich web pages with semantically structured data. Annotations in Pundit encode machine readable seman-

[7] http://lucene.apache.org/solr/

[8] http://purl.org/net/dm2e/search

[9] For example a search for "philosophy": http://purl.org/net/dm2e/search/query?q=philosophy

[10] https://thepund.it/

tic connections among images, texts and LOD entities in form of RDF triples (using the Open Annotation data model[11]), consumable via SPARQL or a dedicated REST API [10]. The Feed REST API provides access to Pundit "as a service" by using the URL of a Web page to be annotated as call parameter. An extension to the Feed API has been developed in DM2E, allowing this call parameter to also be a dereferenceable URL of an RDF description of a digitized object. Feed parses this RDF metadata to create a customized annotation environment based on the object's annotatable Web URL encoded in the DM2E property *dm2e:hasAnnotatableVersionAt*.The *dcterms:isPartOf* and *edm:isNextInSequence* properties are used to provide basic navigation functionalities such as reaching a specific page of a manuscript or going to the next/previous pages. In case of the presence of links to popular LOD datasets such as DBpedia,[12] Feed is able to gather additional metadata (e.g. full names, descriptions, categories) in order to provide additional context to scholars. For this purpose, we established automated contextualization processes, as described in the next section.

As of September 30, 2014, about 6,600 annotations for about 900 digital objects from the DM2E dataset have been created by scholars using our research platform.

5. Contextualization

Linking our datasets to external sources like GND,[13], DBpedia,[14] Geonames,[15] or the Library of Congress Subject Headings[16] enables to easily get information about a resource, either directly by following the link to the external source or by detecting connections between resources based on the same links. While the links to the GND often are already present in the original metadata, links to all other sources are generated automatically. To create the links, we use the link discovery framework Silk.[17] Silk generates links based on a linkage rule that is provided by the user. Such a linkage rule specifies the conditions which have to hold to create a link, e.g. the names of two resources need to have a Jaccard measure value above 0.8. All links that are currently in our system are generated with the same configuration which compares the labels using the Jaro Winkler distance and requires a confidence value of 0.9. With this configuration a high precision is desired while tolerating spelling variations. This might seem like a very simple method, but in most cases we have no further information in the metadata besides a simple string. Our evaluation suggests, however, that even this simple method leads to good results, as many of these strings are not ambiguous.

Table 2
Number of links per external source

dbpedia	freebase	geonames	judaica	lcsh	geodata	nytimes	GND
12287	1868	1571	1474	141	5770	570	22698

Table 2 shows the number of generated links to each external sources. We link agents, places, and subjects.

Altogether, about 24,000 links have been automatically generated. With a manual analysis of 150 random links from agents to DBpedia and 150 random links from places to Linked Geodata, we evaluate the quality. For agents, 125 correct links have been detected which results in a precision of 0.83. Since DBpedia covers several labels, we can for example correctly link "Jakobä" to the DBpedia agent "Jacqueline Countess of Hainaut." The incorrect links result either from ambiguous names, e.g. "Heinrich Fischer" who refers to a Swiss rower in DBpedia and not to an author, or from incomplete information, e.g. if only the first name or surname is given.

For places, 128 correct links can be detected, resulting in a precision of 0.85 which is similar to the precision for agents. Since Linked Geodata includes labels in various languages, even places with a German label such as "München" can be linked to "Munich." The reasons for incorrect links, however, are the same to the ones for agents, e.g. the German city "Heidelberg" is mapped to the city "Heidelberg" located in South Africa due to identical labels.

Across all datasets, about 18% of all agents and 60% of all places are linked on average. With a different linkage rule it is possible to detect more links but with the risk to reduce the precision. Further, the amount of detected links as well as their quality highly depend on the popularity and currency of the resources. Since more than one linkset can be available in our system and the user can track their provenance, more liberal

[11] http://www.openannotation.org/spec/core/
[12] http://dbpedia.org
[13] http://www.dnb.de/DE/Standardisierung/GND/gnd_node.html
[14] http://www.dbpedia.org
[15] http://www.geonames.org
[16] http://id.loc.gov/authorities/subjects.html
[17] http://wifo5-03.informatik.uni-mannheim.de/bizer/silk/

linkage rules can also be applied and the user can be informed about its quality.

