Int. J. Appl. Math. Comput. Sci., 2019, Vol. 29, No. 1, 7–16
DOI: 10.2478/amcs-2019-0001

INTRODUCING NARRATIVES IN EUROPEANA: A CASE STUDY

CARLO MEGHINI [a,*], VALENTINA BARTALESI [a], DANIELE METILLI [a], FILIPPO BENEDETTI [a]

[a] Networked Multimedia Information Systems Laboratory
Institute of Information Science and Technologies "Alessandro Faedo"—CNR
Via Giuseppe Moruzzi 1, 56124 Pisa, Italy
e-mail: {carlo.meghini,valentina.bartalesi}@isti.cnr.it,
{daniele.metilli,filippo.benedetti}@isti.cnr.it

We present a preliminary study to introduce narratives as a first-class functionality in digital libraries. The general idea is to enrich those libraries with semantic networks of events providing a meaningful contextualisation of the digital libraries' objects. More specific motivations are presented through a set of use cases by different actors who would benefit from using narratives for different purposes. Then, we consider a specific digital library, Europeana, the largest European digital library in the cultural heritage domain. We discuss how the Europeana Data Model could be extended for representing narratives, and we introduce an ontology for narratives. We also present a semi-automatic tool, which, on the basis of the ontology, supports the creation and visualisation of narratives, and we show how the tool has been employed to create a narrative of the life of the painter Gustav Klimt as a case study. In particular, we focus our attention on the functionality of the tool that allows extracting and proposing to the user specific digital objects for each event of the narrative.

Keywords: digital libraries, narratives, Europeana, ontologies.

1. Introduction

Europeana[1] is the largest European digital library (DL), containing descriptions of about 54 million cultural heritage objects in various formats such as images, texts, audio, movie files, and 3D objects. It aggregates metadata from more than 3,500 European institutions such as libraries, archives, galleries, and museums. Europeana has several different kinds of users, from scholars and researchers to professionals and general users. The DL provides a single access point to metadata descriptions of European cultural heritage material. The access functionalities of the current DLs are largely influenced by Web search engines, whereby users express their information needs in the form of natural language queries and the DL returns a ranked list of relevant objects. However, DLs do not report any semantic relations between objects, i.e., relations connecting objects in a way that reflects their meaning and context. Europeana is no exception.

In order to enrich Europeana with semantic relations, in the Europeana Data Model (EDM)[2] the concept of event was introduced. Indeed, it is well known (Ruotsalo and Hyvönen, 2007; Shaw *et al.*, 2009; Van Hage *et al.*, 2011) that events play a major role in the contextualisation of objects, by providing an account of happenings concerning the objects, in terms of people, physical objects, places, and times that are involved in such events. Moreover, events are linked to one another to form stories. However, although events are included in the EDM, they do not appear in the Europeana database since they do not show up in the catalogues of libraries, archives, museums, and galleries where the descriptions collected by Europeana come from. Indeed, events are typically found in historical documents. The automatic capturing of events and their characteristics from such documents is an unsolved problem given that it is difficult for machines to grasp the meaning of the languages used for inter-human communication.

Our current research aims at overcoming the limitations of the search functionality of the current DLs by introducing the *narrative* as a first-class concept in the

*Corresponding author

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

[2] http://pro.europeana.eu/page/edm-documentation.

data model of such DLs. The vision is that a user wishing to know what Europeana has about the Austrian painter Gustav Klimt would obtain in response not only the ranked list of objects concerning Klimt that the DL knows about, but also a narrative about Klimt, that is, a semantic network linking such objects in a story that would work as a contextualisation of the objects themselves, and as such would provide the user with a larger and more significant amount of information.

Narrative is a concept studied in several fields, from literary studies to cognitive science. As a matter of fact, "narrative can be viewed under several profiles—as a cognitive structure or way of making sense of experience, as a type of text, and as a resource for communicative interaction" (Herman, 2011). In our research, we perceive a narrative as a semantic network, meaningful to the user, consisting of events related to one another, to the entities that compose the events (e.g., agents, places, time, physical objects) and to the digital objects through semantic relations.

