

# The Rijksmuseum Collection as Linked Data

**Editor(s):** Harith Alani, The Open University, UK

**Solicited review(s):** Eetu Mäkelä, Aalto University, Finland; Dana Dannélls, University of Gothenburg, Sweden; Mariana Damova, Mozaika, Bulgaria; One anonymous reviewer

Chris Dijkshoorn<sup>a,b,\*</sup>, Lizzy Jongma<sup>b</sup>, Lora Aroyo<sup>a</sup>, Jacco van Ossenbruggen<sup>c</sup>, Guus Schreiber<sup>a</sup>, Wesley ter Weele<sup>d</sup>, Jan Wielemaker<sup>a</sup>

<sup>a</sup> *Department of Computer Science, The Network Institute, VU University Amsterdam, the Netherlands*

<sup>b</sup> *Rijksmuseum Amsterdam, the Netherlands*

<sup>c</sup> *Centrum Wiskunde & Informatica, the Netherlands*

<sup>d</sup> *History Department, VU University Amsterdam, the Netherlands*

**Abstract.** Many museums are currently providing online access to their collections. The state of the art research in the last decade shows that it is beneficial for institutions to provide their datasets as Linked Data in order to achieve easy cross-referencing, interlinking and integration. In this paper, we present the Rijksmuseum linked dataset (accessible at <http://datahub.io/dataset/rijksmuseum>), along with collection and vocabulary statistics, as well as lessons learned from the process of converting the collection to Linked Data. The version of March 2016 contains over 350,000 objects, including detailed descriptions and high-quality images released under a public domain license.

**Keywords:** Linked Data, Open Data, Image Collections, Cultural Heritage, Museums

## 1. Introduction

Publishing cultural heritage collections as Linked Data improves reusability of the data and allows for easier integration with other data sources [1,14]. Concepts providing context for collection items are often shared among multiple cultural heritage organisations, which is an ideal basis for creating connections between collections and allowing reuse of information [10,6]. The availability of data models tailored towards publishing cultural heritage data helps to make the data available in an interoperable way [3,4]. These benefits have become apparent to the sector, resulting in an increase of attention and the development of methodologies to help institutions overcome the hurdles involved in publishing data according to the Linked Data principles [1,9,14].

The Linked Data version of the Rijksmuseum collection has some unique features. The data is a re-

sult of a joint effort between the museum, CWI and VU University Amsterdam and has evolved with input from many research projects [13,16,17]. Nowadays, employees of the museum are in control of the publishing process, creating and maintaining a conversion layer from collection management system to Linked Data. The museum's digitisation process includes the use of external datasets for adding contextual concepts (e.g. creator or technique), creating manually curated links towards external datasets [8]. The data is continuously extended: every day new objects and descriptions are added and both metadata and images are released under open licenses when possible.

This paper describes the current state of the Rijksmuseum Linked Data and provides insights into the lessons learned during its creation. In the next section, we describe the characteristics of the Rijksmuseum collection and its digitisation process. The historical development of the dataset is given in Section 3 and the conversion approach in Section 4. Sections 5 and 6 provide details on the data model and the number of digital objects currently available. In Section 7 we give

---

\*Corresponding author, e-mail: [c.r.dijkshoorn@vu.nl](mailto:c.r.dijkshoorn@vu.nl)

an overview of the links from collection objects to external data sources. In Section 8 we illustrate uses of the data, before we conclude in Section 9 with a discussion.

## 2. The Rijksmuseum in a digital age

The Rijksmuseum Amsterdam is one of the most visited museums in the Netherlands, with a mission to provide a representational overview of Dutch art from the Middle Ages onwards. It is well known for its Golden Age paintings, including artworks by Rembrandt and Vermeer. The collection comprises over a million objects, of which only a fraction can be on display at a given time. To open up the remaining collection the museum started digitising objects and publishing them online.

Digitising large collections is a time consuming and costly endeavour. To address the backlog of items to be digitised, the Rijksmuseum started a dedicated digitisation project, employing catalogers and a professional photographer. The catalogers register objects in the collection management system and describe the objects, using structured vocabularies if available [8]. The photographer takes high-quality images which are released under a public domain license when possible, waiving the rights of the museum.

The digitised collection items are accessible through the website of the museum. Online visitors can explore the collection using categories or they can search for specific keywords. The presentation of the website focusses on high-quality images of collection objects, encouraging users to save, manipulate, and share them [7]. Developers can use an Application Programming Interface (API) to get access to information about the collection objects, sub-collections created by users, and event information<sup>1</sup>.

## 3. History of the Rijksmuseum Linked Data

The Linked Data version of the Rijksmuseum dataset has a long history, influenced by a number of research projects. A first Resource Description Framework (RDF) version comprising 750 top pieces was created by converting a datadump from an educational database [5]. As a next step, in an effort to integrate Dutch cultural heritage collections, the datamodel was

changed to follow the VRA Core specification<sup>2</sup>, with the key advantages of allowing the use of Dublin Core constructs<sup>3</sup> and making a distinction between the physical artwork and its digital representations. The metadata values of objects were represented in plain text.

In a next version, contextual concepts from in-house thesauri of the Rijksmuseum were aligned with the Getty thesauri and WordNet, resulting in a dataset of 27,993 triples [13]. At the time, the Getty vocabularies were only available under license and in XML format, which resulted in the need for an internally maintained conversion to RDF. In a similar effort, the vocabulary Iconclass was converted and aligned using the Simple Knowledge Organization System (SKOS) to formalise its structure [6]. The experiences gained served as input for the SKOS specification.