6. Implementation

At the core of DM2E's infrastructure is a Jena TDB[18] triplestore, accessible by a Jena Fuseki SPARQL endpoint. After evaluating a few different RDF storage solutions, this combination offered the perfect balance between maintainability, scalability and versatility. In fact, all DM2E's internal applications and the infrastructure are interfacing with the data exclusively through the SPARQL endpoint, making, in theory, the actual triplestore implementation interchangeable. The RDF data is partitioned into Named Graphs that correspond to individual ingestions (see also Section 6.3–Dataset Provenance), hence exporting and importing N-Quad-serialized dumps of the full data store is fairly straightforward.

6.1. Data Ingestion

There are two user interfaces that allow data providers/data mappers to deliver data to DM2E: A Linked Data-based workflow engine with an HTML5 web interface allows casual users to test their transformations and the ingestion process (Omnom)[19] while a set of command line tools is targeted at power users doing large-scale ingestions and conversions (dm2e-data.sh).[20]

Omnom is centered on the idea that RDF's flexible graph-based structure combined with the semantic expressivity of ontologies[21] not only allows the definition and execution of intelligent workflows, automating tedious, long-running and error-prone tasks, but solves the problem of tracking data provenance [9]. Combined with the simple Web User Interface,[22] Omnom can be very helpful for the technically-non-too-savvy to understand the processes of data mapping, data transformation and data ingestion and iteratively improve their own workflow, though Omnom's approach to use and persist RDF for all data does lead to suboptimal performance when doing full-scale transformations/ingestions.

The command line suite of tools is developed with a server environment in mind and consists of a set of Java tools for DM2E validation, provenance-tracking data ingestion, DM2E-EDM-conversion and EDM validation, as well as shell scripts encapsulating XSLT transformers and RDF serializers and for orchestrating the various operations.

The authoritative source of the DM2E model is the textual/tabular "DM2E Data Model Specification" [7], which contains not only the definitions of all properties and classes to be used, but illustrates their usage with examples. The specs are synchronously formalized as an dereferenceable[23] OWL ontology. However, the DM2E model puts restrictions on the usage of properties and classes that cannot be expressed under OWL's Open World Assumption. These restrictions are targeted towards structural validation of subgraphs of DM2E data rather than inference of new facts. While DM2E is involved in the development of community standards for RDF validation [5], we implemented a custom solution using Java, available on GitHub.[24] While the validation tool is "hard-wired" to DM2E's model, it is rather meticulous and has proven useful not only for discovering outright model violations (e.g. wrong cardinality of properties or missing conditional statements) but stylistic problems such as unwise characters in URIs and labels or variations in the UTF-8 normalization.

6.2. Delivery to Europeana

Being a domain aggregator for Europeana, DM2E has a strong focus on interoperability with the EDM, both on the model and data level. To ensure EDM-compliance, conversion of data described with the DM2E data model to EDM for ingestion into Europeana is based on a synthesis of the DM2E OWL and an updated EDM OWL.[25] As the last step before delivery to Europeana, the produced EDM representations are validated using a combination of XML Schema and Schematron.[26]

Due to its ubiquitious deployment in the GLAM sector and its proven track record for scalability,

[18] http://jena.apache.org
[19] http://omnom.dm2e.eu
[20] https://github.com/DM2E/dm2e-ontologies/blob/master/src/main/bash/dm2e-data.sh
[21] http://onto.dm2e.eu/omnom, http://onto.dm2e.eu/omnom-types
[22] https://github.com/DM2E/dm2e-gui
[23] http://onto.dm2e.eu/schemas/dm2e
[24] https://github.com/DM2E/dm2e-ontologies
[25] https://github.com/DM2E/dm2e-ontologies/blob/master/src/main/resources/edm/edm.owl
[26] https://github.com/DM2E/edm-validation

DM2E and Europeana agreed on OAI-PMH as the preferred mode of delivery of data for ingestion into Europeana. Using a multi-step process of extracting per-ore:Aggregation-subgraphs from the triplestore, validation against the DM2E model, conversion to EDM, data massaging and validation against the EDM model,[27] an EDM dump of all data in DM2E is created monthly. With OAI-PMH set names corresponding to datasets, these EDM RDF/XML files are then served using the Repox OAI-PMH repository.

