Museum Linked Open Data: Ontologies, Datasets, Projects

Vladimir Alexiev [0000-0001-7508-7428]

Ontotext Corp, Sofia, Bulgaria
vladimir.alexiev@ontotext.com

Abstract. The Galleries, Libraries, Archives and Museums (GLAM) sector deals with complex and varied data. Integrating that data, especially across institutions, has always been a challenge. Semantic data integration is the best approach to deal with such challenges. Linked Open Data (LOD) enable large-scale Digital Humanities (DH) research, collaboration and aggregation, allowing DH researchers to make connections between (and make sense of) the multitude of digitized Cultural Heritage (CH) available on the web. An upsurge of interest in semtech and LOD has swept the CH and DH communities. An active Linked Open Data for Libraries, Archives and Museums (LODLAM) community exists, CH data is published as LOD, and international collaborations have emerged. The value of LOD is especially high in the GLAM sector, since culture by its very nature is cross-border and interlinked. We present interesting LODLAM projects, datasets, and ontologies, as well as Ontotext's experience in this domain.
An extended version of this paper is available. It has 77 pages, 67 figures, detailed info about CH content and XML standards, Wikidata and global authority control.

Keywords: semantic technologies, museum data, LODLAM, CIDOC CRM.

1 Introduction

The Galleries, Libraries, Archives and Museums (GLAM) sector deals with complex and varied data. Integrating that data, especially across institutions, has always been a challenge. There is growing consensus in GLAM that Semantic Data Integration and Linked Open Data (LOD) are the best approach to deal with such challenges. LOD enables large-scale Digital Humanities (DH) research, collaboration and aggregation, allowing DH researchers to make connections between (and make sense of) the multitude of digitized Cultural Heritage (CH) available on the web.

An upsurge of interest in semtech and LOD has swept the CH and DH communities. An active Linked Open Data for Libraries, Archives and Museums (LODLAM) community exists, CH data is published as LOD, and international collaborations have

Digital Presentation and Preservation of Cultural and Scientific Heritage. Conference Proceedings. Vol. 8, Sofia, Bulgaria: Institute of Mathematics and Informatics – BAS, 2018. ISSN: 1314-4006, eISSN: 2535-0366

emerged. Significant investments were made by the EU (Europeana), US (DPLA), various other countries (e.g. Finland's CultureSampo), international foundations (e.g. Mellon) and important CH institutions (e.g. the Getty Trust).

Ontotext is a Bulgarian software company that has worked on semantic technologies since 2000, and on CH LOD since 2010. Ontotext has 65 staff (7 PhD, 30 MS, 20 BS, 6 university lecturers) It is part of Sirma Group Holding, the largest Bulgarian public software group, and is a core part of Sirma Strategy 2022 that focuses on cognitive computing. Ontotext works on semantic modelling, data integration and Knowledge Graph creation, semantic repositories (Ontotext GraphDB), semantic text analysis (entity, concept, relation extraction, document classification), machine learning (entity disambiguation, deep learning in graphs), recommendations, sentiment analysis, etc. In addition to numerous commercial projects, Ontotext is one of the most innovative Bulgarian software companies, with over 40 EU-funded research projects (6 currently active) and various innovation awards.

Ontotext has participated in these CH/DH LOD projects:

- ResearchSpace: British Museum, Yale Center for British Art. Largest museum collection converted to CIDOC CRM, semantic search…
- (with Sirma Enterprise) ConservationSpace, Sirma MuseumSpace
- Medieval Cultures and Technological Resources (VCMS) COST action
- Europeana Creative, Europeana Food and Drink, OAI PMH, SPARQL, Europeana members council, 5 work groups, Data Quality Committee
- Initiator of the Bulgariana national aggregator
- Getty Research Institute: vocabularies LOD
- Carnegie Hall LOD
- American Art Collaborative: consulting 14 US museums integrating data using CIDOC CRM
- European Holocaust Research Infrastructure: semantic archive integration
- Canadian Heritage Information Network: consulting the Canadian national aggregator's transition to LOD
- Wikidata: frequent contributions, mostly to authority control
- DBpedia: contributions, association member, data quality/ontology committee
- CLADA BG: key participant in both CLARIN (NLP) and DARIAH (CH/DH)

We present some interesting museum projects, datasets and ontologies. Other LODLAM domains (archives and libraries) have also made significant progress in LOD adoption, but are out of scope of this paper.

2 Ontologies, Datasets, Semantic Projects

GLAM data is complex and varied: data comes from a variety of systems, it is not regular (exception is the rule, e.g. there may be several Father relations for a person representing different opinions), and many metadata format variations are in use. To enable efficient interoperation, standardization in several areas is needed: content (what to record about objects), interchange (how to transfer data), metadata schemas (how to encode them in technical formats such as XML).

While XML schemas enable the exchange of information, they carry a lot of syntactic baggage (there are many different ways to structure the same information) and do not enable global information sharing (objects are not required to have URLs). RDF and semantic technologies eliminate these shortcomings, enabling the global accumulation and reuse of museum and authority data LOD.

In my opinion, currently there is no dominant and commonly accepted ontology for describing artworks and museum objects. There is tension between several communities. Below are the strongest candidates per my subjective opinion:

- **CIDOC CRM**. Pros: strong foundational ontology, used by numerous projects especially in Europe. Cons: many consider it complicated, some shortcomings for describing relations between people and between objects, not friendly for integrating with other ontologies, the community (SIG) is slow to adopt practically important issues, few application profiles for specific kinds of objects (e.g. coins vs paintings).
- **linked.art**. Pros: a simplified CRM profile created under the moniker "Linked Open **Usable** Data (LOUD)", more developer friendly through an emphasis on JSONLD, used by some projects especially in the US. Cons: various simplifications that are not vetted by the CRM SIG, rift with European CRM developments.
- **Schema.org**. Pros: supported by the major search engines thus ensures semantic SEO and findability, used by the largest amount of LOD (on billions of websites), pragmatic and collaborative process for data modeling with a lot of examples, possible extensions as exemplified by bibliographic (SchemaBibEx) and archival extension. Cons: not yet proven it is sufficient to represent
- **Wikidata**. Pros: universal platform for data integration, richer model than RDF (but also exposed as RDF), pragmatic and versatile collaborative process for data modeling (property creation) with a lot of examples and justifications, used by some GLAMs and crowd-sourced projects (e.g. Authority Contorl,Sum of All Paintings, Wiki Loves Monuments). Cons: institutional endorsement is not yet strong enough, concerns of institutions how they can be masters of "their own" data.

I open this section with two ontologies that are not limited to CH, but are used widely in CH applications.

2.1 Web Annotation

Web Annotation (Open Annotation, OA) is an important W3C standard that covers all kinds of interactions between users and resources: bookmarking, commenting, editing, highlighting, sharing, making relations between resources, etc. Together with ontologies for advanced citation (the SPAR ontologies), it is by now considered crucial for supporting structured scholarly collaboration on the web, and used widely in life sciences, CH and DH. It is also the foundation of advanced IIIF applications such as Shared Canvas, see below.

OA Specifications. The most recent specifications (Feb 2017) include the following.

- Web Annotation Data Model: description of the ontology, different use cases and combinations

- Web Annotation Protocol: defines interactions between annotation servers and clients
- Selectors and States: how to select part of a resource (e.g. section of a HTML document, a particular sentence, rectangle from a PNG image, structural part of an SVG image, page of a PDF) or specify a particular version of a resource as it existed at a certain time. The specification can be done as RDF triples or as URL "fragment selectors" (e.g. "#page=100" for a PDF or "#xywh=100,100,300,300" for an image).
- Embedding Web Annotations in HTML.

OA Resources. specifications are rather dry, but there are excellent illustrations in:

- The Open Annotation Collaboration site
- Slideshares of Rob Sanderson and Paolo Ciccarese
- The Open Annotation Cookbook

Annotation has always been an interesting topic of development, starting with the W3C Annotea project (2001-2003). Within ResearchSpace, Ontotext implemented an old versions of OA for Data and Image Annotation with Deep Zoom (see the next figure for the used OA RDF data model). The availability of an open and stable OA specification has spurned renewed interest, and a large number of implementation efforts. Some interesting examples (mostly from GLAM domain):

- Annotorious image and text annotator by Austrian Institute of Technology, developed as part of the EuropeanaConnect project
- Lorestore server and Annotator OA client by University of Queensland, Australia
- OACVideoAnnotator by UMD MITH and Alexander Street Press
- The LombardPress annotator of ancient manuscripts that works over canonic text representations in the Scholastic Commentaries and Texts Archive
- Annotopia by MIND Informatics group, Massachusetts General Hospital

To reach Recommendation status, every W3C specification requires test suites, and a certain number of independently developed conforming implementations. Compliant implementations listed on the Annotation Model testing report include:

- Reference Implementation of an Annotation protocol server that implements the new Collection and Page portions of the annotation data model.
- Conquering Corsairs (MangoServer) by Rob Sanderson
- Emblematica Online by University of Illinois Library
- Hypothes.is, perhaps the largest OA project and development community. It implements the core AnnotatorJS project. A number of tools, plug-ins and integrations are available, including Drupal, WordPress and Omeka integrations. Omeka is a popular light-weight CMS and virtual exhibition system
- Europeana Annotation Server
- Mirador client (a well-known IIIF viewer, see below) with MangoServer
- Wellcome Quilt, funded by the Wellcome Trust
- Pundit by Net7, developed through several EU projects (e.g. SemLib, DM2E)
- Image Annotator by KANZAKI Masahide
- Page Notes
- Re-narrations and SWeeT Web (source)

We expect the list of implementations to grow quickly, e.g. a new one is:

- Annotation module for Omeka-S, the new generation of Omeka implemented over JSONLD RDF. It allows tag, comment, rate, highlight, draw, etc.

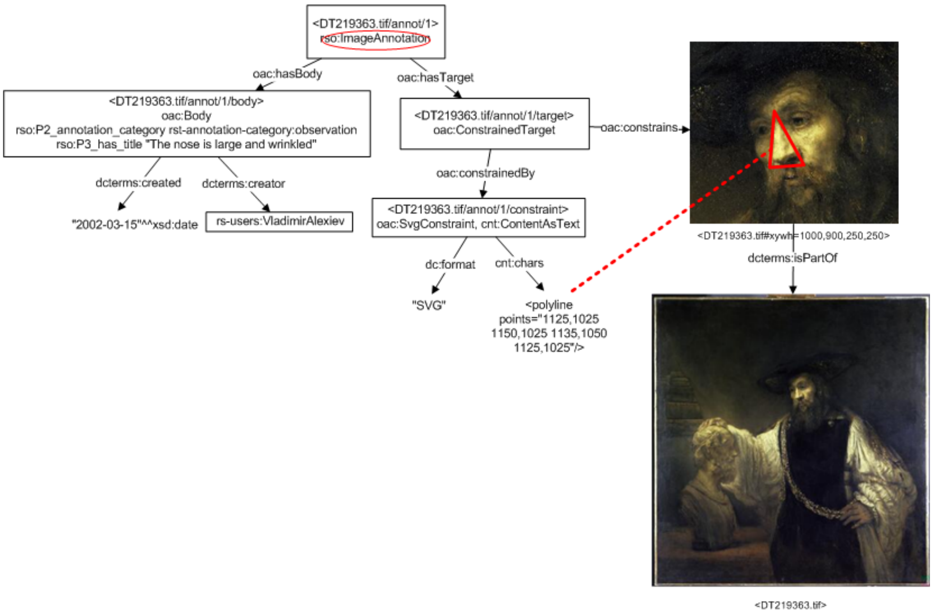


Figure 1. ResearchSpace Image Annotation: Annotating Part of Image with SVG

2.2 IIIF

The International Image Interoperability Framework (IIIF, http://iiif.io/) enables handling of deep zoom (very large resolution) images and applications based on them: book viewers, image composition, image annotation, etc. By defining a client-server protocol, it enables interoperability between image servers (Digital Asset Management) and clients (viewers, annotators). It specifies 4 APIs:

- Image: semantic description of images (available resolutions, features, credit line, conformance level, etc) and serving features such as zooming, gray-scaling, cropping, rotation, etc
- Presentation (Shared Canvas): laying images side by side, assembling folios and books (using so-called IIIF Manifests), image annotation. This has been very popular for virtual reconstruction of manuscripts, book viewers, etc
- Authentication: describes modes or interaction patterns for getting access to protected resources (e.g. Login, Click-through, Kiosk, External authentication)
- Search: search of full-text embedded or related to image resources (e.g. OCRed or manually annotated text of some old book).