The Rijksmuseum dataset was one of the first entries in the Europeana Thought Lab<sup>4</sup>, an initiative for showcasing experimental technologies. This entry marks the first conversion of all available Rijksmuseum collection data: 46,000 objects with images were obtained from the collection management system and converted to comply with the VRA data model. The experience of modelling the complete collection and integrating it with collections from other institutions required the ability to model different (potentially conflicting) metadata records from different sources describing the same artwork. These and other requirements which were gathered influenced the creation of the Europeana Data Model [4].

The Europeana Data Model today has a set of core and contextual classes that can capture collection information. The data model is designed with reuse of existing classes and properties in mind. It includes elements from the Dublin Core metadata initiative and the Object Reuse and Exchange definition of the Open Archives Initiative<sup>5</sup>. Cultural heritage organisations can extend the set of classes and properties when needed, reusing elements of other data models or by defining their own. The possibility of making the collection available on the Europeana portal led to the museum taking matters into their own hands. We describe the resulting conversion of collection data into Linked Data in the next section.

<sup>2</sup><http://www.loc.gov/standards/vracore/schemas.html>

<sup>3</sup><http://dublincore.org/>

<sup>4</sup><http://labs.europeana.eu/apps/SearchEngineEuropeana>

<sup>5</sup><http://www.openarchives.org/ore/>

<sup>1</sup><https://www.rijksmuseum.nl/en/api>

#### 4. Data conversion

To create Linked Data, a conversion needs to take place from the data contained in the collection management system into RDF. As of March 2016, the collection management system includes 597,193 registered objects which can be described using 597 available fields. Multiple steps are taken to select and convert a subset of fields and objects, which we will describe in the remainder of this section.

Data from the collection management is harvested daily and loaded into a database which serves the website. Not all of the 597 available metadata fields are included in the output of the collection management system, a subset of 245 fields is specified in a dedicated file<sup>6</sup>. Fields that are no longer used or contain sensitive data such as insurance values are excluded. The selected fields are transformed to form field names which better reflect their content, omit empty values and generate links to other databases maintained by the Rijksmuseum. This conversion is accomplished using an Extensible Stylesheet Language Transformations (XSLT) file<sup>7</sup>.

On top of the database runs an API, which is used for outputting RDF. Not all of the 597,193 registered collection objects are included in the output, a subset is selected based on copyright statements and the ownership of the object. This results in a set of 351,814 objects which are under management of the museum and are free of rights. Whether a collection object is under management of the museum is loosely defined, it includes objects owned by the museum, the state and the city of Amsterdam, but also objects which are on permanent loan. Objects which are on loan for a period shorter than six months are not considered.

Selected collection objects are converted into RDF with a second XSLT file<sup>8</sup>. Every relevant metadata field of a collection object is mapped to a property in the Europeana Data Model that most closely resembles the values of the field. Since many of these properties originate from the Dublin Core Metadata Initiative, they often describe the data in more generic terms as the original field, causing some loss of precision. We

describe the resulting data model for the Rijksmuseum collection in the next section.

For some fields only textual values are available, others are described using contextual concepts. These concepts are manually added in the collection management system by employees documenting the collection objects. Employees can select concepts from a combination of Rijksmuseum thesauri and external datasets and all concepts have a unique identifier. If for selected fields such an identifier is encountered during the conversion, a reference to the resource is added as well as text in the form of the label of the resource.

The output of the API is used to obtain a complete harvest of the data, which is in turn loaded into a graph database, also known as a triple store. These harvests are run on a monthly basis by an employee of the museum, who updates the triple store by loading the latest version and who provides links to downloads of older datadumps, which are versioned according to the year and month they were obtained. The file `201603-rma-edm-collection.ttl.gz` is used to obtain statistics in Section 6.

#### 5. Data model and URIs

The Linked Data version of the Rijksmuseum collection is modelled according to the Europeana Data Model (EDM). EDM reuses elements from existing models such as Dublin Core. The structure of the model is expressed with RDF Schema, using constructs like subclass and subproperty relations. The Web Ontology Language is used to relate EDM elements to other data models.

The data model makes a distinction between a collection item and its digital representation(s). This is achieved with three core classes: *edm:ProvidedCHO* for cultural heritage objects, *edm:WebResource* for web resources and *ore:Aggregation* for aggregations of resources. Figure 1 shows the metadata of a Rembrandt painting and its core and contextual classes.

An *ore:Aggregation* is used to connect the metadata of a cultural heritage object to web resources. Every collection item in the collection management system gets an aggregation object with its persistent identifier as URI. Information can be added to the *ore:Aggregation*, Figure 1 for example shows that the Rijksmuseum served as data provider.

Every *ore:Aggregation* is connected to a resource with class *edm:ProvidedCHO*, representing a description of the physical cultural heritage object. Figure 1

<sup>6</sup>[https://github.com/Rijksmuseum/conversion\\_adlib](https://github.com/Rijksmuseum/conversion_adlib) includes file `adlibweb.xml` which identifies the metadata fields that are included.

<sup>7</sup>[https://github.com/Rijksmuseum/conversion\\_adlib](https://github.com/Rijksmuseum/conversion_adlib) includes file `rijksstudio.xslt` which transforms the data.