6.3. Linked Data API

The Linked Data API is implemented using a significantly advanced version of Pubby.[28] The source code for this DM2E-specific version is available via GitHub.[29] An integration of the additional features – which are of general interest – into the main branch of Pubby is planned. The basis for all of them is unleashing the power of SPARQL by allowing arbitrary URI patterns to be mapped to customized SPARQL queries. In the following, we describe how the requirements regarding data access have been accomplished for the DM2E data within Pubby.

Multiple resource handling. DM2E implements the OAI-ORE resource map, i.e., whenever the URI of a resource or an aggregation is requested, the client gets redirected to the URI of a resource map. The resource map contains both information about a resource and information about the aggregation – which roughly represents a metadata record in EDM. This implementation also follows practical considerations from the point of view of application developers, as, more often than not, the data about a resource and the data about the aggregation are used together. So this leads to a substantial reduction of necessary requests to the API.

Versioning. All DM2E data is versioned, i.e., the data provided under the URI of a resource map never changes. When updated data is ingested, the API redirects to the new resource map, but the new resource map gets a new URI and contains links to earlier versions of the data. This allows the stable identification of triples within the data, a prerequisite for the data to become a trusted subject of scholarly work.

Dataset provenance. The full provenance of the DM2E data is provided by linking resource maps and other data pages to superordinate datasets using the VoID vocabulary [1]. The datasets are versioned and all data in a dataset shares the same provenance, following the idea of a common provenance context to support provenance-aware Linked Data applications [8]. The version of a resource map and the provided provenance information then simply corresponds to the version and provenance of the dataset.

Statement-level provenance. To support contextualized resources with statements from various enrichment processes, a special approach has been implemented using statement annotations [8]. Subject URIs are created for all statements and these URIs are linked to the datasets the statements originate from. The statement URIs are identified and described as statements using RDF reification. All reification triples are created on the fly, only where necessary, and can safely be ignored by applications not interested in the provenance of the statements. The HTML representation of the contextualized resources makes use of this information and provides an "Oh Yeah?" button for all – possibly wrong – links to external resources, leading to the provenance information of the statement.[30] To the best of our knowledge this is the first implementation of this button as envisioned by Tim Berners-Lee [3].

7. Discussion

Several aspects make the DM2E dataset an interesting and unique source of information. First of all – following the goals of the DM2E project – it contains data from many, carefully selected collections of not only manuscripts, but also old prints, posters, books and old journals with historic value. The data model was developed specifically for this domain where no suitable comprehensive data models existed yet. The DM2E model is also an example of an application profile, an application-specific specialization of the EDM. As such, the data blends well with the huge amount of EDM data available through Europeana. In contrast to many other Linked Datasets, the model and the API have both been tailored to the original data as well as to consuming applications. From a technical point of view, the use of multiple resource representations, ver-

[27] https://github.com/DM2E/dm2e-ontologies/blob/master/src/main/bash/dm2e-data.sh

[28] http://wifo5-03.informatik.uni-mannheim.de/pubby/

[29] https://github.com/dm2e/pubby

[30] For example the city Nancy: http://data.dm2e.eu/data/html/place/onb/abo/Nancy

sioning and the provision of a full provenance chain have to be mentioned, particularly the proper separation of original, curated metadata from data enrichments generated by automated processes with varying quality. Shortcomings of the dataset of course should not be concealed: here, the limited search functionality is the first to be mentioned. A publicly available SPARQL endpoint is currently not provided, mainly out of performance considerations. Fast response times for the scholarly research platform have higher priority. The SOLR-based search and browse interface, however, is provided as convenient entry point to the data and provides a RESTful search API sufficient for most use cases. The data itself also has some shortcomings due to the heterogeneity of the original data. The quality of the metadata ranges from rich descriptions with unambiguous identifiers from authority files for agents, places and subjects to sparse descriptions with few information hidden in free-text fields. It is insofar a dilemma that the contextualization works best for the better data and particularly the poor data is hard to improve. A remedy might be the feedback of data from the annotations provided by the scholars. This will be investigated in the future, when more annotations are hopefully available.