As a first step towards the introduction of narratives in the search functionality of DLs, this paper reports an experiment we performed on Europeana that is an extension of the study presented by Meghini *et al.* (2017). Using a semi-automatic tool to build and visualise narratives (NBVT) we developed, we constructed a narrative about the life of the Austrian artist Gustav Klimt.[3] The narrative can be visualised in the form of a timeline, graphs, or tables. The NBVT is based on an ontology for narratives we developed as an extension of the CIDOC CRM (CRM for short) standard vocabulary (Doerr, 2003). The tool uses Wikidata[4] as an external knowledge base of entities and imports images from Wikimedia Commons.[5] The users of NBVT can populate the ontology either by choosing entities automatically imported from Wikidata, or by defining new entities. The NBVT assigns URIs to the selected entities, makes it possible to link entities to related events, and events to one another through semantic relations. The tool also allows events to be connected to related digital objects stored in external DLs. In addition, users can also provide primary and secondary sources for the events of the narrative. The knowledge acquired by the tool is exported as Linked Data.

The narrative of Klimt's life could be imported in Europeana and shown as a search result, thereby placing the objects related to Klimt in the more general context of the biography (or a part thereof) of their creator. This would enrich the search functionalities of the DL by providing semantic links between its digital objects.

The paper is structured as follows. Section 2 presents some use cases providing further evidence of the usefulness of narratives as well as input to the future technical development. Section 3 reports some related projects that developed tools to enrich and explore the Europeana digital collections. Section 4 describes how the Europeana Data Model needs to be extended in order to represent narratives and reports a description of an ontology for representing narratives. In Section 5 we present a semi-automatic tool for building and visualising the narratives we built. Section 6 describes the functionality we developed in order to automatically enrich the events of a narrative with the Europeana digital objects. In Section 7 we describe how we built a narrative on the life of the Austrian painter Gustav Klimt and how it could be imported in Europeana. Finally, in Section 8 we report our final remarks.

2. Identifying users for narratives

The Europeana portal serves as a single access point to the metadata of the digital objects collected in the DL (Petras *et al.*, 2017). Each object page reports a description of the object, a thumbnail preview, the link to the owner institution and information about the legal terms of the reuse of the object. Furthermore, users can explore the DL via curated collections that aggregate objects on the basis of a specific topic, e.g., Natural History or Art. Various studies have been conducted in order to identify the users of Europeana (Europeana, 2014; Clough *et al.*, 2017). There are several different kinds of users, from scholars to cultural heritage professionals (such as curators, librarians, archivists) to teachers and students and, finally, general users, e.g., casual tourists. In this section, we consolidate our motivations by presenting some use cases of how narratives could improve the use of Europeana by the kinds of users reported below. With the NBVT, these users can create and visualise their own narratives or access the narratives built by somebody else.

Scholar. Scholars, such as historians or biographers, could create and access narratives about the life and the works of the authors they study. They may provide their own texts from which the narrative was extracted, and they may also be interested in expressing the primary sources supporting the narrative. To this end, the construction of a narrative by a historian or biographer can be viewed as an inferential process, using evidence collected from sources to infer propositions that have been narrated in a text. Different scholars can create different narratives on the same topic.

Exhibition or museum curator. A narration timeline could be used during a monographic exhibition in order to associate the works of an artist to her/his biography and help the visitors to better understand the life and works of the artist. A timeline could also be used in a museum context in order to show to the visitors the history of the

[3] https://dlnarratives.eu/timeline/klimt.html.
[4] https://wikidata.org.
[5] https://commons.wikimedia.org.

museum and the acquisition of the main artworks. In the interface it could be useful to have a functionality that allows visualising only those artworks that are part of the museum's collections. Both for exhibitions and museums, the curator who builds a narrative should have the possibility to add her/his own new images of the objects kept in the museum. Furthermore, information about the location of the objects in the museum rooms could be added through a specific field of the tool.

Digital curator. A digital curator would be able to create narratives for the DL objects she/he would like to promote. For example, a narrative about the Yalta conference may be defined by a digital librarian linking the digital objects (e.g., photographs of the people, the final declaration, etc.) to the event representing the conference. Such an event may further be divided into sub-events which will have to be properly placed on a temporal axis. Curators may also be interested in representing the provenance of the digital objects.