Various open source IIIF clients are available, most based on Javascript and HTML5:

- Diva.js, especially suited for use in archival book digitization initiatives
- IIPMooViewer, for image streaming and zooming

- Mirador, implementing a workspace that enables comparison of multiple images from multiple repositories, widely used for manuscripts
- OpenSeadragon, enabling smooth deep zoom and pan
- Leaflet-IIIF, a plugin for the Leaflet framework that also includes display of geographic maps
- Universal Viewer, widely used by CH institutions

Examples of IIIF servers include:

- Cantaloupe, enabling on-demand generation of image derivatives
- IIPImage Server, fast C++ server also used for scientific imagery such as multispectral or hyperspectral images
- Loris, a server written in Python
- ContentDM, a full-featured digital collection management (DAM) system
- Djatoka, a Java-based image server
- Digilib, another Java-based image server

Two examples of IIIF applications:

- Biblissima is the French national manuscript library, based on CIDOC CRM and FRBRoo metadata and IIIF digital asset handling. An IIIF Mirador Viewer is configured to view and compare manuscript images of mermaids from various sources.
- Europeana can search for CHO with IIIF representations by using the search term **sv_dcterms_conformsTo:*iiif***. This returns 2.5M objects.

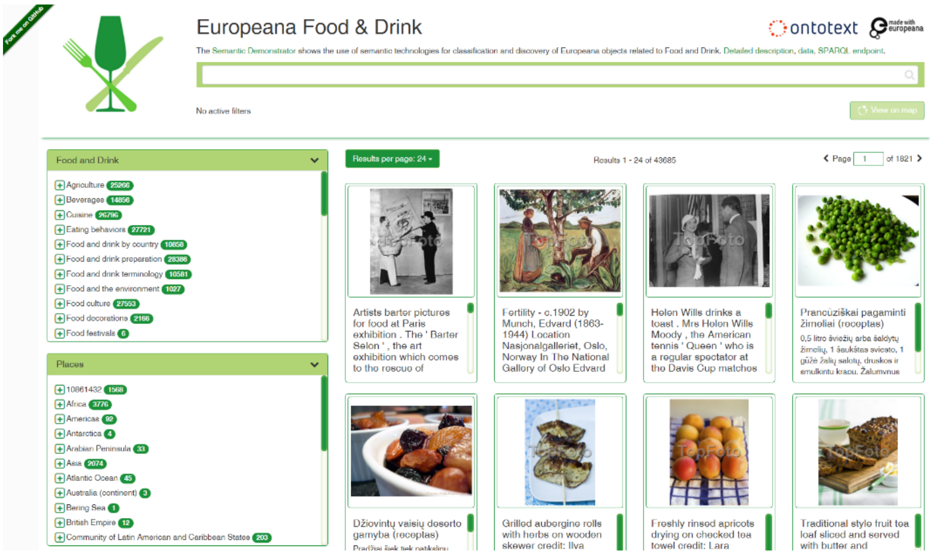
2.3 Europeana and EDM

Europeana is a large-scale CH aggregation that covers CH from institutions in Europe (not only EU member states), Israel and some other countries. It includes artefacts from all over the world (not limited to Europe). It started in 2008 and has aggregated 58M objects at present, described using the Europeana Data Model (EDM), an RDF ontology. Europeana has a general search and display mechanism. The search is not semantic (e.g. won't catch different multilingual names, unless they are included in enriched object data) and includes a set of fixed facets (including image characteristics). Europeana has been criticized for providing a similar look to all kinds of objects, thus not respecting provider wishes and established practices in different domains. So Europeana has created several Thematic Collections (Art, Fashion, Music) that have their own look and features.

Europeana is a long-term program (over 10 years), with perhaps 50 associated projects that aggregated data in particular domains, e.g.:

- APE and APEx aggregated archival information (see below)
- Europeana Regia collected royal illuminated manuscripts
- DM2E (Digital Manuscripts to Europeana) contributed medieval manuscripts, and developed an EDM extension for manuscripts
- PartagePlus collected Art Nouveau
- Europeana Fashion collected artefacts about fashion and garments, which resulted in the establishment of a professional association to continue the project
- Europeana Judaica collected artefacts related to the Judaic tradition
- ECLAP aggregated objects describing performance art

- Europeana Inside developed connectors for several popular Collection Management Systems (CMS) to ease the aggregation of Europeana objects.
- Europeana Creative developed several creative applications, paving the way for reuse of Europeana data by the creative industries.
- Europeana Sounds collects music and other audio and developed an EDM extension for music.
- Europeana Food and Drink collected food and drink related heritage and developed several applications, including a semantic app (Vladimir Alexiev, Andrey Tagarev, & Laura Tolosi, 2016). It includes semantic hierarchical facets for food and drink topics (based on Wikipedia categories) and places (based on Geonames).



Europeana Food & Drink

The Semantic Demonstrator shows the use of semantic technologies for classification and discovery of Europeana objects related to Food and Drink. Detailed description, data, SPARQL endpoint.

No active filters

Food and Drink

- Agriculture 25266
- Beverages 14856
- Cuisine 26796
- Eating behaviors 27721
- Food and drink by country 10858
- Food and drink preparation 28386
- Food and drink terminology 10581
- Food and the environment 1027
- Food culture 27553
- Food decorations 2166
- Food festivals 6

Places

- 10861432 1568
- Africa 3776
- Americas 92
- Antarctica 4
- Arabian Peninsula 33
- Asia 2074
- Atlantic Ocean 45
- Australia (continent) 3
- Bering Sea 1
- British Empire 12
- Community of Latin American and Caribbean States 203

Results per page: 24

Results 1 - 24 of 43685

Page 1 of 1821

Artists barter pictures for food at Paris exhibition . The ' Barter Selon ' , the art exhibition which comes to the roscue of

Fertility - c.1902 by Munch, Edvard (1863-1944) Location Nasjonalgalleriet, Oslo, Norway In The National Gallery of Oslo Edvard

Helen Wills drinks a toast . Mrs Helen Wills Moody , the American tennis ' Queen ' who is a regular spectator at the Davis Cup matches

Prancūziškai pagaminti žirneliai (receptas)
0,5 litro šviežių arba šaldytų žirnelių, 1 šaukštas sviesto, 1 gūžė žalių salotų, druskos ir smulkintų krapų. Žalumynus

Džiovintų vaisių deserto gamyba (receptas)
Pradžiai šiek tiek patikslinu

Grilled aubergine rolls with herbs on wooden skewer credit: Ilva

Freshly rinsed apricots drying on checked tea towel credit: Lara

Traditional style fruit tea loaf sliced and served with butter and

Figure 2. Europeana Food and Drink (EFD) Semantic Application

In addition, various national aggregators have emerged, e.g.:

- Collection Trust established CultureGrid in the UK
- The German Digital Library (DDB) is the aggregator for DE
- DigitaleCollectie is the aggregator for NL
- The Varna library established the first BG aggregator, and Ontotext established Bulgariana, an aggregator with a more technological orientation. E.g. Bulgariana submitted a BG traditional recipes collection to EFD, including semantic enrichment.

Many Europeana satellite projects have faced sustainability problems, i.e. inability to continue collecting and updating objects after the project finishes. Good exceptions are Apex and Europeana Fashion that have established respective associations to continue the work. Not coincidentally, these projects (especially Apex) often collect richer metadata and submit a subset of it as EDM to Europeana. Even some national aggregators faced sustainability problems.

Aggregating a collection often takes a long time by Europeana (several weeks) because of slow iteration cycles of test ingestion, previews, checking object quality. Europeana has changed several aggregation approaches and software: SIP ingest, Unified Ingest Manager (UIM); Europeana Inside (connectors to various Collection Management Systems, CMS); Operation Direct (announced at Europeana's 2016 AGM): an API-based ingestion approach, where a CMS can submit and update individual objects directly to Europeana, and it adds them to the search index incrementally; and is currently working on Metis.

The Europeana API allows applications to search for objects, using a large selection of search fields. However, it does not allow complex queries (e.g. across objects, result aggregation such as count or sum and group by, searching by author characteristics such as nationality, by concept or place hierarchy, etc). Although EDM is an RDF ontology, semantic technologies are not used in the core of Europeana. Instead, it uses SOLR to index all search fields.

Europeana Labs provides a gallery of datasets and apps. Several Europeana projects (starting with Europeana Creative and Food and Drink) have organized competitions, provided prizes and start-up support, in an effort to increase creative reuse of CH materials.

Europeana uses the OAI PMH protocol to aggregate content from aggregators. In 2015 it also established an OAI PMH server developed by Ontotext (Vladimir Alexiev & Dilyana Angelova, 2015) to allow mass-downloading of metadata. Ontotext also created the Europeana SPARQL endpoint allowing complex queries, which was later replaced by an open source RDF repository. However, the SPARQL endpoint is not supported well (there is a google group with little traffic) and is not widely used.

Europeana is currently funded by the EC as a Digital Service Infrastructure (DSI) under the Connecting Europe Facility (CEF). Although the funding is smaller than in previous years. This ensures Europeana's longevity. Recent targeted funding includes projects for creating more collections on Migration, Rise of Literacy, Byzantine Art.

Critiques. For long Europeana focused on quantity rather than quality, leading to:

- Low metadata quality of some of the collected objects: poor or incomplete metadata, mistakes in metadata structure, broken links, etc.
- Uneven content selection criteria. For example, AskAboutIreland contributed yellow pages (phone books) from 1975, every page as a separate object; LGMA contributed photos of common foods like carrots and jelly, etc
- Aggregation through one-off projects, leading to inability to update the aggregated collections (provide new content) and low availability of images and institutional websites

EDM has been criticized by some in the CIDOC CRM community (Dominic Oldman, Martin Doerr, Gerald de Jong, Barry Norton, & Thomas Wikman, 2014) for being a least-common-denominator model that shoe-horns CH institutions into providing a poorer version of their metadata. Since aggregation initiatives are expensive, data should be aggregated in a rich format to begin with, and the Synergy Reference Model is proposed to that end. While EDM allows richer modelling such as events, this is not supported by Europeana and many existing metadata collections have little more than Dublin Core.

In the last two years Europeana has put Data Quality in the middle of its Strategic agenda. In particular:

- Two task forces have focused on Enrichment, since semantic enrichment of metadata is one of the ways to increase the value of metadata.
- A Data Quality task force (May 2015) analysed and outlined problems.
- A permanent Data Quality Committee was formed to define and validate quality rules (using mechanisms such as RDF Shapes) and measure metadata coverage.
- The Europeana Publishing Framework established tiers of participation, where some institutions can benefit more by providing higher-quality collections, better-resolution images, and richer metadata.

Despite the progress, a lot of work remains to make Europeana objects most useful for consumers and researchers.

Pros. One of the most important achievements of Europeana is increasing the level of networking of CH institutions in Europe. Europeana has also been very strong in user engagement, developer engagement (hackathons and Europeana Lab), lobbying for digitization and CH in Europe.

Europeana has a strong distributed organization. It operates through several interconnected groups:

- About 3500 CH institutions contribute content through a network of Aggregators, reducing the load on the Europeana office.
- Funding is sought by and provided through the Europeana Foundation
- The Europeana Association is a voluntary organization with about 3000 individual members. It meets yearly at the Europeana AGM and elects a Members Council that participates in setting Europeana strategy, selecting task forces, etc
- Task Forces are temporary groups assembled to elaborate and make recommendations on issues of importance. Working Groups are more permanent.

Europeana has set some technological examples (e.g. the EDM) that have been followed by DPLA. Also, Europeana is cooperating with DPLA and other organizations on license standardization (RightsStatements.org), IIIF images, schema.org representation for better findability by search engines, etc.