<sup>8</sup>[https://github.com/Rijksmuseum/conversion\\_oai\\_formats](https://github.com/Rijksmuseum/conversion_oai_formats) includes file `europaana_edm.xslt` which provides an overview of the mappings.

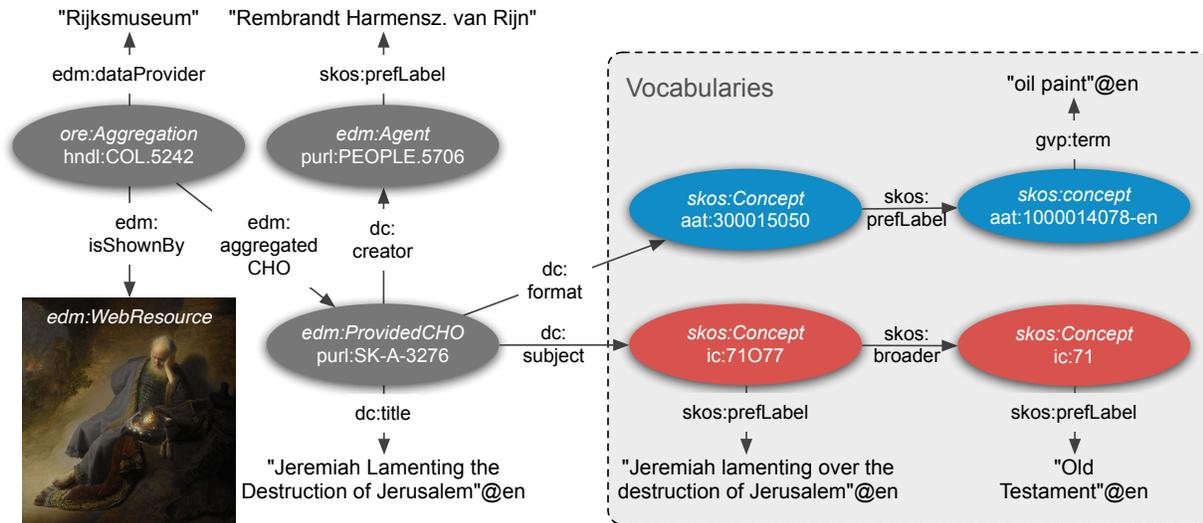


Fig. 1. Example of the painting "Jeremiah Lamenting the Destruction of Jerusalem" modelled according to the EDM data model.

shows four of the properties used to describe objects in the Rijksmuseum dataset: *dc:creator*, *dc:title*, *dc:format* and *dc:subject*. When possible, concepts are used to describe aspects of the artwork, such as the thesaurus term *purl:PEOPLE.5706* for Rembrandt and the concept *aat:300015050* for oil paint. Section 6 lists the occurrences of predicates used to describe objects in the Rijksmuseum dataset.

When a digital representation is available, the aggregation points to the URL where the image can be obtained. This URL is of type *edm:WebResource* and can in turn be described with metadata, adding for example information about its creator. Note that the creator of the image most often differs from the creator of the artwork. The Rijksmuseum dataset currently includes information about the date of creation and the file format of the image.

Not all subtleties of the collection data can be captured by using constructs included in the Europeana Data Model description. While the original data includes detailed information about creator roles and fields like 'rejected creator', no such properties exist in Dublin Core. EDM allows for refining and extending the data model, so in the future the museum can choose to introduce its own more specific constructs or find others to reuse. This could increase the coverage of data in the collection management system included in the Linked Data version.

Persistent identifiers in the form of handles<sup>9</sup> are used for the URIs of the *ore:Aggregation*. Since an aggregation connects metadata of the artwork and its digital representation, the persistent identifier is not related to the object number of the artwork. The URI of the cultural heritage object descriptions is based on the purl scheme<sup>10</sup> and consist of five elements: purl prefix, dataset type, country code, organisation, and object number. This results in the following URI for the *edm:ProvidedCHO* resource of the Rembrandt in Figure 1: <http://purl.org/collections/nl/rma/SK-A-3276>. When values refer to one of the thesaurus databases of the museum, a URI is generated based on the internal reference used, linking the collection object with the thesaurus.

## 6. Rijksmuseum dataset statistics

As of March 2016 the Linked Data version of the Rijksmuseum collection<sup>11</sup> comprises 22,846,996 triples, describing 351,814 objects, of which 207,441 have a graphical depiction. Metadata about the collection is made available using the Vocabulary of a Friend (VOAF)<sup>12</sup>. Ten sub-collections are maintained, including sculptures (29,782 objects), historical items (19,936 objects), paintings (3,949 objects) and Asian

<sup>9</sup><http://www.handle.net/>

<sup>10</sup><https://purl.org/>

<sup>11</sup><http://datahub.io/dataset/rijksmuseum>

<sup>12</sup><http://purl.org/vocommons/voaf>

art (3,722 objects). The print collection has 280,047 objects and is by far the largest sub-collection, it includes prints, drawings and photos.

Table 1 lists the predicates used to describe collection items. A title is provided for almost all artworks, of which the majority is unique. Although most of the titles are in Dutch, some of them are also available in English. Over half of the objects have the predicate *dc:description*, which includes textual information about the subject matter and art-historical background of the object. For example, the description of Figure 1 includes the following text: “Downcast, the biblical prophet Jeremiah leans his tired head on his hand” and “Rembrandt used powerful contrasts of light and shadow to heighten the drama of the scene”.