References

[1] Keith Alexander, Richard Cyganiak, Michael Hausenblas, and Jun Zhao. Describing linked datasets - on the design and usage of void, the "vocabulary of interlinked datasets". In *In Linked Data on the Web Workshop (LDOW 09), in conjunction with 18th International World Wide Web Conference (WWW 09*, 2009.

[2] Konstantin Baierer, Evelyn Dröge, Vivien Petras, and Violeta Trkulja. Linked Data Mapping Cultures: An Evaluation of Metadata Usage and Distribution in a Linked Data Environment. In *To appear in: Proceedings of the International Conference on Dublin Core and Metadata Applications, DC-2014*, 2014.

[3] Tim Berners-Lee. Linked data: Design issues. http://www.w3.org/DesignIssues/LinkedData.html, 2006.

[4] Kristin Dill, Evelyn Dröge, Øyvind Liland Gjesdal, Doron Goldfarb, Esther Guggenheim, Julia Iwanowa, Marko Knepper, Gerhard Müller, Alois Pichler, Kilian Schmidtner, Klaus Thoden, and Jorge Urzúa. D1.2 – final integration report. http://dm2e.eu/files/D1.2_2.0_Final_Integration_Report_140214_final.pdf, 2014.

[5] Evelyn Dröge, Thomas Bosch, Valentine Charles, Robina Clayphan, Mark Matienzo, Stefanie Rühle, Adrian Pohl, Mika Alonen, Lars Svensson, and Karen Coyle. Report on the current state: use cases and validation requirements [editor's draft], September 2014. Dublin Core RDF Application Profile Task Group.

[6] Evelyn Dröge, Julia Iwanowa, and Steffen Hennicke. A specialisation of the Europeana Data Model for the representation of manuscripts: The DM2E model. In *Libraries in the Digital Age (LIDA) Proceedings*, 2014.

[7] Evelyn Dröge, Julia Iwanowa, Steffen Hennicke, and Kai Eckert. DM2E Model V 1.1 Specification. http://dm2e.eu/files/DM2E_Model_V1.1_Specification.pdf, 2014.

[8] Kai Eckert. Provenance and Annotations for Linked Data. In Kai Eckert and Muriel Foulonneau, editors, *Proceedings of the DCMI International Conference on Dublin Core and Metadata Applications 2013 (DC-2013), September 2-6, 2013, Lisbon, Portugal*. Dublin, OH : Published by the Dublin Core Metadata Initiative, 2013.

[9] Kai Eckert, Dominique Ritze, Konstantin Baierer, and Christian Bizer. RESTful Open Workflows for Data Provenance and Reuse. In *WWW14 Companion, April 7–11, 2014, Seoul, Korea*, pages 259–260. ACM, New York, NY, 2014.

[10] Marco Grassi, Christian Morbidoni, Michele Nucci, Simone Fonda, and Francesco Piazza. Pundit: augmenting web contents with semantics. *Literary and Linguistic Computing*, 28(4):640–659, 2013.

[11] Antoine Isaac and Bernhard Haslhofer. Europeana Linked Open Data - data.europeana.eu. *Semantic Web*, pages 291–297, 2013.

[12] Krzysztof Janowicz, Pascal Hitzler, Benjamin Adams, Dave Kolas, and Charles Vardeman II. Five Stars of Linked Data Vocabulary Use. *Semantic Web*, 5:173–176, 2014.

[13] J. Unsworth. Scholarly primitives: What methods do humanities researchers have in common, and how might our tools reflect this? Symposium on Humanities Computing: Formal Methods, Experimental Practice, May 2000.