The Europeana Data Model (EDM) (Europeana, 2017) is an RDF ontology used by Europeana for harvesting and managing CH objects (CHO). EDM builds upon:

- Dublin Core (DC): descriptive metadata
- OAI ORE (Open Archives Initiative Object Reuse & Exchange): organizing object metadata and digital representations (WebResources)
- SKOS (Simple Knowledge Organization System): contextual objects (concepts, agents, etc)

EDM is inspired by CIDOC CRM (see below): events, some relations between objects. EDM describes:

- CHOs (ProvidedCHO)
- The real-world things related to them (Non-Information Resources, also called contextual entities or contextual objects).
- Metadata records (aggregations)
- Associated media (WebResources)

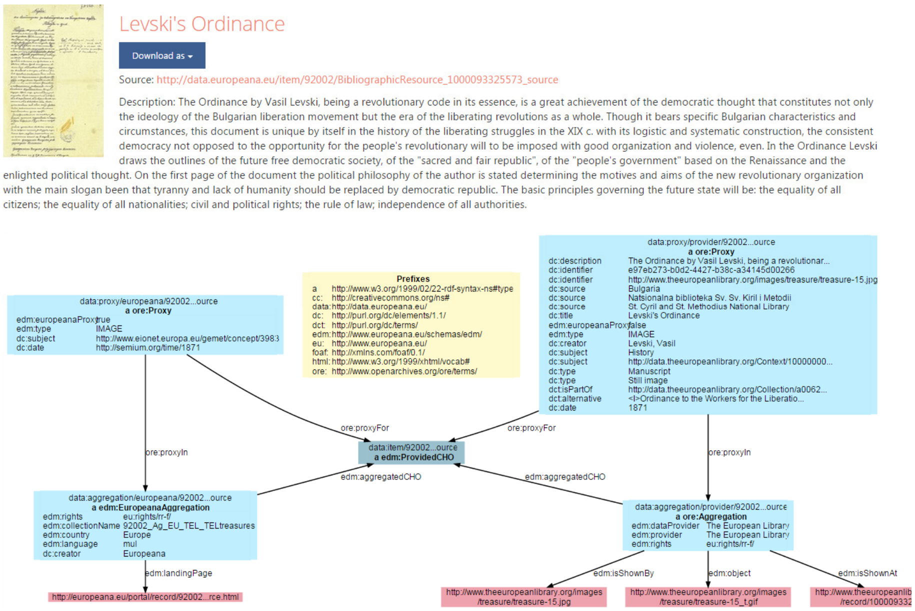
EDM includes two auxiliary classes from ORE, which are used to split the information into clearly delineated nodes:

- Proxy carries object information, as provided by a certain agent (the data Provider or Europeana)
- Aggregation carries information about the provider, collection, metadata rights, etc

EDM has two flavors:

- **External** as served by the Provider (aggregator). It has only 2 nodes, ProvidedCHO and Aggregation.
- **Internal**: after Europeana ingests the object, it splits the object info to the Provider Proxy and adds extra info in the Europeana Aggregation and Proxy. The ProvidedCHO node itself does not carry information.

A typical EDM graph is shown below, highlighting the nodes centered around CHO.



Levski's Ordinance

Download as ▾

Source: http://data.europeana.eu/item/92002/BibliographicResource_1000093325573_source

Description: The Ordinance by Vasil Levski, being a revolutionary code in its essence, is a great achievement of the democratic thought that constitutes not only the ideology of the Bulgarian liberation movement but the era of the liberating revolutions as a whole. Though it bears specific Bulgarian characteristics and circumstances, this document is unique by itself in the history of the liberating struggles in the XIX c. with its logistic and systematic construction, the consistent democracy not opposed to the opportunity for the people's revolutionary will to be imposed with good organization and violence, even. In the Ordinance Levski draws the outlines of the future free democratic society, of the "sacred and fair republic", of the "people's government" based on the Renaissance and the enlighted political thought. On the first page of the document the political philosophy of the author is stated determining the motives and aims of the new revolutionary organization with the main slogan been that tyranny and lack of humanity should be replaced by democratic republic. The basic principles governing the future state will be: the equality of all citizens; the equality of all nationalities; civil and political rights; the rule of law; independence of all authorities.

data:proxy/europeana/92002...ource
a ore:Proxy

edm:europeanaProxy	true
edm:type	IMAGE
dc:subject	http://www.eionet.europa.eu/gemet/concept/3983
dc:date	http://semium.org/time/1871

Prefixes

a	http://www.w3.org/1999/02/22-rdf-syntax-ns#type
cc:	http://creativecommons.org/ns#
data:	http://data.europeana.eu/
dc:	http://purl.org/dc/elements/1.1/
dct:	http://purl.org/dc/terms/
edm:	http://www.europeana.eu/schemas/edm/
eu:	http://www.europeana.eu/
foaf:	http://xmlns.com/foaf/0.1/
html:	http://www.w3.org/1999/xhtml/vocab#
ore:	http://www.openarchives.org/ore/terms/

data:proxy/provider/92002...ource
a ore:Proxy

dc:description	The Ordinance by Vasil Levski, being a revolutionar...
dc:identifier	e97eb273-b0d2-4427-b38c-a34145d00266
dc:identifier	http://www.theeuropeanlibrary.org/images/treasure/treasure-15.jpg
dc:source	Bulgaria
dc:source	Natsionalna biblioteka Sv. Sv. Kiril i Metodii
dc:source	St. Cyril and St. Methodius National Library
dc:title	Levski's Ordinance
edm:europeanaProxy	false
edm:type	IMAGE
dc:creator	Levski, Vasil
dc:subject	History
dc:subject	http://data.theeuropeanlibrary.org/Context/10000000...
dc:type	Manuscript
dc:type	Still image
dct:isPartOf	http://data.theeuropeanlibrary.org/Collection/a0062...
dct:alternative	<I>Ordinance to the Workers for the Liberatio...
dc:date	1871

data:item/92002...ource
a edm:ProvidedCHO

data:aggregation/europeana/92002...ource
a edm:EuropeanaAggregation

edm:rights	eu:rights/rr-f/
edm:collectionName	92002_Ag_EU_TEL_TELtreasures
edm:country	Europe
edm:language	mul
dc:creator	Europeana

data:aggregation/provider/92002...ource
a ore:Aggregation

edm:dataProvider	The European Library
edm:provider	The European Library
edm:rights	eu:rights/rr-f/

http://europeana.eu/portal/record/92002...rce.html

http://www.theeuropeanlibrary.org/images/treasure/treasure-15.jpg

http://www.theeuropeanlibrary.org/images/treasure/treasure-15_t.gif

http://www.theeuropeanlib /record/100009332

Figure 3. Typical EDM Graph (from Ontotext's Europeana endpoint)

Despite the complicated graph structure, typical EDM objects used to have little more than DC information: in particular, few if any references to global authorities, rather providing mere strings. However, Europeana has started providing more enrichments against authorities such as Geonames, DBpedia, Getty AAT. The Europeana Entity Base copies relevant authorities from LOD sources (only resources that appear in CHOs or are widely used) and equivalences to the original URLs in those datasets. Also, Europeana implements and promotes the use of IIIF for deep zoom images.

Several EDM extensions and profiles have been proposed:

- Describing Hierarchical Objects, such as books
- Extending EDM with properties from FRBRoo
- EDM Profile for Sound

The EDM mappings, refinements and extensions task force published a report (2014) on various extensions and extension approaches. The EDM Extensions workshop (2015) developed directions for future extensions.

An important EDM feature was identified by the Europeana Data Quality Committee as required to improve the precision of describing author contributions to artworks: edm:Event with dc:type being Production or a specific "business sub-type" such as design, gilding, decoration, translation, etc. Although EDM includes such class, it is not implemented in the Europeana portal and consequently is not used by data providers.

2.4 CIDOC CRM

The CIDOC Conceptual Reference Model (CRM) (Patrick Le Boeuf, Martin Doerr, Christian Emil Ore, & Stephen Stead, 2018) is a foundational ontology for history, archeology and art. It is developed by ICOM, CIDOC (International Committee for Documentation), CRM Special Interest Group (http://www.cidoc-crm.org). It has been in development for 17 years (since 1999) and standardized as ISO 21127:2006 in 2006. The ontology continues to evolve: the current version with RDF representation is CRM 6.2.1 (Oct 2015), the version in progress is CRM 6.2.3 (May 2018). It has about 85 classes and 285 properties (about 140 object properties and their inverses, and a few that don't have inverses).

Many resources are available to learn CIDOC CRM, e.g.:

- Video Tutorial (2008) that explains the logic of CIDOC CRM, especially the event orientation.
- Graphical Representation: presents "typical situations" or CRM constructs. Includes a comprehensive property and class index that allows you to lookup a certain ontology element in all typical situations.
- The CRM Primer presents CRM in brief by presenting the representation of typical museum information.

Most CRM classes fall in the following fundamental divisions (see red lines in the following figure:

E77 Persistent (endurant): whenever it exists, it exists with all its parts simultaneously. This does not preclude changes in time (e.g. part additions/removals)

- **E18 Physical**. Includes physical things such as objects, features (e.g. scratches, marks, inscriptions), collections, and even persons.
- **E28 Conceptual**. Includes ideas, text, images, formulas and other "information" entities that can be easily copied communicated in many different formats, with some variations of physical rendition that still keep them recognizable. Includes information found on museum objects (e.g. inscriptions, text, images) but also museum documentation info such as titles, identifiers, types, languages, etc.
- **E39 Actor**. Please note that E21 Person has two super-classes. If you study the actions of a person, that corresponds to his role as Actor. But if you study his remains, that would be under his role as E19 Physical Thing.

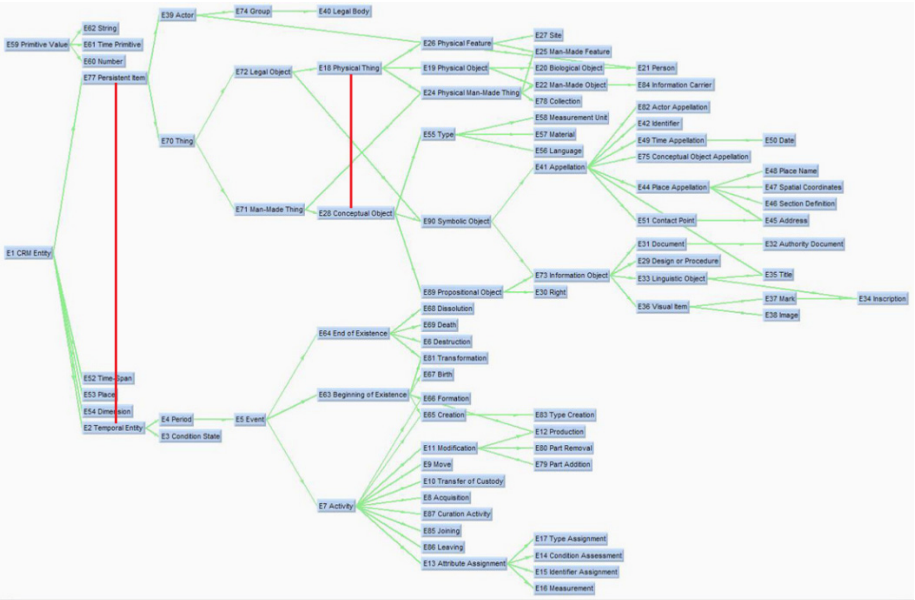


Figure 4. CRM Class Hierarchy

E2 Temporal (perdurant): progresses through time. This includes such large temporal entities like **E4 Period** (a whole cultural period), shorter specific **E5 Events**, and **E7 Activity** (which is caused by an actor). Specific events/activities include:

- **Beginning of Existence**: **Birth** of a person, **Formation** of a group, **Production** of a physical object, **Creation** of a conceptual object.
- **End of Existence**: **Death** of a person, **Dissolution** of a group, **Destruction** of a physical object. Conceptual objects cannot be destroyed, since they exist separately from any and all physical carriers.
- **Transformation**, which is both the End of an old object, and the Beginning of a new one
- **Move, Acquisition** (Transfer of Ownership), **Transfer of Custody**. CRM distinguishes between owner and custodian (keeper/curator).
- **Modification** (of an object), **Part Addition, Part Removal** (of an object or collection); **Joining/Leaving** (of a group)
- Activities related to museum documentation: **Attribute Assignment** and its subclasses. E.g. **Measurement** records the details of how a Dimension was obtained, **Identifier Assignment** records when an identifier or title started to be used (assignment) and stopped to be used (deassignment).

A few classes outside these branches:

- **Primitive Value** and its subclasses are not used in RDF. Instead, appropriate RDF literals are used (e.g. xsd:string, rdf:langString, xsd:decimal, xsd:date, xsd:gYearMonth, xsd:gYear)
- **Place**: can be a place on Earth or on an object, identified through a "Section Definition"

- **Dimension**: some dimension of an object, comprising type, unit and value
- **Time-Span**: temporal info (see below)

CRM Time. Historic/archeological time intervals are expressed in CRM using E52 Time-Span, which allows fuzzy intervals and comprises:

- A label (e.g. "started circa 1520, finished no later than 1610")
- Duration (minimum, maximum): **P83 had at least duration, P84 had at most duration**
- Up to 4 dates (see below) that are refinements of **P82 at some time within, P81 ongoing throughout**. These define the outer and inner bounds of the interval.