There are over thirty thousand unique creators, which are mostly described using resources from the person database of the museum. Half of the *dc:creator* literals are based on labels from resources in the person database, while the other half is used for adding nuances to the creator field which are difficult to capture in an resource of type *edm:Agent*. This includes textual descriptions such as "Anonymous", "possibly Rembrandt" and "follower of Rembrandt". The predicate *dc:contributor* refers to names of additional persons involved in the creation process.

The *dc:subject* predicate provides information about the subject matter, where resources from both the Iconclass vocabulary as well as the Rijksmuseum thesaurus are used. Subjects are also described using Dutch literals, since not all subject matter falls within the scope of the available vocabularies. The predicate *dcterms:spatial* refers to places, for which both terms from the thesaurus as well as language agnostic literals are used.

The museum describes temporal aspects using the predicates *dc:coverage* for periods and *dcterms:created* for creation dates. Dutch as well as English literals are used for both predicates, where creation dates are expressed using a year or estimated years (e.g. "1630" or "c.1600-c.1625") and periods are expressed using textual descriptions (e.g. "second quarter 17th century" or "18th century").

The predicate *dc:type* uses a mixture of concepts from the Art & Architecture thesaurus, the museum's thesaurus and literals to identify the type of artwork (e.g. print or painting). The same applies to *dc:format*, which is used to specify materials such as the resource *aat:300015050*, which stands for "oil paint". In the next section we describe in more detail how many of such connections are made to external datasets. Phys-

ical dimensions of the work are recorded using *dcterms:extent*, specifying the height and width of objects in centimeters. The painting in Figure 1 has for example "height 58 cm" and "width 46 cm".

Objects can be connected to each other with the *dcterms:hasPart* and *dcterms:isPartOf* predicates. These relations are for example used to link photographs to their album. Related sources (often books) are linked to the object using *dcterms:isReferencedBy*. These three predicates currently refer to literals representing identifiers, which in a later stage can be converted to resources matching the objects indicated by the identifiers. Every object has two identifiers, one for internal use (e.g. "SK-A-3276") and one persistent identifier in the form of a handle (e.g. "hdl:RM0001.COLLECT.5242").

The predicate *dc:provenance* encodes the provenance in a literal enumerating the present and past owners of the artwork. Most of the intellectual property rights are part of the public domain, while sometimes specific persons are specified who own the copyrights. Europeana requires some values to be present and limited to a set range. The publisher is always the "Rijksmuseum", while *dc:language* is the language code of the country of the institution, in this case "nl". The *edm:type* is "IMAGE".

As outlined in Section 4, some of the literals are based on the labels of resources. Although adding both the literal and resource introduces redundant information, this can support applications that do not handle the added complexity of resources well. The literals from *dc:type*, *dc:format* and *dcterms:spatial* all directly originate from the museum's thesaurus. 77 percent of the literals of the *dc:subject* field come from either resources contained in thesaurus or the person database. The remainder of the subject literal values mainly describe specific dates and periods such as "1701 - 1703". We describe the resources contained in the thesauri and links to external datasets in the next section.

## 7. Contextual concepts and links to external datasets

Institutions often maintain their own vocabularies containing their perspective on contextual concepts. When the contextual concepts of collection items are replaced with objects from standardised vocabularies such as the Getty vocabularies, these nuances in perspectives are in danger of disappearing. So while col-

Table 1  
Overview of the predicates that describe collection items

predicate	distinct artworks	distinct resources	vocabularies	distinct literals	en literals	nl literals
dc:contributor	89,796	0	-	9,146	0	9,146
dc:coverage	138,141	0	-	212	106	106
dcterms:created	340,865	0	-	46,255	19,837	19,837
dc:creator	349,787	27,904	rma	38,851	97	38,754
dc:description	217,202	0	-	176,786	2,487	174,299
dcterms:extent	288,318	0	-	55,241	15,229	40,012
dc:format	322,152	593	aat, rma	924	322	602
dcterms:hasPart	487	0	-	13,646	0	0
dc:identifier	351,814	0	-	703,429	0	0
dcterms:isPartOf	59,197	0	-	2,754	0	0
dcterms:isReferencedBy	98,815	0	-	80,770	0	0
dc:language	351,814	0	-	1	0	0
dcterms:provenance	13,239	0	-	1,155	0	1,155
dc:publisher	351,814	0	-	1	0	0
dcterms:spatial	214,752	3,339	rma	3,361	0	0
dc:subject	221,868	44,840	ic, rma	32,452	0	32,452
dc:title	351,789	0	-	297,271	6,784	290,487
dc:type	351,749	3,541	aat, rma	3,700	155	3,545
edm:type	351,814	0	-	1	0	0

Table 2  
Types in the Rijksmuseum thesaurus with more than 500 values

type	distinct resources	nl labels	en labels
person	38,939	27,904	27,904
place	17,174	17,174	152
object name	6,074	6,074	298
keyword	5,021	5,021	166
event	1,982	1,982	43
technique	1,401	1,401	73
occupation	1,044	1,044	15
material	882	882	428
location RMA	808	808	5

lection objects and contextual concepts in the thesauri of the Rijksmuseum are linked to an increasing number of available datasets maintained by other institutions, the Rijksmuseum chooses to also maintain and use its own. This allows the museum to preserve its own perspective and in a later stage vocabulary alignment tools can be used to match the concepts with similar concepts in external datasets [15].