High school professor. The NBVT could be used by a professor as a learning tool. The professor may create a narrative on a topic of study and show it to the students through a timeline visualisation. For each created event, the professor should add her/his description of it using a specific field of the tool. She/he could also insert fragments from a text book, which are taken into account as secondary sources. The professor could also add some primary sources for the main events, in order to examine them in depth. Using the tool, the professor can share her/his timeline with the students. The professor could also enrich the narratives with Wikimedia Commons images or links to related digital objects included in DLs, in order to make the narrative more attractive for the students.

High school student. The NBVT could be used to verify students' comprehension of history or literature taught by a professor, who could ask them to create a narrative on a particular topic using the tool. The students could enrich the narration by adding appropriate entities and providing textual descriptions of each event. Alternatively, a quiz could be proposed to the students on the narrative created by the professor.

3. Related projects

In the past years, several projects have used Europeana data. They have proposed some tools to explore collections of Europeana digital objects. For example, EAGLE (Mannocci *et al.*, 2014) is a European project (2013–2016), which was part of Europeana. The aim of the project was to provide a world reference portal for the epigraphic images and texts of the ancient world. EAGLE enriched Europeana with many digital objects representing epigraphs and developed a story-telling application which allows teachers and experts to assemble epigraphy-based narratives.

Another example is the PATHS project (Fernie *et al.*, 2012), which provides a system that allows users to create, edit, publish, and share pathways or narratives starting from digital objects contained in Europeana and the Alinari Archives. Also the GLAM Wiki Toolset[6] uses Europeana objects. This is a set of tools developed through a collaboration between Europeana and Wikimedia. These tools allow GLAMs (galleries, libraries, archives and museums) to upload their content to Wikimedia Commons, meaning that their collection items can be featured in Wikipedia.

Historiana[7] is an on-line educational multimedia platform that offers historical content, learning activities, and digital tools for exploring events from European history. To enrich its contents, Historiana uses digital objects imported from Europeana. Another example is the Wikidata knowledge base, which links its entities to related digital objects from Europeana[8]. When Europeana launched its Art History collections, Wikidata organised a challenge to enrich its entities with metadata related to the objects contained in these collections.

Our research aims at enriching Europeana with narratives in which each event can be associated with one or more digital object(s) included in the DL. Different from the PATHS project, which defines events linked to each other by semantic similarity relations, the events included in our narratives are linked via richer semantic relations (i.e., temporal, causal and mereological relations), providing a more significant semantic network. Furthermore, as in the EAGLE project, the narrator can add new digital objects to Europeana in order to use these in her/his own narration. Finally, adding digital objects to Europeana automatically enriches the Wikidata knowledge base, since Wikidata links its entities to the Europeana objects. In this way, these new objects can be used by all the projects based on the Wikidata knowledge.

4. From the EDM to an ontology for narratives

The Europeana Data Model (EDM) (Doerr *et al.*, 2010) is an ontology for representing data according to the practices of the various organisations that provide cultural heritage material to Europeana. The EDM proposes two different approaches for descriptive metadata: "object-centric" and "event-centric." The former approach focuses on the descriptive metadata, e.g., a title or creator, which are linked to the related digital object. In the latter approach, the relations between

[6] https://commons.wikimedia.org/wiki/Commons:GLAMwiki_Toolset_Project.

[7] http://historiana.eu.

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

objects are defined by the use of the events in which objects have been involved. In this case, the metadata are related to each event. The idea is that, using this approach, it is possible to create richer networks of entities by representing the events that constitute an object's history. The event-centric approach underlies models such as the CRM (Doerr, 2003) and may suit the data of some (but of course not all) Europeana providers. An event is represented in the EDM as an instance of the class edm:Event. The relations connecting events with the resources that characterise the events are represented in the EDM using three properties: (i) edm:happenedAt, which links an event to the place of its occurrence; (ii) edm:occurredAt, which links an event to the time span of its occurrence; and (iii) edm:wasPresentAt, holding between any resource and an event it is involved in.

The concept of event is a core element of the narratives and more generally of narratology theory (Bal, 2009). Conventionally, an event is defined as an occurrence taking place at a certain time at a specific location. Even if the EDM provides the concept of event, it does not define any relation between one event and another. Furthermore, the EDM does not provide a classification of events by type, nor does it define roles for the participants in an event. These are limitations that have to be overcome to represent narratives in Europeana.