Table 1. CRM Time-Span Bounds

CRM property	**Meaning**	**Latin phrase**	**Meaning**
P82a_begin_of_the_begin	started after this moment	terminus post quem	limit after which
P81a_end_of_the_begin	started before this moment	terminus a quo	limit from which
P81b_begin_of_the_end	finished after this moment	terminus ad quem	limit to which
P82b_end_of_the_end	finished before this moment	terminus ante quem	limit before which

Representing Objects and Features Museum objects are mapped to **E22 Man-Made Object** (or **E19 Physical Object** if they are natural such as a rock). Further distinctions are introduced with **P2_has_type** which points to a thesaurus (**E55 Type** or **skos:Concept**); this is a universal property that applies to any **E1 Entity**. This underlies the universality of CIDOC CRM.

It may be tempting to define more specific classes like Painting or Sculpture. But museums hold all kinds of weird and wonderful things; e.g. the Getty AAT Object hierarchy has 20k concepts.

CRM has sufficient universal constructs to model more specialized domains. E.g. consider Numismatics. Coins use specific dimension types (e.g. die-axis, o'clock) that can be modeled with P2_has_type, referring to a specialized thesaurus (e.g. AAT or BM thesaurus). We need to describe separately the images and inscriptions on the Obverse and Reverse sides of the coin. To model this, consider the CRM Graphical diagram below (double arrows show sub-class and sub-property relations, single arrows are properties). We model Coins as follows (see CRM Graphical: Mark and Inscription Information, parts 1 and 2):

- E22_Man-Made_Object (with standardized P2_has_type Coin) P56_bears_feature E25_Man-Made_Feature (with standardized P2_has_type Obverse or Reverse). These classes can be related by P56 because they are sub-classes of E19 respectively E26, which are the defined domain & range of P56
- E25_Man-Made_Feature (obverse/reverse) P65_carries_visual_item E38_Image (e.g. of a ruler) or E34_Inscription (some text). These classes can be related by P65

because they are subclasses of E24 respectively E36, which are the defined domain & range of P65.

- E38_Image P138_represents (some ruler, e.g. from ULAN). You can find this relation on graphical diagram Image Information Objects and Carriers
- E34_Inscription P3_has_note "the text" and P72_has_language (some language from a thesaurus, e.g. Latin from AAT). We could also record P73_has_translation to another node (Linguistic Object), e.g. a translation to English

Since Features are considered Things, one can represent these situations:

- Represent a wax seal on a parchment, or an ink stamp or signature on a paper document, and use P45_consists_of to designate the material
- Record the specific technique (e.g. incised) or creator of a mark or inscription by using E12 Production or E11 Modification, recording P32_used_general_technique and P14_carried_out_by

CRM has "part of" relations for various entities (physical object, conceptual object, place, temporal object including event, actor). It has title/ identifier/ image (representation) for objects; who (actor)/ when (time span)/ where (place) for events/activities. CRM includes limited object relations (shows features of, motivation/influence), and it has been criticized for that. CRM is strongly event-oriented. One cannot attach person, place and date information to an object directly: there are no simple properties like "creator", "created on", "created at": one must create Events, e.g. Production. But this allows richer representation of more complex cases, e.g. different kinds of contribution as production sub-events, Attribution Qualifiers (workshop of, circle of, attributed to), etc.

CRM Short Cuts and Long Paths are an important CRM notion that allows recording of information with different levels of detail. E.g. CRM Graphical "Measurement Information" shows that Dimension records the direct info about an object, while the Measurement node allows to record extra info about it: who did the measurement, when, what tools were used, what was the precision, etc. E13_Attribute_Assignment is the prototypical class that participates in long-paths, and activities such as Measurement, Type Assignment are sub-classes thereof.

There are several CRM RDF definitions, the two most important being:

- CRM SIG: RDFS. It defines classes, properties (with multiple language translations), and sub-class and sub-property relations.
- Erlangen CRM: OWL-DL. It tracks the official definition and adds inverse and transitive property declarations and class restrictions (owl:Restriction). It is developed on github and full version history is available. Since some of these additions (especially the restrictions) are controversial, I provided a script ecrm-simplify.xq that can generate CRM "application profiles", e.g. leave only the inverse declarations, which are an innate feature of CRM.

CRM Extensions. The CIDOC CRM specification, section "Modelling principles: Extensions", defines how to extend CRM in a compatible way, so that an application that understands only the core ontology, still can consume data conforming to the extension. The guidance is to create extension properties and classes as sub-properties and sub-classes of the core. The following CRM extension ontologies have been developed. See (Martin Dörr, 2018) for an overview:

- FRBRoo: bibliographic information following the FRBR principles (Work-Expression-Manifestation-Item), artistic performances and their recordings
- PRESoo: periodic publications
- DoReMus: music and performances
- CRMdig: digitization processes and provenance metadata
- CRMinf: statements, argumentation, beliefs
- CRMsci: scientific observations
- CRMgeo: spatiotemporal modeling by integrating CRM to GeoSPARQL
- Parthenos Entities: research objects, software, datasets
- CRMeh (English Heritage): archeology
- CRMarchaeo: archeology, excavation, stratigraphy
- CRMba: buildings
- CRMx: proposed extension for museum objects, including simple properties such as main depiction of an object, preferred title, extent, etc

Benefits

- Provides a strong ontological foundation
- Being event-based, it is well suited for representing deeper details, such as separate contributions to an artwork, object parts, etc.
- Used by a large number of (especially European) projects, e.g. UK Claros, UK ResearchSpace, H2020 Gravitate, H2020 Parthenos, etc
- Has extensions in various domains, most importantly archeology and bibliography

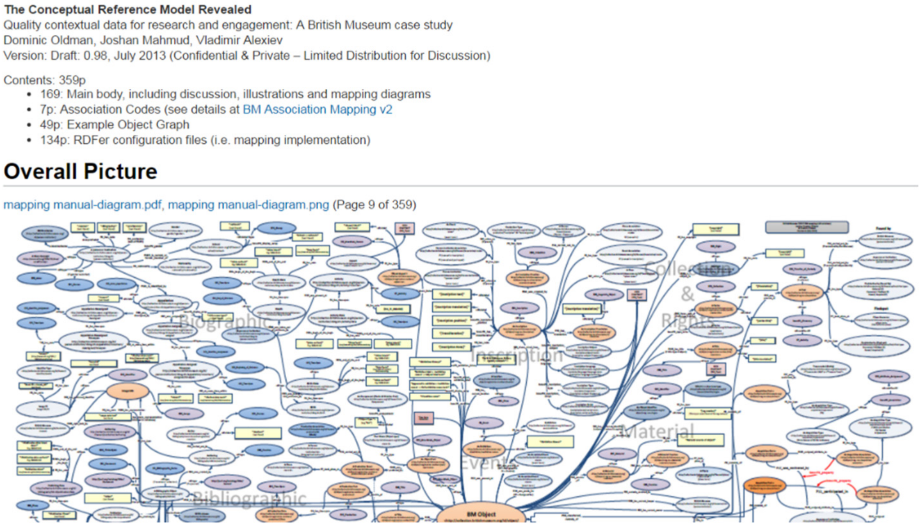
Cons

- Somewhat complicated
- Some CRM SIG members are somewhat theoretical, with little regard for practical implementation
- Most collaboration happens in face to face meetings (not so strong electronic collaboration)
- Overly deep class hierarchy with a lot of abstract and not so useful classes
- Strict (monomorphic) domains and ranges, which leads to modeling complications

2.5 UK ResearchSpace (British Museum)

CRM was used in projects since about 2000 (e.g. CLAROS-Net at Oxford started in 2009). But the first large-scale CRM-based effort was ResearchSpace (RS). It is a Mellon-funded project that started in 2010 and is ongoing. The purpose of the project is to develop a web-based Virtual Research Environment (VRE) where art researchers can collaborate on different projects, import and interlink semantic data, coreference thesauri, use semantic search, annotate data and images, etc. The project is strongly based on CIDOC CRM and has provided CRM consulting and mapping advice in various summer schools and other fora.

British Museum Data as CIDOC CRM. As part of the RS project, the British Museum data was mapped to CRM and published semantically. In Oct 2015 the Open Data Institute and NESTA organized the Heritage+Culture Open Data Challenge, and as part of that initiative released a Data Guide and a comparison of CH open datasets. In that comparison, the BM SPARQL Endpoint received a perfect score, for depth of data representation and other indicators. The mapping documentation (Oldman, Mahmud, & Alexiev, 2013) is very comprehensive but is monolithic and has imprecisions. There is a lot more technical information at the Ontotext RS confluence. This model of mapping museum data to CIDOC CRM has been followed by some US museums: Yale Center for British Art (YCBA) and Smithsonian American Art Museum (SAAM).



The Conceptual Reference Model Revealed
Quality contextual data for research and engagement: A British Museum case study
Dominic Oldman, Joshan Mahmud, Vladimir Alexiev
Version: Draft: 0.98, July 2013 (Confidential & Private – Limited Distribution for Discussion)

Contents: 359p

- 169: Main body, including discussion, illustrations and mapping diagrams
- 7p: Association Codes (see details at BM Association Mapping v2
- 49p: Example Object Graph
- 134p: RDFer configuration files (i.e. mapping implementation)

Overall Picture

mapping manual-diagram.pdf, mapping manual-diagram.png (Page 9 of 359)

Figure 5. ResearchSpace British Museum Mapping to CIDOC CRM

CIDOC CRM Semantic Search. RS implemented semantic search based on CRM Fundamental Relations (FR). It was based on GraphDB Rules and is an example of large-scale reasoning over CH data (Alexiev, 2012; Alexiev, Manov, Parvanova, & Petrov, 2013), with 4.7x reasoning expansion ratio and 900M statements. FR (Katerina Tzompanaki & Martin Doerr, 2012) is an approach of creating a set of "indexing" relations that abstract over complex CIDOC CRM networks. A number of FRs are defined across 5 types of Fundamental Classes (what, who, where, when).

As an example, the FR Thing From Place codifies the notion that a Thing may have its origin at Place if the thing (or a part of it) was used for an important activity at place, or was created at place, or was made by someone born at place, or who had its residence at place, etc.

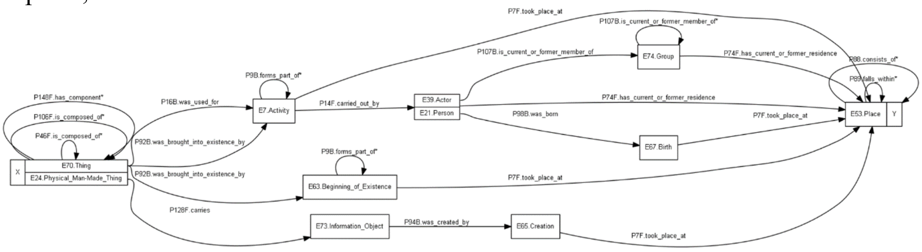
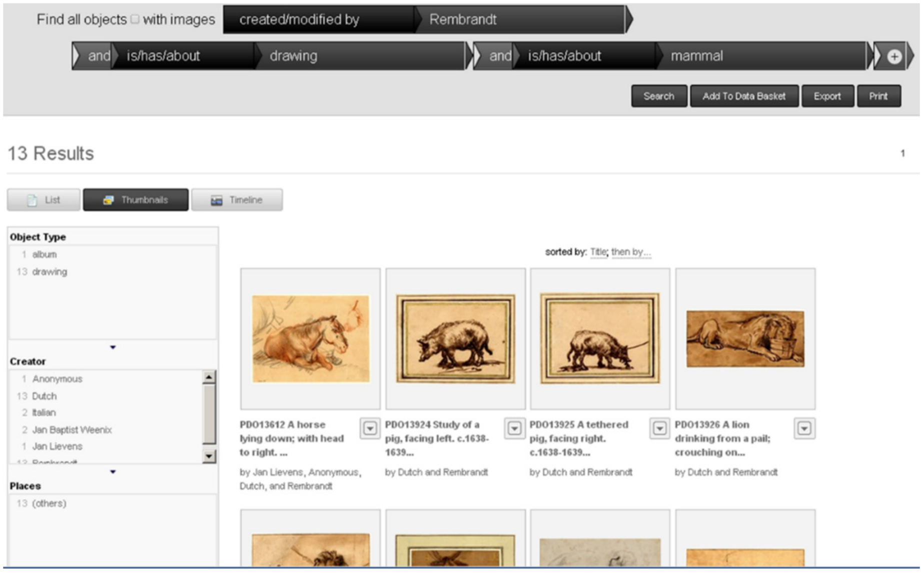


Figure 6. CRM Fundamental Relation: Thing From Place

The first version of RS semantic search implemented 23 FRs, all of them about Thing. It also included several semantic hierarchical facets: object type, creator, place, date created, etc. It implemented a natural-language-like interface for defining the query. It employs query expansion across hierarchical thesauri, e.g. searching for "Mammal" finds drawings of horses and pigs.