Five contextual classes are defined in the Europeana Data Model for relating collection items to contextual information: *edm:Agent*, *edm:Place*, *edm:TimeSpan*, and *skos:Concept*. These classes correspond to the types of thesaurus records in the databases of the Rijksmuseum: the person database maps to the agent

class and the general thesaurus database contains information about places, historical events, and other concepts. However, the type of concepts in the museum's thesaurus are divided into finer grained types. An overview of the type of terms is presented in Table 2 along with the number of available resources and labels.

The thesaurus forms a hierarchy of terms using relations such as broader and narrower, which are represented using SKOS. Figure 2a shows two object terms, where the type of term is indicated using the *rma:term\_type* predicate. All of the 33,800 concepts in the thesaurus have a Dutch label, only 1,539 have an English label. For 3,254 terms a *skos:scopeNote* is available, describing the appropriate use of the term. Every term has its own unique *skos:externalID* and the last modification date of the term is recorded using *rma:modification*.

The person database contains concepts of type *edm:Agent*, Figure 2b shows "Rembrandt" as an example. The names of persons are indicated using *skos:prefLabel* and every person has a name, either represented as a Dutch, English, or language independent literal. Additional information about persons is added using the the Resource Description and Access

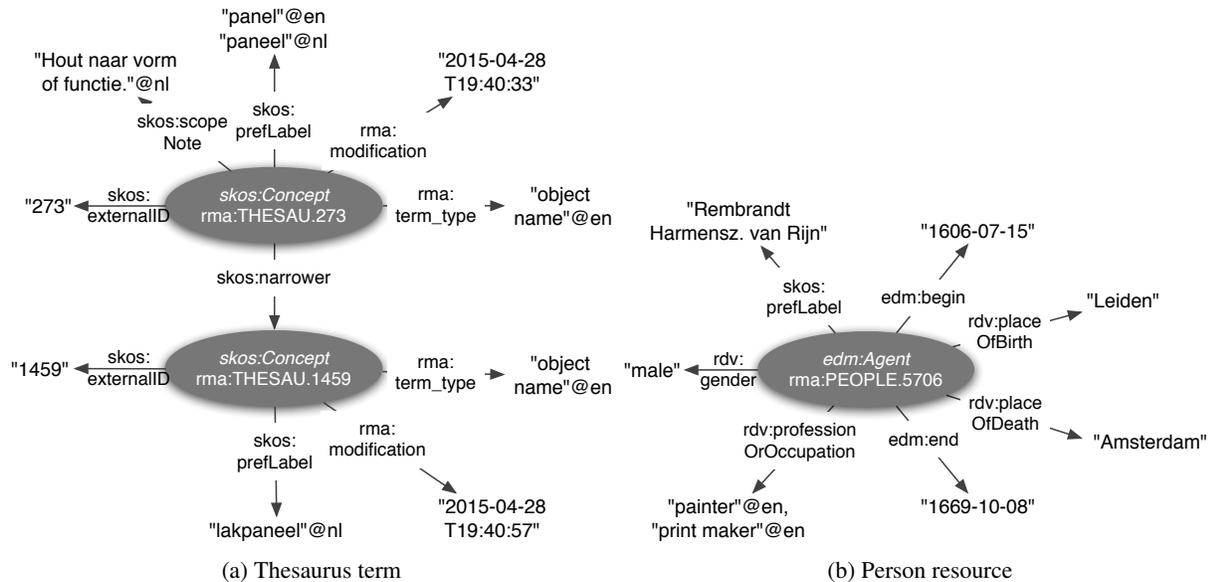


Fig. 2. Diagrams of contextual concepts representing "panel" thesaurus terms and the person "Rembrandt".

(RDA)<sup>13</sup> vocabulary. A `rdv:professionOrOccupation` is provided when known, Rembrandt has for example multiple listed profession. Besides being a painter he also made prints. When available, information is added about the birth and death of the person. In the remainder of this section we describe how external datasets are used to extend the thesauri and annotate the collection data of the Rijksmuseum.

The Art & Architecture Thesaurus<sup>14</sup> (AAT) consists of concepts about arts from antiquity to the present. Concepts include art styles, materials and agents. It is maintained by the Getty foundation, which released a Linked Data version in February 2014 with 38,619 concepts. The focus of the thesaurus lies on generic concepts: instead of for example describing individual artists, it includes the concept "printmakers". New concepts originate from cataloging and documentation projects and labels of concepts are available in multiple languages.

The Rijksmuseum uses the Art & Architecture Thesaurus for the `dc:type` and `dc:format` metadata fields. A small subset of the available concepts is used: 305 distinct formats and 124 distinct types. As can be seen in the type frequency distribution in Figure 3a, a small number of concepts is often used. This is also the case for the format field. For example, the top three types

are prints (183,916), stereoscopic photographs (3,480) and plates (1,617). The museum refrains from assigning art styles to objects, since it is often debatable to which art style an object belongs.

The Iconclass vocabulary<sup>15</sup> contains 39,578 concepts, providing 'a systematic overview of subjects, themes and motifs in Western art'. An official Linked Data version was released in 2012. Concepts are identified with codes and SKOS relations are used to create a hierarchy between them. Labels of concepts are available in English, German, French, Finnish and Italian. An example of a code used in Iconclass is 7, which refers to the "Bible" and is connected to the concept 7107, "the book of Jeremiah", using `skos:narrower` predicates. Context dependent modifiers can be added to the codes: for 71C131(+3), the code 71C131 indicates "the sacrifice of Isaac", while the modifier (+3) indicates that one or more angels are depicted on the artwork.