The Europeana portal provides a search engine that works on the metadata of digital objects. It is possible to refine the search through some facets provided by the portal: (i) collection (e.g., art, fashion, music); (ii) media type; (iii) copyright (free re-use, limited re-use, or not re-usable); (iv) providing country; (v) language; (vi) aggregator (organisations performing a data aggregation role); and (vii) institution that provides the data. The portal does not currently allow a search based on events, nor does it supply a visualisation of the digital objects based on the events to which the objects are related. We performed a SPARQL query through the Europeana SPARQL endpoint[9] in order to extract all the instances of the edm:Event class in the current Europeana knowledge base. The query returned zero results. This suggests, as already pointed out in Introduction, that the data collected by Europeana are not organised according to the event-centric approach, nor is this information introduced by the data aggregators that work in the Europeana network. In order to introduce narratives in Europeana, we developed an ontology for narratives[10] (Bartalesi *et al.*, 2017). To maximise its interoperability, our ontology was developed as an extension of the CIDOC CRM standard ontology. Since the CRM underlies the EDM, our ontology automatically extends the EDM, re-using its definition of event and introducing new relations between events. In particular, the ontology for narratives allows:

- Representing the events that compose a narrative, linked to each other using three types of semantic relations:
  - Temporal occurrence relation, associating each event with a time interval during which the event occurs. We reused the intervals defined in Allen's temporal logic (Allen, 1984), which are adopted by the CRM.
  - Causality relation, linking events that in normal discourse are predicated to have a cause–effect relation, e.g., *the tsunami caused the destruction of towns, villages and livelihoods*. The causality concept is represented by introducing a new relation of causal dependency, named *causallyDependsOn*. The only causal property of the CRM is *P17 was motivated by*, but it cannot be used for narratives since it relates activities but not events. We are not interested in modelling the mechanical causal relationships that connect events in a physical or chemical process. Rather, we are interested in a more generic notion of causality as defined by the narrator according to her/his own interpretation of the events.
  - Mereological relation, connecting events to other events that include them as parts, e.g., the birth of Klimt is part of the life of Klimt, represented using the CRM property *P9 consists of*.
- Linking an event with the related digital objects included in a DL.
- Representing the data provenance in order to describe the inferential process of a narrator who reconstructs a narrative starting from the primary and secondary sources.

We used the NBVT to populate the ontology. The tool is described in the next section.

5. Narrative building and visualising tool

We developed a narrative building and visualising tool (NBVT for short)[11] that allows users to construct and visualise narratives. The NBVT is Web-based and written in HTML5, CSS, and JavaScript with jQuery.[12] It interfaces with a CouchDB[13] database to store and retrieve data on subsequent loadings. The communication with the

[9] `http://sparql.europeana.eu`.

[10] A description of the classes and relations of our ontology is available at `https://dlnarratives.eu/ontology`.

[11] `https://dlnarratives.eu/tool.html`.

[12] `https://jquery.com`.

[13] `https://couchdb.apache.org`.