Find all objects □ with images created/modified by Rembrandt

and is/has/about drawing

and is/has/about mammal

13 Results

Object Type
1 album
13 drawing

Creator
1 Anonymous
13 Dutch
2 Italian
2 Jan Baptist Weenix
1 Jan Lievens

Places
13 (others)

sorted by: Title; then by...

PD013612 A horse lying down; with head to right. ...
by Jan Lievens, Anonymous, Dutch, and Rembrandt

PD013924 Study of a pig, facing left. c.1638-1639...
by Dutch and Rembrandt

PD013925 A tethered pig, facing right. c.1638-1639...
by Dutch and Rembrandt

PD013926 A lion drinking from a pail; crouching on...
by Dutch and Rembrandt

Figure 7. RS First Semantic Search: Hierarchical Query Expansion

The current version of RS semantic search implements a lot more FRs, "stored queries", ability to join against such queries, and a nicer user interface.

Pros: RS has pioneered several novel approaches in CH: CIDOC CRM representation, powerful semantic search, image annotation, saved searches, data basket, etc. It intends to be a generic art research system that can be adapted for various needs and projects.

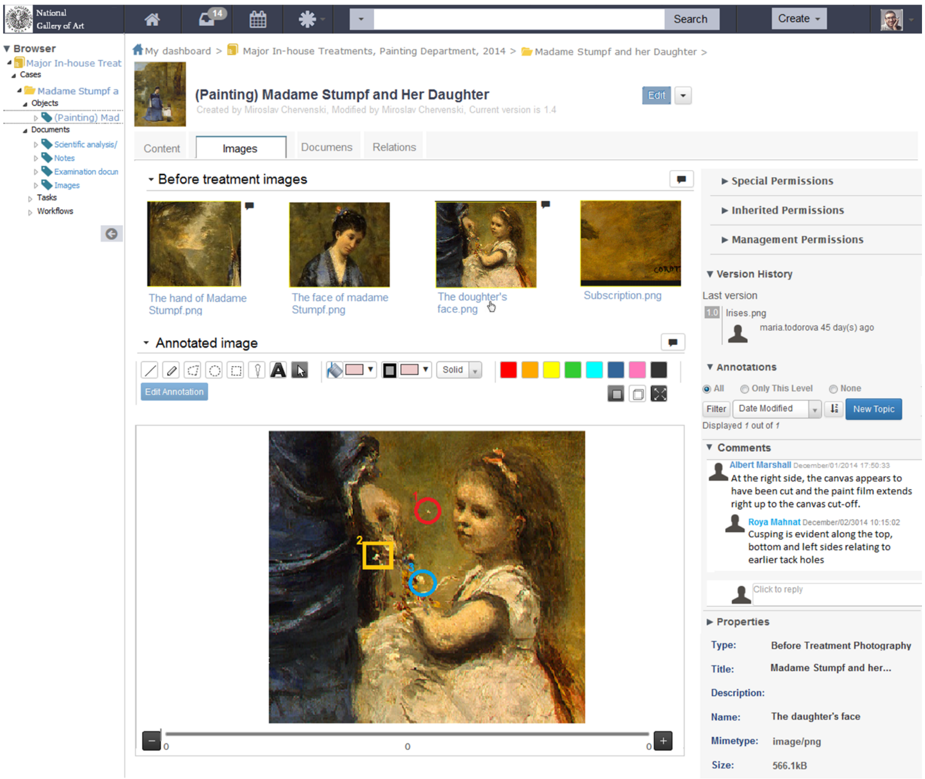
Cons: RS still has very few production users to use the system on a daily basis.

2.6 US ConservationSpace, Sirma MuseumSpace

Like RS, ConservationSpace (CS) is a Mellon-funded project that started in 2009 and spent about 3 years defining requirements, creating UI mockups, designs and RFP documentation (see old project site). The project is led by the US National Gallery of Art and includes a strong consortium. The goal is to create a system for conservation specialists, including tasks such as object examination, image annotation, process and work flows, intelligent documents, etc

Development started in 2013 and a production system was completed in 2015-16. Several Bulgarian companies were involved: Sirma Enterprise developed the system, Ontotext provided semantic database and semantic consulting. CS is now in production in all partner institutions, several other deployments are in progress, and it is being adopted for two MS programs in conservation. CS features include:

- The ability to import data from collection and digital asset management systems
- Storing data in a semantic database (Ontotext GraphDB)
- Generating user interfaces from ontologies and declarative descriptions
- Flexible access control and user rights model
- Cloud-based deployment (Software as a Service) with a full multi-tenant model (each tenant institution operates completely independently from the others)
- Capabilities that facilitate both enterprise-level and user-level customization of system object templates and code lists
- Role-based security management controls, specific to each institution's standards
- System object security controls permitting controlled access to sensitive documentation or data
- Adoption of image annotation standards in conformance with established protocols such as the International Image Interoperability Framework (IIIF)
- Extended Mirador viewer for working with images
- Dashboard customization capabilities for individual users
- Full workflow management capabilities to support the unique business processes of each institution
- Capabilities to support the use of locally preferred terminology by institutions
- Version management and rollback capabilities for key system objects
- Cultural and digital object record management and search/retrieval independent from the project/case/task system object hierarchy
- Reports on system status and activity
- Intelligent Documents (iDoc) which incorporate data entry forms and can query information from the system.
- Ability to print and export iDoc-based documents



(Painting) Madame Stumpf and Her Daughter

Created by Miroslav Chervenski, Modified by Miroslav Chervenski, Current version is 1.4

Last version

1.0 Irises.png

maria.todorova 45 day(s) ago

Displayed 1 out of 1

Albert Marshall December/01/2014 17:50:33

At the right side, the canvas appears to have been cut and the paint film extends right up to the canvas cut-off.

Roya Mahnat December/02/3014 10:15:02

Cusping is evident along the top, bottom and left sides relating to earlier tack holes

Type:	Before Treatment Photography
Title:	Madame Stumpf and her...
Description:	
Name:	The daughter's face
Mimetype:	image/png
Size:	566.1kB

Figure 8. ConservationSpace Painting Examination

CS spent significant time on user/requirements workshops, ensuring the applicability and longevity of the project. Several institutions use the system in production, and there is a thriving user community. Deployment for a new institution involves defining specific objects, work-flows and customizations, but little programming.

ConservationSpace is based on the **Sirma Enterprise Platform**, a flexible software solution that includes semantic data modelling, process management, work flow, (BPMN process definition), collaboration (contextual comments, email notifications, etc). It was deployed in a variety of domains, including Sirma MuseumSpace. That system (to be demonstrated at DiPP 2018) includes modules for curation/collection management, exhibition and loan management, conservation management, etc.

2.7 US AAC (American Art Collaborative) and linked.art

The publication of semantic data by the BM, YCBA and SAAM generated enough interest, so the American Art Collaborative (AAC, http://americanartcollaborative.org) was established as a 2-year project (from Oct 2015 to Nov 2017) with Mellon Foundation funding. 14 US museums and galleries participate in this collaboration to publish their data in RDF. Although the Getty Trust is not formally affiliated with AAC, it had

a crucial role, as the project was started by the former founder of the Getty Vocabulary Program, and two Getty staff (the semantic architect and data architect) had core involvement in developing the data model.
A lot of the technical work was done by external consultants: data conversion mostly by USC ISI students using the ISI Karma tool. Design for Context created UI mockups and implemented the Browse and Mapping/Review apps. Vladimir Alexiev and Stephen Stead provided CIDOC CRM mapping advice. Vladimir Alexiev provided detailed bug reports and semantic data publishing advice. Towards the end of the project a couple of the institutions took charge of their transformations, aiming to establish their own sustainable RDF publication.

The project did a lot of its work in the open (http://github.com/american-art/): this site has 26 repositories with common tools, data and issues (e.g. aac-alignment, aac_mappings, AAT-Term-Mappings, Semantic-UI, AAC-Instructions, linking, semantic-hosting, pubby) per-museum repositories (e.g. **DMA** for Dallas Museum of Art, **npg** for National Portrait Gallery) with source data, Karma mapping models and converted data. Semantic resources at http://data.americanartcollaborative.org/:

- Per-museum data, e.g. http://data.americanartcollaborative.org/npg
- Per-agent data, e.g. http://data.americanartcollaborative.org/npg/person-institution/44424
- Per-object data, e.g. http://data.americanartcollaborative.org/npg/object/29
- SPARQL endpoint http://data.americanartcollaborative.org/sparql. There is no editor there, so it's best to use YASGUI. E.g. http://yasgui.org/short/H1u89bnJG is a query that returns NPG objects and their images

Mapping/review app: http://review.americanartcollaborative.org/. This tool uses Ontotext's rdfpuml visualization tool (Vladimir Alexiev, 2016) and is used to both define the desired mapping and check certain semantic URLs for conformance.

The web browse app http://browse.americanartcollaborative.org shows an overview of aggregated collections, simple full-text search, individual object pages, artist pages, and statistics about number of objects per artist across collections.

(Craig Knoblock et al., 2017) describe project challenges, volumetrics and semantic conversion experience. (Fink, 2018) describes lessons learned and an overview of good practices.

Pros

- The project aggregated artwork data from 14 institutions: 233,666 Objects, 28,882 Artists and 20,446 other agents (Related Parties), comprising about 15M triples. (For comparison, the British Museum semantic data comprises 2.5M objects and 960M triples.)
- Used a harmonized data model so the data can be shown together.
- Harmonized not only data models but also value sets. AAC standardized on using Getty AAT concepts for "business classification" of various aspects as the value of crm:P2_has_type, e.g. http://vocab.getty.edu/aat/300055147 for "Gender". Furthermore, an USC ISI tool was used successfully by the institutions for linking artists to ULAN (though later a comparison to Wikidata Mix-n-Match showed that tool could have been used to better effect).

- Raised LOD awareness with the target institutions and a wider audience and mobilized inter-institutional collaboration.
- Towards the end of the project a lot of IT people and data curators from the institutions became deeply involved in the details of the semantic representation. Some of the institutions took charge of their transformations to establish a sustainable LOD publication process.
- The project created excellent use cases and UI mockups for browsing and exploration, e.g. comparing artists by style, material and genres; artwork timelines, etc.

Cons

- Started mapping without having a proper mapping specification. As a result, some mappings were reworked up to 6 times (Craig Knoblock et al., 2017). A lot of bugs were filed (total 592 issues). A lot of these are still open (107 open issues as of Mar 2018). Some were postponed for a future version, and then closed without being implemented, i.e. dismissed (e.g. mapping Exhibitions). Many issues were replicated between the different institutions, so had to be posted and fixed several times. Perhaps the most important lesson learned was that one should not attempt a massive mapping effort without having an agreed data model and strong mapping specifications (prototypical mappings): bug reports are no substitute for a proper specification.
- Some data submitted by the institutions was left unmapped and therefore not published semantically (e.g. Exhibitions, Publications/bibliographic info, Videos, etc). A lot of the use cases and mockups could not be implemented because of data omissions or insufficient harmonization of the data.
- Various details were glossed over, e.g. the Actor Image mapping disregards the SAAM flag PrimaryDisplay. This means that when an artist has many images (e.g. a photo and a self-portrait), a random one needs to be selected to display just one image (e.g. in search results). But even the old SAAM mapping had that, e.g. see Ivan Albright at SAAM: it has two links P138i_has_representation but only one of them is PE_has_main_representation.
- The mapping specification omits important details, such as URL patterns. As a result, many conversion implementers (ISI students) have made mistakes.
- Since the adopted data model (linked.art) was derived post-factum, various problems still remain.