The museum uses the Iconclass vocabulary to describe subject matter. Iconclass codes are added by catalogers during the registration process described in Section 2. Out of the 39,578 concepts in the vocabulary, 10,434 are used to add information to an object. Of the 351,814 collection objects, 172,059 have one or more Iconclass annotations. As Figure 3b shows, many

<sup>13</sup><http://rdvocab.info>

<sup>14</sup><http://www.getty.edu/research/tools/vocabularies/aat/>

<sup>15</sup><http://www.iconclass.nl/>

of the concepts are often used, while on average a code is used 27 times.

The Short-Title catalogue Netherlands (STCN) is ‘the retrospective national bibliography of the Netherlands in the period 1540-1800’<sup>16</sup>, maintained by the National Library of the Netherlands. A Linked Data version is available, containing records of 196,396 publications. This dataset contains many books that are the source of objects in the print collection of the Rijksmuseum and linking the two collections provides valuable contextual information.

The catalogers of the Rijksmuseum add references to the National Library by adding textual descriptions of the books in a notes field. To create links these descriptions are scanned for objects from the STCN that match the title, publication date and publisher. This automated matching process resulted in 3598 links from the Rijksmuseum collection to 501 publications in the STCN catalogue. The links are encoded as *dc:hasPart* relations from the STCN vocabulary to the Rijksmuseum collection. Catalogers estimate that roughly 14,000 works can potentially be linked to STCN. A more rigorous way of referring to STCN titles is in the making to support this.

## 8. Data usage

Uses of the Linked Data of the Rijksmuseum include search, recommendation, collection integration and browsing. In this section we give an overview of how the museum data has been used in various research projects and provide statistics about the Rijksmuseum API. Most projects that contributed to the process of data development had demonstrators illustrating the power of Linked Data. The *MultimediaN E-Culture* project showcased a semantic search system, which won the 1st prize in the 2006 International Semantic Web Conference Challenge [13]. It clustered search results based on the graph path leading from matching literal to artwork. The dataset was extended from 750 artworks to the entire Rijksmuseum collection in a search prototype of the *Europeana Thought Lab*<sup>4</sup>, showing advanced search functionality to be included in the portal at a later stage.

Other ways of accessing data were introduced in subsequent years. The *CHIP* demonstrator recommended artworks based on graph patterns [17]. The

*STITCH* project took a different approach with facets based on Iconclass concepts, allowing the user to browse the collection based on different topics [6]. The *Agora* demo provided access to the collection with an emphasis on the events related to objects [16]. The *Accurator* crowdsourcing tool of the *SEALINCMedia* project uses graph patterns to recommend people artworks to which they can contribute information, gathering more accurate subject matter descriptions [2].

The Rijksmuseum maintains an API<sup>1</sup> for application developers, optionally returning data formatted according to the Europeana Data Model. 587 people have registered for access to the API as of August 2015 and many different applications have been build on top of it<sup>17</sup>. The API is used by Europeana to harvest collection data, making all Rijksmuseums structured data available through the Europeana portal. Europeana logs the page views of this portal and during a period of 20 weeks (starting from the 1st of May 2015) Rijksmuseum collection objects got 42,156 page views of which 34,206 were unique.

## 9. Discussion

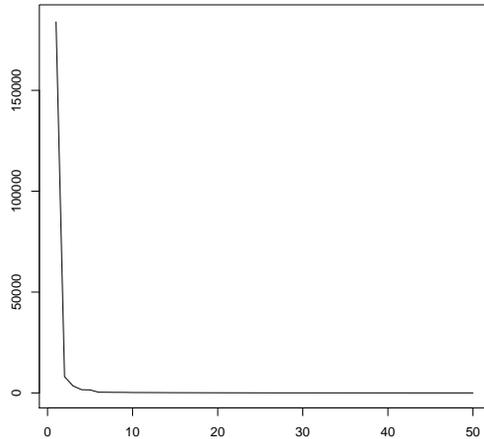
For decades, Linked Data has been a promise for data publication and integration in the cultural heritage sector. Despite widespread interest and apparent advantages, only a limited number of institutions have managed to make their collection available as Linked Data<sup>18</sup>. After a period of development influenced by many research projects, the Rijksmuseum is one of them. Furthermore, the museum is in control of the entire publication process of its own collection as Linked Data.

The majority of the Rijksmuseum collection items are part of the public domain since their intellectual property rights have expired. Although general understanding is that digitised representations of public domain works should again be released under the same license terms, many institutions are hesitant to do so, in fear of losing a possible revenue stream. The Rijksmuseum did release their high-quality images in the public domain in 2013, arguing that the increase in attention and exposure would result in a higher number

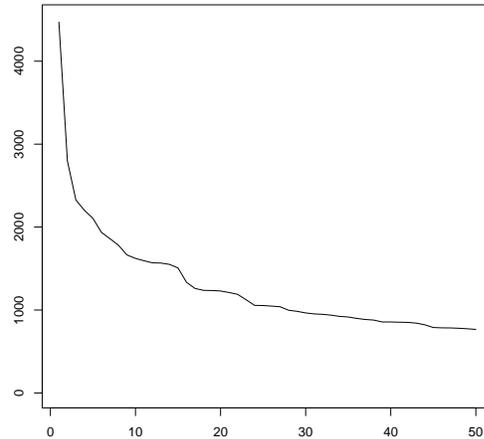
<sup>16</sup><http://www.kb.nl/expertise/voor-bibliotheken/short-title-catalogue-netherlands> (accessed on 04-07-2014)

<sup>17</sup><http://www.opencultuurdata.nl/category/apps/> provides an overview of applications that use cultural heritage data, including applications that are build on top of the Rijksmuseum API.