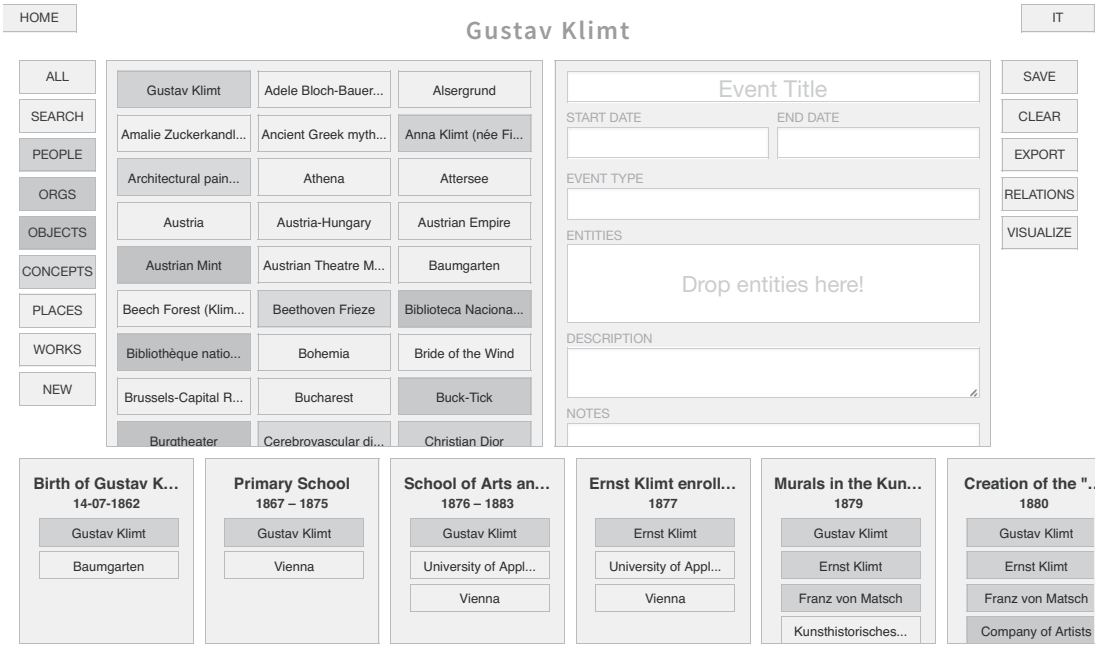

Fig. 1. Main view of the NBVT: on the left-hand side the entities automatically extracted from Wikidata, on the right-hand side the form for constructing an event, and at the bottom the created events in chronological order.

database is handled by the PouchDB[14] library. Figure 1 shows the main interface of the NBVT.

To simplify the insertion of data in a narrative by the user and the subsequent population of the ontology, we decided to build the NBVT on top of an existing knowledge base. The knowledge base we adopted as reference is Wikidata, a free collaborative knowledge base operated by the Wikimedia Foundation (Vrandečić and Krötzsch, 2014). Wikidata was built by extracting structured knowledge from Wikipedia and other collaborative projects (e.g., Wikisource, Wikibooks). It currently contains more than 25 million entities and allows exporting the data, through the Wikidata API and a SPARQL endpoint called Wikidata Query Service.[15] Wikidata encourages collaborative addition and editing of the data by its users through manual and automated means. The NBVT takes as input resources inserted manually by the user or imported automatically from Wikidata. It also initially imports a few default events from Wikidata, such as births, deaths, and company foundations. Then, the user adds the remaining events of the narrative one by one, by inserting the following information: (i) the title of the event; (ii) the start and end dates of the event; (iii) the event type; (iv) a set of entities imported from Wikidata or defined by the user (e.g., person, location, object); (v) for each entity, one or more primary or secondary sources and, in the case of people, the role they played in the event; (vi) a textual description of the event; (vii) optional textual notes; (viii) one or more digital objects.

At any moment during the narrative construction, the user can switch to the relations view, which allows her/him to link the events through causal or mereological relations through a drag-and-drop interaction. An example of mereological relation in the narrative about the life of Gustav Klimt is the one that links the events of the attendance of the primary school and later of the School of Arts and Crafts with a super-event which can be called "Education of Klimt." On the other hand, an example of a causal relation is the one that connects the painting of the murals in the Burgtheater with the caused event of the awarding of Klimt with the Golden Order of Merit by Emperor Franz Josef I of Austria.

The output of the Web interface is an intermediate JSON (JavaScript Object Notation)[16] representation of the narrative, later converted to an OWL (Web Ontology Language) graph (McGuinness and Van Harmelen, 2004) by a triplifier. The resulting representation is uploaded into a Blazegraph[17] triple store.

The NBVT allows users to have different visualisations of the narrative.[18] First of all, the events of a narrative are placed on a timeline, implemented using the TimelineJS library.[19] Each event on the timeline has a

[14] `https://pouchdb.com`.
[15] `https://query.wikidata.org`.
[16] `http://json.org`.
[17] `https://www.blazegraph.com`.
[18] On our Web site, four narratives are available: `https://dlnarratives.eu/narratives.html`.
[19] `https://timeline.knightlab.com`.