E.g. regarding Title Types I have posted the following github issues (aac_mappings/48 and cbm/58):

- an object may have several titles of the same type, in which case their labels get mixed together
- all title types of an object are mixed together
- commonality of title types across objects is not captured
- there is no relation from title type to AAT (whereas now AAT has related concepts such as "Group Title")
- The use of aat:300404670 "preferred name" is wrong, e.g. for Group Title
- Use crm:P48_has_preferred_identifier instead of crm:P1_is_identified_by for the title id
- No need to use aat:300404012 Unique Identifier for the title id

- The title type mixes SKOS and CRM in an undisciplined way
- DisplayOrder of titles is not captured
- Group Title reflects a collection of objects so it should be modelled as crm:E78_Collection

http://linked.art emerged from the AAC effort as an application profile for CRM, i.e. a particular way of using CRM. It was created out of frustration with the complications of applying CRM (Robert Sanderson, 2016) and is promoted under the moniker Linked Open **Usable** Data (LOUD). linked.art steps on the following principles:

- CIDOC-CRM as the core ontology, giving an event-based paradigm
- The Getty Vocabularies (see next) as core sources of identity, i.e. specific object types (e.g. painting), activity types (e.g. book binding, gilding, etching), title types (e.g. artists vs repository title), etc
- JSON-LD as the primary RDF serialization. Being JSON, it is more developer-friendly than other serializations.

linked.art makes a number of simplifying assumptions, defining "a stream-lined profile of CRM for better consistency and comprehension". Its CRM Class Analysis uses 28 of the CRM classes, dismisses about 60 classes (under headings Overly Abstract, Overly Specific, Datatypes, Ineffective, Unnecessary, Incomprehensible), and introduces 7 new classes. It similarly dismisses a number of CRM's properties.

One of the most useful features of linked.art is the large number of examples (model components) that guide the semantic representation of museum data. The count of examples (Aug 2018) per area is: 42 activity, 1 concept, 2 group, 2 identifier, 2 legal, 1 name, 46 object, 12 person, 6 place, 7 set, 11 text, 2 value.

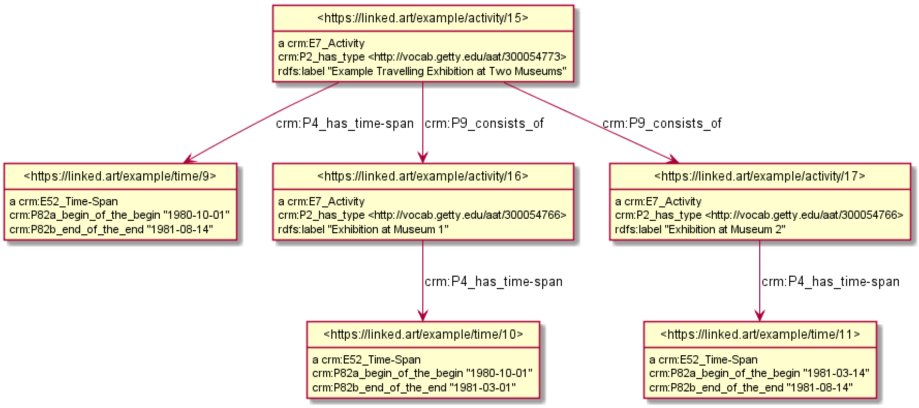

<https://linked.art/example/activity/15>
a crm:E7_Activity
crm:P2_has_type <http://vocab.getty.edu/aat/300054773>
rdfs:label "Example Travelling Exhibition at Two Museums"
crm:P4_has_time-span
crm:P9_consists_of
crm:P9_consists_of
<https://linked.art/example/time/9>
a crm:E52_Time-Span
crm:P82a_begin_of_the_begin "1980-10-01"
crm:P82b_end_of_the_end "1981-08-14"
<https://linked.art/example/activity/16>
a crm:E7_Activity
crm:P2_has_type <http://vocab.getty.edu/aat/300054766>
rdfs:label "Exhibition at Museum 1"
<https://linked.art/example/activity/17>
a crm:E7_Activity
crm:P2_has_type <http://vocab.getty.edu/aat/300054766>
rdfs:label "Exhibition at Museum 2"
crm:P4_has_time-span
crm:P4_has_time-span
<https://linked.art/example/time/10>
a crm:E52_Time-Span
crm:P82a_begin_of_the_begin "1980-10-01"
crm:P82b_end_of_the_end "1981-03-01"
<https://linked.art/example/time/11>
a crm:E52_Time-Span
crm:P82a_begin_of_the_begin "1981-03-14"
crm:P82b_end_of_the_end "1981-08-14"

Figure 9. linked.art Representation of Traveling Exhibition

Pros: linked.art is used in a number of US-based projects: AAC (post-factum), Getty Provenance Index, Getty Museum data mapping (upcoming), Pharos.net photographic consortium.

Cons: the linked.art simplifications are controversial and have not been accepted by the "mainstream" CRM SIG. Therefore, it creates a rift in the CRM community: European projects using full CRM, and US projects using linked.art.

2.8 US GVP (Getty Vocabulary Program)

The Getty Research Institute (part of the Getty Trust) manages the Getty Vocabulary Program (GVP), which publishes some of the core and most respected CH thesauri. Getty's vision for the GVP thesauri:

- The thesauri are interconnected with each other. For example, TGN uses AAT for place types, ULAN uses AAT for artist types (roles) and event types, and TGN for places of birth, death, etc.
- The thesauri provide shared data for Getty's own databases and systems: Arches (see next section), Provenance Index, Getty Museum. Getty site-wide web search, AATA Online (bibliography of art and architecture).
- AAT is translated internationally to Dutch, Spanish, German, Chinese, and work on a Swiss Art and Architecture thesaurus is pending. The International Terminology Working Group (ITWG) coordinates this work.
- The thesauri are coreferenced to other relevant thesauri, for example LCSH and VIAF (ULAN is completely incorporated in VIAF, though with a narrower scope of data, e.g. VIAF doesn't include artist relations).
- The thesauri are used by various external databases, projects, search engines and Collection Management Systems.

GVP started a LOD publication program in 2013. To date it has published the following thesauri as LOD at http://vocab.getty.edu, sharing the same basic semantic representation, and publicized in blog posts by J.Cuno, CEO of the Getty Trust.

- Art and Architecture Thesaurus (AAT): Feb 2014
- Thesaurus of Geographic Names (TGN): Aug 2014
- Union List of Artist Names (ULAN): Mar 2015

GVP LOD was presented at the CIDOC Congress in Dresden in 2014 (Vladimir Alexiev, 2014). Ontotext provided the following services as part of this project:

- Semantic/ontology development (Alexiev, 2015b)
- Contributed to the ISO 25964 ontology, which is the latest standard on thesauri. Provided implementation experience, suggestions and fixes. Published on varieties of Broader relations (Vladimir Alexiev, Jutta Lindenthal, & Antoine Isaac, 2015)
- Complete mapping specification. Helped implement R2RML scripts working off Getty's Oracle database, contribution to Perl implementation (RDB2RDF), R2RML extension (rrx:languageColumn)
- Worked with a wide External Reviewers group (people from OCLC, Europeana, ISO 25964 working group, etc.)
- GraphDB semantic repository, clustered for high-availability
- Semantic application development, user interface, technical consulting
- SPARQL 1.1 compliant endpoint, comprehensive documentation (Vladimir Alexiev, Joan Cobb, Gregg Garcia, & Patricia Harpring, 2015), sample queries (Alexiev, 2015a). Per-entity export files, explicit/total data dumps.
- Help desk / support on twitter and google group (continuing until now)

GVP Ontologies. GVP LOD uses 11 ontologies to represent all data present in the thesauri. In addition, it features the GVP LOD ontology that has 10 classes and 177 properties. The ontology is documented with Parrot and registered in Linked Open Vocabularies to facilitate discovery. The GVP ontology captures specific Getty classes and properties that are not available in SKOS, SKOS-XL and ISO 25964. Nevertheless, it maps to these established ontologies, so one can also consume the data using only these ontologies.

- Includes these specific node types: gvp:Facet, gvp:Hierarchy, gvp:GuideTerm, gvp:Concept, gvp:ObsoleteSubject. These are implemented as subclasses of skos:Concept, skos:Collection, iso:ThesaurusArray.
- Most of the properties are GVP Associative Relations, defined as sub-properties of skos:related. These were described by GVP domain experts in Excel, and we generated the ontological definitions from that.
- The inference from GVP custom properties to standard properties is shown below (blue=standard relation, black=GVP relation, bold=transitive closure, red=restriction)

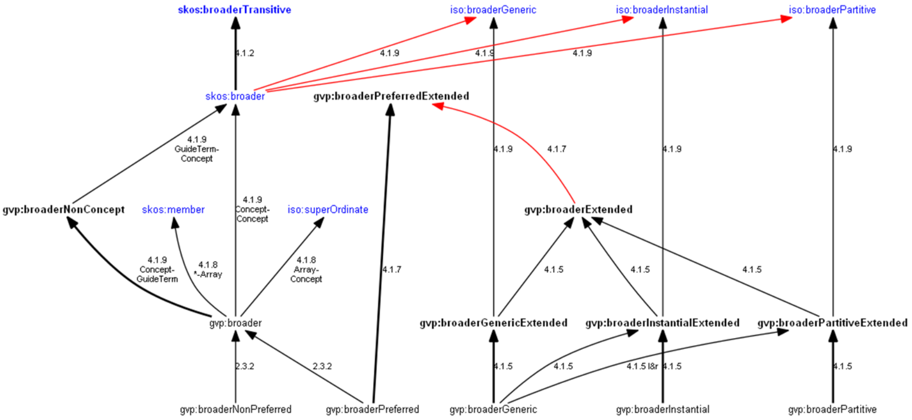
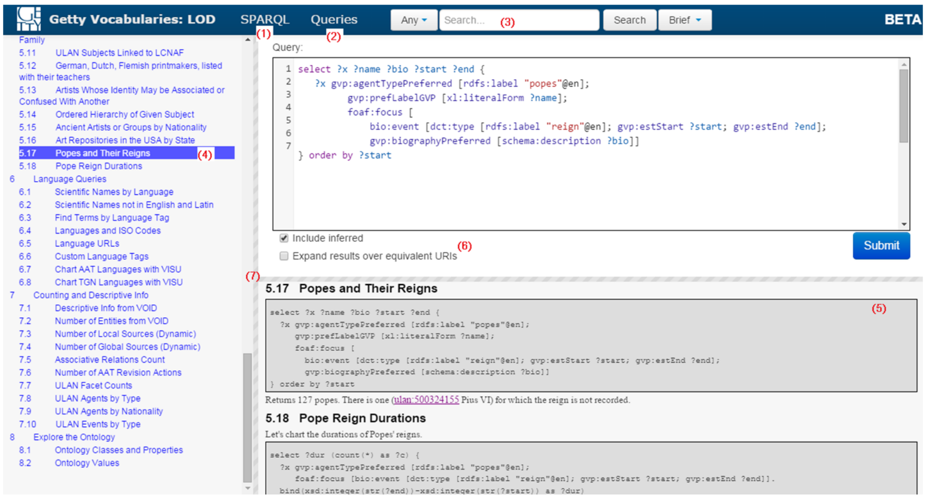


Figure 10. GVP Hierarchical Relation Inference

Documentation, Sample Queries, Support. GVP LOD has set best-practice standards for good quality CH LOD semantic publishing.

- Comprehensive documentation (100 pages) that describes all aspects of representation, semantic resolution, URLs, content negotiation, It is kept up to date, with complete revision notes.
- There are about 100 sample queries, covering topics such as Full-Text Search (many external systems use GVP LOD for auto-completion), getting various kinds of information, TGN and ULAN specific queries (e.g. by geographic proximity), language-related queries, making graphs and charts, etc. There is a special Sample Queries UI that shows the outline of queries (TOC), the description of each query, and allows the user to easily select and execute the query.



Family
5.11 ULAN Subjects Linked to LCNAF
5.12 German, Dutch, Flemish printmakers, listed with their teachers
5.13 Artists Whose Identity May be Associated or Confused With Another
5.14 Ordered Hierarchy of Given Subject
5.15 Ancient Artists or Groups by Nationality
5.16 Art Repositories in the USA by State
5.17 Popes and Their Reigns
5.18 Pope Reign Durations
6 Language Queries
6.1 Scientific Names by Language
6.2 Scientific Names not in English and Latin
6.3 Find Terms by Language Tag
6.4 Languages and ISO Codes
6.5 Language URLs
6.6 Custom Language Tags
6.7 Chart AAT Languages with VISU
6.8 Chart TGN Languages with VISU
7 Counting and Descriptive Info
7.1 Descriptive Info from VOID
7.2 Number of Entities from VOID
7.3 Number of Local Sources (Dynamic)
7.4 Number of Global Sources (Dynamic)
7.5 Associative Relations Count
7.6 Number of AAT Revision Actions
7.7 ULAN Facet Counts
7.8 ULAN Agents by Type
7.9 ULAN Agents by Nationality
7.10 ULAN Events by Type
8 Explore the Ontology
8.1 Ontology Classes and Properties
8.2 Ontology Values

Query:

```
select ?x ?name ?bio ?start ?end {
  ?x gvp:agentTypePreferred [rdfs:label "popes"@en];
      gvp:prefLabelGVP [xl:literalForm ?name];
      foaf:focus [
         bio:event [dct:type [rdfs:label "reign"@en]; gvp:estStart ?start; gvp:estEnd ?end];
         gvp:biographyPreferred [schema:description ?bio]]
} order by ?start
```

☑ Include inferred

☐ Expand results over equivalent URIs

5.17 Popes and Their Reigns

```
select ?x ?name ?bio ?start ?end {
  ?x gvp:agentTypePreferred [rdfs:label "popes"@en];
     gvp:prefLabelGVP [xl:literalForm ?name];
     foaf:focus [
       bio:event [dct:type [rdfs:label "reign"@en]; gvp:estStart ?start; gvp:estEnd ?end];
       gvp:biographyPreferred [schema:description ?bio]]
} order by ?start
```

Returns 127 popes. There is one (ulan:500324155 Pius VI) for which the reign is not recorded.