<sup>18</sup>As of March 2016 <http://datahub.io> lists 7 cultural heritage datasets formatted in RDF of which 3 provide a SPARQL endpoint.



(a) AAT type concept frequency distribution



(b) Iconclass subject concept frequency distribution

Fig. 3. Frequency distributions of the top 50 concepts of AAT and Iconclass that the collection objects are linked to.

of visitors [12]. In turn it allowed the museum to gain more control over the digital representations that had appeared online, replacing many inferior versions by its high-quality images.

The quality and correctness of metadata is of paramount importance to museums [14]. The Rijksmuseum has an extensive quality control process in place to ensure the correctness of metadata. By adding a direct conversion layer to the collection management system it ensures that the same level of quality is translated to the Linked Data version. All the criteria for five star Linked Data as defined in [11] are met. There is a description of the data online<sup>19</sup>, the data is available in RDF, there are many links to structured vocabularies and metadata about the collection is made available. Furthermore, concepts in the Linked Data version of the Smithsonian American Art Museum are linked to the thesauri of the Rijksmuseum [14].

Data aggregators such as Europeana enticed many institutions to provide digital versions of their collection, often relying on external expertise for the conversion process. This led to an increase of available collections, although providing access to data through aggregators has the major drawback that it creates a gap between the institution and its data [1]. We believe it is therefore still desirable that institutions publish their own data, if the required expertise is available. They

thereby remain in control of choosing the most suitable data model, URI naming schemes, links to other datasets, and update processes.

The data of the Rijksmuseum is subject to constant change: newly digitised objects are added on a daily basis and employees extend and refine information regularly. The Linked Data version places the artworks in a broader context, allowing others to benefit from the progress made through easy reuse and the possibility to add new perspectives to the data.

*Acknowledgements* This publication was supported by the Dutch national program COMMIT/.

## References

- [1] Viktor de Boer, Jan Wielemaker, Judith van Gent, Michiel Hildebrand, Antoine Isaac, Jacco van Ossenbruggen, and Guus Schreiber. Supporting Linked Data production for cultural heritage institutes: The Amsterdam Museum case study. In Elena Simperl, Philipp Cimiano, Axel Polleres, Óscar Corcho, and Valentina Presutti, editors, *The Semantic Web: Research and Applications - 9th Extended Semantic Web Conference, ESWC 2012, Heraklion, Crete, Greece, May 27-31, 2012. Proceedings*, volume 7295 of *Lecture Notes in Computer Science*, pages 733–747. Springer, 2012. 10.1007/978-3-642-30284-8\_56.
- [2] Chris Dijkshoorn, Mieke H. R. Leysen, Archana Nottamkandath, Jasper Oosterman, Myriam C. Traub, Lora Aroyo, Alessandro Bozzon, Wan Fokkink, Geert-Jan Houben, Henrike Hovelmann, Lizzy Jongma, Jacco van Ossenbruggen, Guus