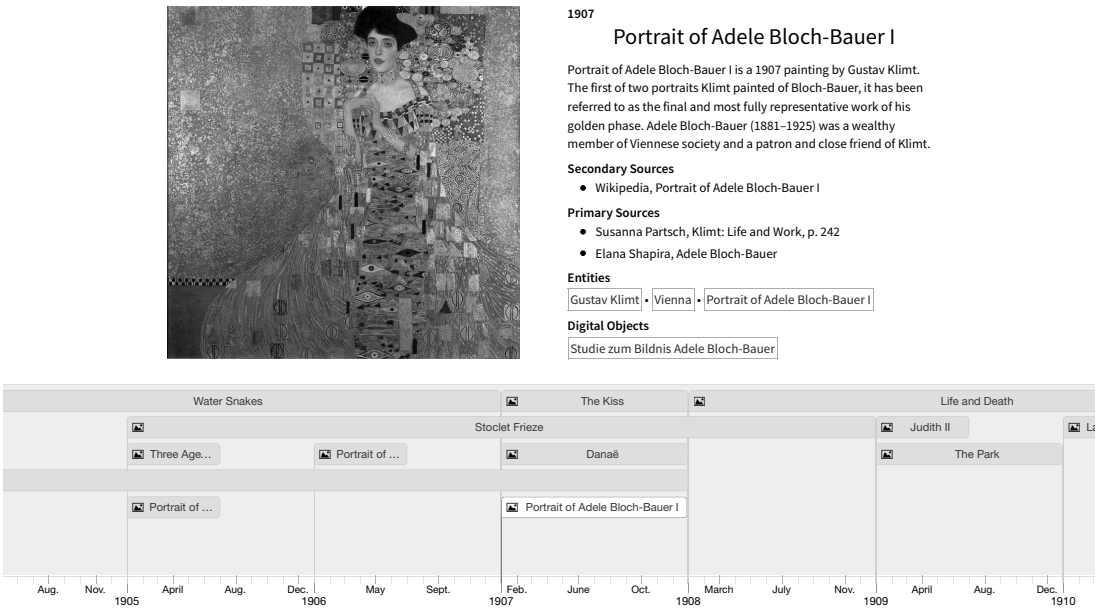

Fig. 2. Event in the timeline of Klimt's biography, showing the textual description of the event, the related digital object, the primary and secondary sources, the related entities and an image from Wikimedia Commons.

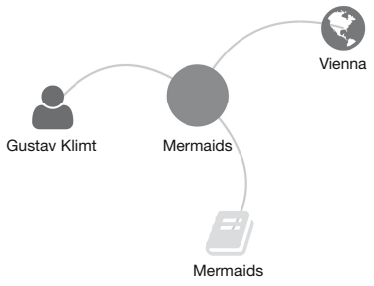

Fig. 3. Graph visualisation showing the event of the painting of Klimt's *Mermaids*.

title and a list of its components, in particular: (i) date, (ii) factual components, (iii) primary and secondary sources, (iv) related digital objects, (v) related images imported from Wikimedia Commons. Figure 2 shows an event of the timeline about Gustav Klimt's life.

The NBVT is also able to generate three network graph visualisations, showing (i) an event chosen by the user with all entities that participated in it (this visualisation is shown in Fig. 3); (ii) an entity chosen by the user with all events in which it participated; (iii) the digital objects related to a specific event selected by the user. By clicking on the event or on the entities, it is possible to access a description of them. These network graphs are implemented using the vis.js library.[20]

An important requirement for a scholar who studies historical events is knowing their primary sources. For each event of the narrative, the NBVT allows visualising information about primary sources, in particular (i) the title and the author of the primary source; (ii) the textual fragment of the primary source that describes the event; (iii) the bibliographic reference of the textual fragment. This information is visualised in a tabular format.

Finally, the user can visualise all events that occurred in a specified period of time, chosen by the user. The results of the query are shown in the form of a table, reporting the dates of each event.

6. Enriching the events of the narrative through Europeana

When a user creates an event, the NBVT suggests a list of digital objects extracted from Europeana that are deemed relevant to the event. The objects are ranked according to a score from 1 to 100, where 100 indicates the maximum relevance. In order to retrieve all the exact matches between the Europeana digital objects and the entities used in the narrative, the NBVT performs several queries. Since it uses Wikidata as reference and both Europeana and Wikidata contain links to each other, the first logical step is to look for matching URIs between the two knowledge bases. Currently, Europeana contains 31,479 links to distinct Wikidata entities (owl:sameAs and dc:relation properties), while Wikidata contains 76,151 links to distinct Europeana objects (P727 property). Given that these two sets do not coincide, the NBVT performs two queries: (i) one on the Europeana SPARQL

[20] http://visjs.org/.

endpoint, looking for the URIs of all Wikidata entities contained in the event and classified in the CRM as E73 Information Object, which may be part of Europeana; (ii) the other on the Wikidata Query Service, looking for links to Europeana from the Wikidata entities. If an exact match is found, the corresponding digital object is added to the list of results with the maximum score (100).