5.18 Pope Reign Durations

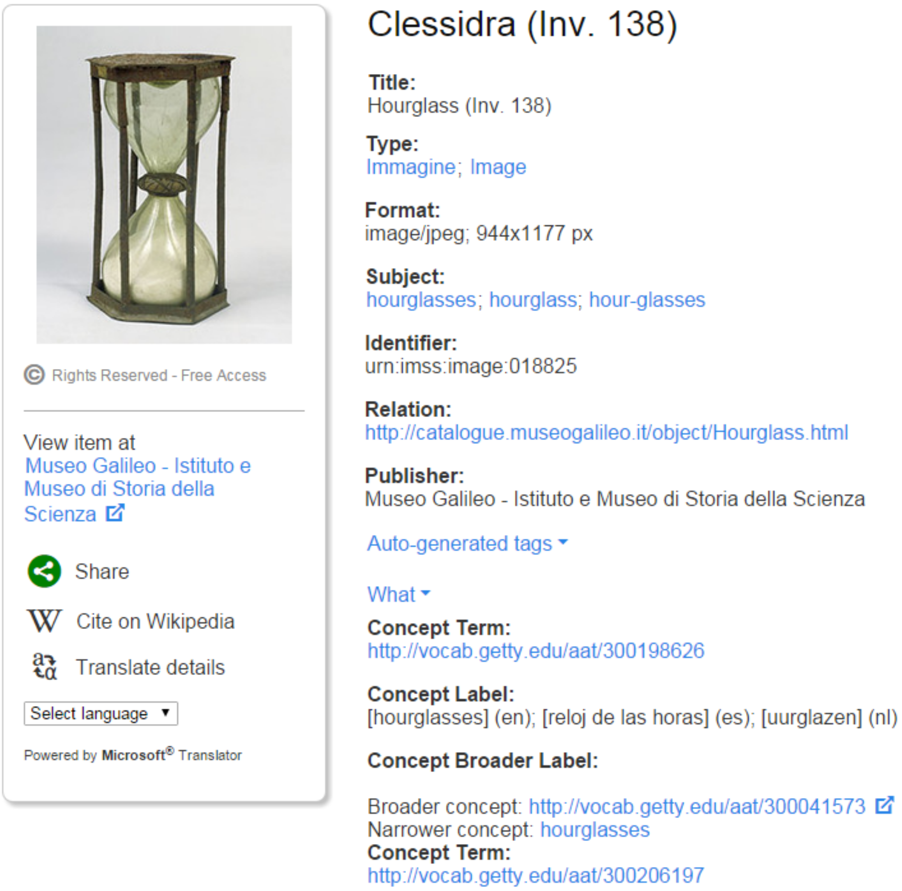
Let's chart the durations of Popes' reigns.

```
select ?dur (count(*) as ?c) {
  ?x gvp:agentTypePreferred [rdfs:label "popes"@en];
     foaf:focus [bio:event [dct:type [rdfs:label "reign"@en]; gvp:estStart ?start; gvp:estEnd ?end]].
  bind(xsd:integer(str(?end))-xsd:integer(str(?start)) as ?dur)
```

Figure 11. GVP LOD Sample Queries UI

- The GVP UI includes other convenient features, such as full-text search, exploring data, download in a variety of semantic formats (RDF/XML, NTriples, RDF/JSON, Turtle, JSONLD), bidirectional links between LOD and the traditional website. There is a community support group that is monitored regularly, questions are answered, additional queries are added, and issues are resolved.
- GVP has a comprehensive URL strategy that covers all objects and sub-objects. The stability and permanence of URLs is guaranteed by Getty and doesn't change over time (e.g. with new versions), which is extremely important for the consumers of this data (CH institutions that embed GVP thesaurus references in their own data). Obsolete concepts are not deleted for 5 years, rather they are marked as obsoleted, with potentially a dct:isReplacedBy link to the new concept.
- The various semantic formats can be downloaded by extension or through content negotiation (Accept header with appropriate MIME type). All URLs have proper semantic resolution, which was validated with Vapour. GVP provides per-entity download, which includes not just the immediate triples but all nodes and triples of the "business object". In addition, complete downloads (dumps) per thesaurus are available.
- Dataset Description: GVP LOD uses 15 external ontologies for machine-readable description of the dataset, SPARQL endpoint, preferred prefix, used vocabularies, number of triples per property, number of entities per class, etc. We used several descriptive ontologies to cater to different kinds of software agents, enabling dataset discovery and crawling. For example, the datasets of each vocabulary are declared void:Dataset, dct:Dataset, dcat:Dataset, adms:Asset, cc:Work, dct:Collection. There is good agreement between the conceptual models of the main descriptive ontologies (VOID, DCAT, ADMS), which makes this possible. Complete licensing info, keywords, subjects, crawling entry points (void:rootResource) are described.

GVP LOD Uses: GVP LOD has found a wide variety of uses in the CH community: over 50 actual and potential uses. The thesauri are used by many CH institutions (including the Google Cultural Institute) and CH-related software (including Gallery Systems TMS, which is widely used in the US). The reliability of the GVP SPARQL endpoint is such that many use the thesauri directly, without a need to copy them locally. Above we saw that AAT is used crucially in the linked.art semantic profile, to describe specific semantic types (e.g. painting, gilding, author's title, etc)


Rights Reserved - Free Access
Share
Cite on Wikipedia
Translate details
Select language ▾
Powered by Microsoft® Translator

Clessidra (Inv. 138)

Title:
Hourglass (Inv. 138)

Type:
Immagine; Image

Format:
image/jpeg; 944x1177 px

Subject:
hourglasses; hourglass; hour-glasses

Identifier:
urn:imss:image:018825

Relation:
http://catalogue.museogalileo.it/object/Hourglass.html

Publisher:
Museo Galileo - Istituto e Museo di Storia della Scienza

View item at
Museo Galileo - Istituto e Museo di Storia della Scienza

Auto-generated tags

What

Concept Term:
http://vocab.getty.edu/aat/300198626

Concept Label:
[hourglasses] (en); [reloj de las horas] (es); [uurglazen] (nl)

Concept Broader Label:

Broader concept: http://vocab.getty.edu/aat/300041573
Narrower concept: hourglasses
Concept Term:
http://vocab.getty.edu/aat/300206197

Figure 12. GVP Use in Europeana

Pros: GVP data has been modelled comprehensively, and auxiliary aspects were taken into account such as proper semantic resolution, serving useful entities, licensing, dataset description. This LOD publication has been praised as a comprehensive example to be followed by other CH publications. Getty took care of all aspects of documentation, hosting and support, so GVP LOD is used widely.

Cons: Some people find the representation too complex, since it exposes all aspects of the data. Thus, Getty is considering serving different profiles of the data, e.g. simple SKOS that conflates the difference between guide terms and concepts, without label metadata, etc.

2.9 Cultural Periods and Styles

Dealing with cultures and periods is of prime importance in art research. Getty AAT considers culture, peoples, ethnic groups, historic periods, art movements, and even religion in a uniform way, since any of these can generate related artworks. Some examples: Stone Age, Christianity, Alhambra style, Reign of the Knight Templars in Malta, Impressionism, Nazism

CRM's E4 Period is a complex cultural phenomenon that has spatial and temporal extent, a cultural/historic dimension, and may be dis-continuous (see more at the CRM Tutorial). Two co-extensive periods are not necessarily the same. E.g. the Nazi occupation of France and the French resistance movement are co-extensive, but these are distinct, opposing cultural phenomena. There are a few projects/datasets that try to build databases of periods:

- Getty AAT Periods and Styles: 5569 ethnic and artistic styles, using this query. Does not include date info.
- British Museum thesauri: over 6000. Does not include date info.
- Wikidata: only about 396 but see discussion on WikiProject Visual arts: Item structure: Art_movements: Matching Periods and Styles for trying to bring AAT, BM and WD together.
- PeriodO: A gazetteer of period definitions for linking and visualizing data.
- STAR.Timeline: treatment of archeological time periods. Has a UI demo and REST API returning JSON. Searching by date-range returns only "correlated" periods, using a measure of closeness that considers relation and the duration of query and found period. E.g. searching for "1701-1800" returns "18TH CENTURY AD" and 9 other periods. One of them is "NAPOLEONIC WARS", which does not intersect with the 18th century, but is right after it, so is considered related.

2.10 Iconography

Iconography studies the identification, description, and interpretation of the content of images: the subjects depicted, the particular compositions and details used to do so, and other "standard" elements that are distinct from artistic style. Thus iconography is the art and science of capturing subjects that often appear in artworks. Two iconographic datasets are available.

Iconclass. This is a well-known Dutch iconographic effort maintained by RKD. Iconclass includes three sets of data:

- **Classification System**: 28,000 hierarchically ordered definitions divided into ten main divisions. Each definition consists of an alphanumeric classification code (notation) and the description of the iconographic subject (textual correlate). The definitions are used to index, catalogue and describe the subjects of images represented in works of art, reproductions, photographs and other sources. Example of a biblical topic:

```
7 Bible
71 Old Testament
71H story of David
71H7 David and Bathsheba (2 Samuel 11-12)
71H71 David, from the roof (or balcony) of his palace, sees Bathsheba bathing
71H713 Bathsheba receives a letter from David
71H7131 Bathsheba (alone) with David's letter
```

- **Alphabetical Index**: 14,000 keywords used for locating the notation and its textual correlate needed to describe and/or index an image.
- **Bibliography**: 40,000 references to books and articles of iconographical interest (not yet online).

Iconclass has a comprehensive and complicated notation system including "auxiliaries" that allow a huge number of combinations (about 1.3 million notations with all keys and children fully expanded):

- **Bracketed text**, e.g. 25G41(**ROSE**) meaning "rose"
- **Key** (+digits), e.g. 25F23(LION)(**+12**) meaning "heraldic lion"
- **Queuing of keys** (catenating +digits), e.g. 25FF241(+5**11**) meaning "unicorn with nose or tusk in an unusual place"
- **Doubling of letter** to modify the meaning, e.g.
  - Animals: 25F Animals vs 25F**F** fabulous animals
  - The (nude) human figure: 31A male vs 31A**A** female
  - Wedding feast/meal: 42D25 indoors vs 42D**D**25 out of doors
- **Structural digit**: indicates important episodes in a character's lifetime, e.g.
  - For saints, 2 means early life, e.g. 11H(FRANCIS)**2** "**early life** of St. Francis of Assisi"
  - For classical gods, 2 means love-affairs, e.g. 92B3**2** "**love-affairs** of Apollo"

Iconclass is available in numerous languages (Dutch, English, French, German, Italian, Finnish) through:

- Iconclass Browser, e.g. http://www.iconclass.org/rkd/94L/ is Hercules
- Iconclass LOD, e.g. http://iconclass.org/94L is the semantic URL for Hercules. It's available as RDF and JSON (but not JSON-LD)
- FINTO (the Finnish Thesaurus and Ontology Service) has an excellent Iconclass browser with alphabetical and hierarchical browsing. E.g. Hercules is at https://finto.fi/ic/en/page/94L, and that page offers RDF/XML, TURTLE and JSON-LD downloads

There are several art search systems based on Iconclass, including institutional and commercial. E.g. Brill Arkyves is a commercial database, a single access point for thematic searches across a wide variety of cultural heritage collections.