<sup>19</sup><http://datahub.io/dataset/rijksmuseum>

- Schreiber, and Jan Wielemaker. Personalized nichesourcing: Acquisition of qualitative annotations from niche communities. In Shlomo Berkovsky, Eelco Herder, Pasquale Lops, and Olga C. Santos, editors, *Late-Breaking Results, Project Papers and Workshop Proceedings of the 21st Conference on User Modeling, Adaptation, and Personalization., Rome, Italy, June 10-14, 2013*, volume 997 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2013. URL [http://ceur-ws.org/Vol-997/patch2013\\_paper\\_13.pdf](http://ceur-ws.org/Vol-997/patch2013_paper_13.pdf).
- [3] Martin Doerr and Nicholas Crofts. Electronic Esperanto: The role of the object oriented CIDOC reference model. In David Bearman and Jennifer Trant, editors, *Cultural Heritage Informatics: Selected Papers from ICHIM99. Washington, DC, USA, September 22-26*, pages 157–173. Archives & Museum Informatics, 1999.
- [4] Martin Doerr, Stefan Gradmann, Steffen Hennicke, Antoine Isaac, Carlo Meghini, and Herbert van de Sompel. The Europeana Data Model (EDM). In *World Library and Information Congress: 76th IFLA general conference and assembly*, pages 10–15, 2010.
- [5] Joost Geurts, Stefano Bocconi, Jacco van Ossenbruggen, and Lynda Hardman. Towards ontology-driven discourse: From semantic graphs to multimedia presentations. In Dieter Fensel, Katia P. Sycara, and John Mylopoulos, editors, *The Semantic Web - ISWC 2003, Second International Semantic Web Conference, Sanibel Island, FL, USA, October 20-23, 2003, Proceedings*, volume 2870 of *Lecture Notes in Computer Science*, pages 597–612. Springer, 2003. 10.1007/978-3-540-39718-2\_38.
- [6] Julio Gonzalo, Costantino Thanos, M. Felisa Verdejo, and Rafael C. Carrasco, editors. *Semantic Web Techniques for Multiple Views on Heterogeneous Collections: A Case Study*, volume 4172 of *Lecture Notes in Computer Science*, 2006. Springer. 10.1007/11863878\_36.
- [7] Peter Gorgels. Rijksstudio: Make your own masterpiece! In Nancy Proctor and Rich Cherry, editors, *Museums and the Web 2013 (MW2013)*, January 2013. URL <http://mw2013.museumsandtheweb.com/paper/rijksstudio-make-your-own-masterpiece/>.
- [8] Michiel Hildebrand, Jacco van Ossenbruggen, Lynda Hardman, and Geertje Jacobs. Supporting subject matter annotation using heterogeneous thesauri: A user study in web data reuse. *International Journal of Human-Computer Studies*, 67(10):887–902, 2009. 10.1016/j.ijhcs.2009.07.008.
- [9] Seth van Hooland and Ruben Verborgh. *Linked data for libraries, archives and museums*. Facet Publishing, 2014.
- [10] Eero Hyvönen, Eetu Mäkelä, Mirva Salminen, Arttu Valo, Kim Viljanen, Sampsa Saarela, Miiikka Junnila, and Suvi Kettula. MuseumFinland - Finnish museums on the semantic web. *Web Semantics: Science, Services and Agents on the World Wide Web*, 3(2-3):224–241, 2005. 10.1016/j.websem.2005.05.008.
- [11] Krzysztof Janowicz, Pascal Hitzler, Benjamin Adams, Dave Kolas, and Charles Vardeman. Five stars of Linked Data vocabulary use. *Semantic Web*, 5(3):173–176, 2014. 10.3233/SW-140135.
- [12] Joris Pekel. Democratising the Rijksmuseum: Why did the Rijksmuseum make available their highest quality material without restrictions, and what are the results? Case Study Europeana 2014, 2014. URL [http://pro.europeana.eu/files/Europeana\\_Professional/Publications/Democratising%20the%20Rijksmuseum.pdf](http://pro.europeana.eu/files/Europeana_Professional/Publications/Democratising%20the%20Rijksmuseum.pdf).
- [13] Guus Schreiber, Alia K. Amin, Mark van Assem, Viktor de Boer, Lynda Hardman, Michiel Hildebrand, Laura Hollink, Zhisheng Huang, Janneke van Kersen, Marco de Niet, Borys Omelayenko, Jacco van Ossenbruggen, Ronny Siebes, Jos Taekema, Jan Wielemaker, and Bob J. Wielinga. MultimediaN e-culture demonstrator. In Isabel F. Cruz, Stefan Decker, Dean Allemang, Chris Preist, Daniel Schwabe, Peter Mika, Michael Uschold, and Lora Aroyo, editors, *The Semantic Web - ISWC 2006, 5th International Semantic Web Conference, ISWC 2006, Athens, GA, USA, November 5-9, 2006, Proceedings*, volume 4273 of *Lecture Notes in Computer Science*, pages 951–958. Springer, 2006. 10.1007/11926078\_70.
- [14] Pedro A. Szekely, Craig A. Knoblock, Fengyu Yang, Xuming Zhu, Eleanor E. Fink, Rachel Allen, and Georgina Goodlander. Connecting the Smithsonian American Art Museum to the Linked Data cloud. In Philipp Cimiano, Óscar Corcho, Valentina Presutti, Laura Hollink, and Sebastian Rudolph, editors, *The Semantic Web: Semantics and Big Data, 10th International Conference, ESWC 2013, Montpellier, France, May 26-30, 2013. Proceedings*, volume 7882 of *Lecture Notes in Computer Science*, pages 593–607. Springer, 2013. 10.1007/978-3-642-38288-8\_40.
- [15] Anna Tordai, Jacco van Ossenbruggen, Guus Schreiber, and Bob J. Wielinga. Aligning large SKOS-like vocabularies: Two case studies. In Lora Aroyo, Grigoris Antoniou, Eero Hyvönen, Annette ten Teije, Heiner Stuckenschmidt, Liliana Cabral, and Tania Tudorache, editors, *The Semantic Web: Research and Applications, 7th Extended Semantic Web Conference, ESWC 2010, Heraklion, Crete, Greece, May 30 - June 3, 2010, Proceedings, Part I*, volume 6088 of *Lecture Notes in Computer Science*, pages 198–212. Springer, 2010. 10.1007/978-3-642-13486-9\_14.
- [16] Chiel van den Akker, Susan Legêne, Marieke van Erp, Lora Aroyo, Roxane Segers, Lourens van der Meij, Jacco van Ossenbruggen, Guus Schreiber, Bob J. Wielinga, Johan Oomen, and Geertje Jacobs. Digital hermeneutics: Agora and the online understanding of cultural heritage. In David De Roure and Marshall Scott Poole, editors, *Web Science 2011, WebSci '11, Koblenz, Germany - June 15 - 17, 2011*, pages 10:1–10:7. ACM, 2011. 10.1145/2527031.2527039.
- [17] Yiwen Wang, Natalia Stash, Lora Aroyo, Peter Gorgels, Lloyd Rutledge, and Guus Schreiber. Recommendations based on semantically enriched museum collections. *Web Semantics: Science, Services and Agents on the World Wide Web*, 6(4):283–290, 2008. 10.1016/j.websem.2008.09.002.