The second phase of the ranking is made on specific metadata fields of the Europeana records. We performed several experiments on the ranking and found that, for our purposes, the solution presented below is a good trade-off between precision and recall (Buckland and Gey, 1994) of the results. However, due to space limitations we cannot account for all the technical details of the ranking method. We just report its main steps and the main characteristics of each step. First of all, the NBVT performs a query to extract all digital objects related to the subject of the narrative (e.g. Gustav Klimt) by searching for all digital objects containing the name or the URI of the entity in one of the following Europeana metadata fields: (i) dc:title, (ii) dc:creator, (iii) dc:contributor, (iv) dc:subject. Then, a ranking is performed on the results of the query using four parameters for each extracted digital object:

1. The title of the event is matched against that of the digital object. Given the current state of Europeana records, we believe that this is the most important clue to the relevance of the object. Indeed, many Europeana records are incomplete, but they always have a title. We compute the similarity between the title of the event and the title of the digital object by splitting that of the latter into a set of substrings, with each string containing a number of words equal to the number of words in the title of the event. For instance, if the title of the event is "Portrait of Sonja Knips" and that of the digital object is "Studie für das Bildnis Sonja Knips" (Study for the Portrait of Sonja Knips), the array will contain the strings "Studie für das Bildnis", "für das Bildnis Sonja", and "das Bildnis Sonja Knips". Then, we apply the Damerau–Levenshtein similarity measure (Damerau, 1964; Levenshtein, 1966) between the title of the event and each of the substrings, and we take the maximum score. For instance, in the previous example the maximum score would be given to "das Bildnis Sonja Knips." Since long strings obtain a disproportionately large score when applying Damerau–Levenshtein, we normalise the strings using the following formula:

$$score = 100 \times \frac{(longLength - damlev(str1, str2))}{longLength},$$

where *str1* and str2 are the two strings, *longLength* is the length of the longer string and *damlev* is the function computing the Damerau–Levenshtein distance. This formula returns a score from 1 to 100, which is taken as the weight of the parameter.

2. For each person participating in the event, its Wikidata URI is matched against the following metadata fields of the digital object: (i) dc:creator, (ii) dc:contributor, (iii) dc:subject. The result of this evaluation is 1 if an exact match is found, i.e., if the URI is present in the record, or 0 if it is not found. Then, the sum of the results is given a weight of 50 because we want to balance this step with the average weighted result of the previous step.

3. For each person participating in the event, excluding those that were matched in the previous step, the person's name in natural language, as reported in Wikidata, is matched against the following metadata fields of the digital object: (i) dc:title, (ii) dc:creator, (iii) dc:contributor, (iv) dc:subject, using the similarity measure described by the formula in Step 1. The reason for excluding the entities that were matched in the previous step is that in the fields (ii), (iii) and (iv) both the URIs of the entities and the related strings are reported. The result of this evaluation is 1 if the similarity is greater than 90 (indicating about 10% differing characters), otherwise it is 0. Then, the sum of the results is given weight of 50 for the reason reported in the previous step.

4. The period of occurrence of the event is matched against the dc:date metadata field of the digital object. If the period of occurrence contains the value of the field, the score is incremented by 20. The semantics of dc:date is too generic to express a higher factor. For instance, dc:date can indicate both the date of creation of a physical object that the digital object represents, and the date of acquisition of the digital object.

Finally, the sum of the four parameters is normalised and the list of digital objects is sorted based on the relevance score and presented to the user through a graphical interface. The interface initially shows the five most relevant digital objects, but the user is able to load more if needed. Figure 4 shows the interface with the five most relevant digital objects for the event "Beethoven Frieze", describing Klimt's painting of a frieze for the 14th Vienna Secessionist Exhibition in 1901. Europeana contains several digital objects about this event, including photographs of the frieze and preparatory drawings. The interface allows the user to select some of these objects and link them to the event.

Through this functionality, we are able to propose to the user the objects that, according to the knowledge

Fig. 4. Example showing the interface to enrich events in the NBVT through Europeana digital objects. The title of the event is "Beethoven Frieze", and the five most relevant digital objects are shown in the interface.

stored in Europeana, may be relevant to the subject of each event of the narrative.