Getty Iconography Authority (Patricia Harpring, 2016). IA includes all subjects except those that belong to AAT (general concepts), TGN (real places), ULAN (real agents) or CONA (real artworks). The scope of IA includes:

- Character, Fictional Person, Named Animal, Event/Narrative, Fictional Place, Allegory/Symbolism, Fictional Built Work, Fictional Literature, Religion/Mythology/Legend (as described in CONA section 3.6.3.18)

- Person (character), animal (character), event, imaginary place (as described in CCO section A.4.2.2.5.2)

While Iconclass is well developed but focuses on ancient mythology and Christian religious iconography, IA is in development and has wider remit. IA includes:

- Multilingual labels and descriptions
- IA hierarchical organization, including Root Record, Facets, Guide Terms
- Associative relations within IA
- Relations from IA to the other Getty vocabularies

The diagram below illustrates a LOD mapping for the following facts about **ia:1000042** Hercules. A lot of info is packed into this graph!

- Part of: IA Thesaurus
- Record Type: Religion/Mythology/Legend
- Concept sources and Locators in those sources
- Same As: iconclass:94L
- Labels (names) and their Sources
- Description: "Probably based on an actual historical figure, a king of ancient Argos. The legendary figure was the son of Zeus and Alcmene ..."
- Hierarchy: Classical Mythology> Greek heroic legends> Story of Hercules
- Birth place: tgn:7010720 Argos, Associated place: tgn:7029383 Thebes. Notice that mythological characters may be related to real historic places
- Father: Zeus (Greek god), with comment: "was his favorite son"
- Mother: Alcmene (Greek heroine)
- Role: Greek hero, king
- Participated in event: Labors of Hercules, Clean the stables of King Augeas
- Additional associative relations:
  - "Zeus" has spouse "Alcmena"
  - "Labors of Hercules" has subevent "Clean the stables of King Augeas" (i.e. we enumerate the 12 labors of Hercules as separate events)
  - "Augeas" participated in event "Clean the stables of King Augeas" (presumably Augeas asked/motivated Hercules to perform this labor)

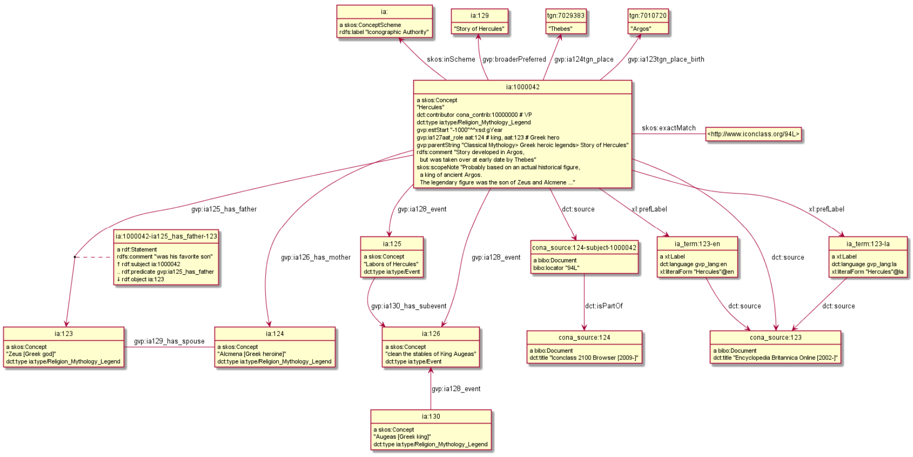

ia:
a skos:ConceptScheme
rdfs:label "Iconographic Authority"
ia:129
"Story of Hercules"
tgn:7029383
"Thebes"
tgn:7010720
"Argos"
skos:inScheme
gvp:broaderPreferred
gvp:ia124tgn_place
gvp:ia123tgn_place_birth
ia:1000042
a skos:Concept
"Hercules"
dct:contributor cona_contrib:10000000 # VP
dct:type ia:type/Religion_Mythology_Legend
gvp:estStart "-1000"^^xsd:gYear
gvp:ia127aat_role aat:124 # king, aat:123 # Greek hero
gvp:parentString "Classical Mythology> Greek heroic legends> Story of Hercules"
rdfs:comment "Story developed in Argos,
but was taken over at early date by Thebes"
skos:scopeNote "Probably based on an actual historical figure,
a king of ancient Argos.
The legendary figure was the son of Zeus and Alcmene ..."
skos:exactMatch
<http://www.iconclass.org/94L>
gvp:ia125_has_father
gvp:ia128_event
dct:source
xl:prefLabel
xl:prefLabel
ia:1000042-ia125_has_father-123
a rdf:Statement
rdfs:comment "was his favorite son"
↑ rdf:subject ia:1000042
.. rdf:predicate gvp:ia125_has_father
↓ rdf:object ia:123
gvp:ia126_has_mother
ia:125
a skos:Concept
"Labors of Hercules"
dct:type ia:type/Event
gvp:ia128_event
cona_source:124-subject-1000042
a bibo:Document
bibo:locator "94L"
ia_term:123-en
a xl:Label
dct:language gvp_lang:en
xl:literalForm "Hercules"@en
dct:source
ia_term:123-la
a xl:Label
dct:language gvp_lang:la
xl:literalForm "Hercules"@la
gvp:ia130_has_subevent
dct:isPartOf
dct:source
dct:source
ia:123
a skos:Concept
"Zeus [Greek god]"
dct:type ia:type/Religion_Mythology_Legend
gvp:ia129_has_spouse
ia:124
a skos:Concept
"Alcmena [Greek heroine]"
dct:type ia:type/Religion_Mythology_Legend
ia:126
a skos:Concept
"clean the stables of King Augeas"
dct:type ia:type/Event
cona_source:124
a bibo:Document
dct:title "Iconclass 2100 Browser [2009-]"
cona_source:123
a bibo:Document
dct:title "Encyclopedia Britannica Online [2002-]"
gvp:ia128_event
ia:130
a skos:Concept
"Augeas [Greek king]"
dct:type ia:type/Religion_Mythology_Legend

Figure 13 Semantic Representation of Hercules Info in Getty IA

3. Conclusions

We presented an overview of CH ontologies, datasets and semantic projects. Following earlier researcher communities in Life Sciences, the CH and DH communities have come to the conclusion that semantic data integration is the key to interlinking CH data across time and borders, future-proofing it, and enabling DH research based on Big-Data, semantic linking, semantic text enrichment, inference, network analysis, network visualization, etc. CH LOD (also called LODLAM) remains an exciting area of research as more and more institutions publish their data in a semantic way, enabling new modes of consumption.

Acknowledgements

Part of this work was done in a consulting project with the Canadian Heritage Information Network (CHIN) to create a strategy for transitioning the Canadian national museum data aggregation to LOD.

References

Alexiev, V. (2012). Implementing CIDOC CRM Search Based on Fundamental Relations and OWLIM Rules. In *Workshop on Semantic Digital Archives (SDA 2012), part of International Conference on Theory and Practice of Digital Libraries (TPDL 2012)*. Paphos, Cyprus: CEUR WS Vol.912. Retrieved from http://ceur-ws.org/Vol-912/paper8.pdf

Alexiev, V. (2015a). *Getty Vocabularies: LOD Sample Queries* (3.2). Getty Research Institute. Retrieved from http://vocab.getty.edu/doc/queries/

Alexiev, V. (2015b). *Getty Vocabulary Program (GVP) Ontology* (3.2). Getty Research Institute. Retrieved from http://vocab.getty.edu/ontology

Alexiev, V., Manov, D., Parvanova, J., & Petrov, S. (2013). Large-scale Reasoning with a Complex Cultural Heritage Ontology (CIDOC CRM). In *Workshop Practical Experiences with CIDOC CRM and its Extensions (CRMEX 2013) at TPDL 2013*. Valetta, Malta. Retrieved from http://www.ontotext.com/sites/default/files/publications/CRM-reasoning.pdf

Craig Knoblock, Pedro Szekely, Eleanor Fink, Duane Degler, David Newbury, Robert Sanderson, … Yixiang Yao. (2017). Lessons Learned in Building Linked Data for the American Art Collaborative. In *International Semantic Web Conference (ISWC)*. Retrieved from http://usc-isi-i2.github.io/papers/knoblock17-iswc.pdf

Dominic Oldman, & Donna Kurtz. (2014). *CRM Primer v1.1*. Retrieved from http://www.cidoc-crm.org/sites/default/files/CRMPrimer_v1.1_1.pdf

Dominic Oldman, Joshan Mahmud, & Alexiev, V. (2013). *The Conceptual Reference Model Revealed. Quality contextual data for research and engagement: A British Museum case study* (p. 359 pages). ResearchSpace Project. Retrieved from http://confluence.ontotext.com/display/ResearchSpace/BM+Mapping

Dominic Oldman, Martin Doerr, Gerald de Jong, Barry Norton, & Thomas Wikman. (2014). Realizing Lessons of the Last 20 Years: A Manifesto for Data Provisioning & Aggregation Services for the Digital Humanities (A Position Paper). *D-Lib Magazine*, *20*(7). https://doi.org/10.1045/july2014-oldman

Europeana. (2017). *Definition of the Europeana Data Model v5.2.8*. Retrieved from https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation//EDM_Definition_v5.2.8_102017.pdf

Fink, E. E. (2018). *American Art Collaborative (AAC) Linked Open Data (LOD) Initiative: Overview and Recommendations for Good Practices*. American Art Collaborative.

Katerina Tzompanaki, & Martin Doerr. (2012). *Fundamental Categories and Relationships for intuitive querying CIDOC-CRM based repositories* (No. TR-429). ICS-FORTH. Retrieved from http://www.cidoc-crm.org/docs/TechnicalReport429_April2012.pdf

Martin Dörr. (2018, May). *CRM Family of Models*. Presented at the 2nd Data for History workshop, Lyon France. Retrieved from http://dataforhistory.org/sites/default/files/dfh20180525_doerr.pdf

Parvanova, J., Alexiev, V., & Kostadinov, S. (2013). RDF Data and Image Annotations in ResearchSpace. In *Collaborative Annotations in Shared Environments: metadata, vocabularies and techniques in the Digital Humanities (DH-CASE 2013). Collocated with DocEng 2013*. Florence, Italy. Retrieved from http://www.ontotext.com/sites/default/files/publications/Parvanova2013-SemanticAnnotation.pdf

Patricia Harpring. (2016, June). Getty Iconography Authority: Introduction and Overview. Getty Vocabulary Program. Retrieved from http://www.getty.edu/research/tools/vocabularies/cona_ia_in_depth.pdf

Patrick Le Boeuf, Martin Doerr, Christian Emil Ore, & Stephen Stead. (2018). *Definition of the CIDOC Conceptual Reference Model v6.2.3*. Retrieved from http://www.cidoc-crm.org/releases_table

Robert Sanderson. (2016, December). *Community Challenges for Practical Linked Open Data*. Presented at the Linked Pasts keynote, Madrid, Spain. Retrieved from https://www.slideshare.net/azaroth42/community-challenges-for-practical-linked-open-data-linked-pasts-keynote

Vladimir Alexiev. (2014). *Getty Vocabulary Program LOD: Ontologies and Semantic Representation*. Dresden, Germany. Retrieved from http://vladimiralexiev.github.io/pres/20140905-CIDOC-GVP/GVP-LOD-CIDOC.pdf

Vladimir Alexiev. (2016). RDF by Example: rdfpuml for True RDF Diagrams, rdf2rml for R2RML Generation. In *Semantic Web in Libraries 2016 (SWIB 16)*. Bonn, Germany. Retrieved from http://vladimiralexiev.github.io/pres/20161128-rdfpuml-rdf2rml/index-full.html

Vladimir Alexiev, Andrey Tagarev, & Laura Tolosi. (2016). *Europeana Food and Drink Semantic Demonstrator Extended* (Deliverable No. D3.20d). Europeana Food and Drink project. Retrieved from http://vladimiralexiev.github.io/pubs/Europeana-Food-and-Drink-Semantic-Demonstrator-Extended-(D3.20d).pdf

Vladimir Alexiev, Cobb, J., Gregg Garcia, & Patricia Harpring. (2015). *Getty Vocabularies Linked Open Data: Semantic Representation* (3.2). Getty Research Institute. Retrieved from http://vocab.getty.edu/doc/

Vladimir Alexiev, & Dilyana Angelova. (2015, July). *O is for Open: OAI and SPARQL interfaces for Europeana*. Presented at the Europeana Creative Culture Jam, Vienna, Austria. Retrieved from http://vladimiralexiev.github.io/pubs/O_is_for_Open_(CultJam_201507)_poster.pdf

Vladimir Alexiev, Jutta Lindenthal, & Antoine Isaac. (2015). On the composition of ISO 25964 hierarchical relations (BTG, BTP, BTI). *International Journal on Digital Libraries*, 1–10. https://doi.org/10.1007/s00799-015-0162-2

Received: June 15, 2018
Reviewed: June 28, 2018
Finally Accepted: July 07, 2018