7. Narrative of Klimt's life as a case study

In order to introduce narratives in Europeana, we performed an experiment creating the narrative of the life of the Austrian painter Gustav Klimt. This artist is well represented in the DL, where a search for the string "Gustav Klimt" currently returns 353 objects.

Since we are not art historians nor experts on Klimt, we decided to build the narrative based on the English Wikipedia page about the painter.[21] In the Wikipedia text we identified the main events of Klimt's life, and we reproduced them using the NBVT. For each event, we defined the related entities (e.g., persons, physical objects, location, time). We also reported the fragment of text that describes the event in Wikipedia, and one or more images from Wikimedia Commons that are related to the entities that compose the event. Furthermore, we reported the primary sources cited by Wikipedia, on the basis of which an event is placed in the narration.

The Klimt narrative is composed of a total of 54 events. 31 of them are connected with Europeana digital objects, and 18 are linked to more than one digital object. The total number of digital objects in the narrative is 127, which is 36% of all Klimt-related objects in Europeana. It should be noted that several Europeana objects are not related to Klimt's biography, e.g., posters or modern objects inspired by Klimt. More than 70% of the entities used in the narrative were imported from Wikidata. The first version of the tool, presented by Meghini *et al.* (2017), allowed the creation of the narrative in about 3 person-days (7 hours per day). After adding the functionality to automatically extract and suggest to the user the relevant Europeana digital objects, the time was reduced to 12 hours.

[21] https://en.wikipedia.org/wiki/Gustav_Klimt.

In order to integrate our narratives into Europeana, the NBVT enriches them with metadata that describe the narrated topic. These metadata could be matched with those contained in Europeana to enhance its search functionality. When a user queries Europeana, she/he could obtain as a response one or more narratives related to the topic of the search, along with the classical ranked list of digital objects. In the Web interface, the timeline of each narrative could be visualised by adding a specific menu section, titled "Narratives", in the "Refine your search" menu on the left side of the page. Europeana also has an "Explore" section in its upper menu, providing a list of particular views on the Europeana data, e.g., views for people, or time periods. A new entry could be added to this section, showing all narratives collected in Europeana. If there is more than one narrative about the same topic, they could be aggregated by the subject. By exploring this page, users could have an immediate idea of all narratives that Europeana stores and also of the digital objects related to each narrative.

8. Conclusions and future work

We have presented a study in order to introduce narratives in digital libraries. The general idea is to enrich those libraries with semantic networks of events providing a meaningful contextualisation of the digital objects collected in them. In particular, we performed an experiment based on Europeana. First of all, we reported some use cases that would call for the introduction of narratives in digital libraries. Then, we have discussed how to extend the Europeana data model in order to realise such introduction in Europeana and we have briefly presented an ontology for representing narratives we created.

A first experiment was reported consisting in the creation of a narrative of the life of Gustav Klimt, the Austrian painter. Such creation was performed by using a tool for building and visualising the narratives we

developed. We also presented a new functionality of the tool that allows automatically suggesting to the users the appropriate Europeana digital objects that can be assigned to the events of a narrative, in order to enrich it. We plan to extend the experiment to other artists and we are also working on improving our research to (i) represent and reason about the temporal relations between events; (ii) represent in a richer way the text of narratives, called narrations, and the relation between narration fragments and their semantic counterparts, events and objects; (iii) introduce narrative templates capturing recurring plots.

References

Allen, J.F. (1984). Towards a general theory of action and time, *Artificial Intelligence* **23**(2): 123–154.

Bal, M. (2009). *Narratology: Introduction to the Theory of Narrative*, University of Toronto Press, Toronto.

Bartalesi, V., Meghini, C. and Metilli, D. (2017). A conceptualisation of narratives and its expression in the CRM, *International Journal of Metadata, Semantics and Ontologies* **12**(1): 35–46.

Buckland, M. and Gey, F. (1994). The relationship between recall and precision, *Journal of the American Society for Information Science* **45**(1): 12.

Clough, P., Hill, T., Paramita, M.L. and Goodale, P. (2017). Europeana: What users search for and why, *International Conference on Theory and Practice of Digital Libraries, Thessaloniki, Greece*, pp. 207–219.

Damerau, F.J. (1964). A technique for computer detection and correction of spelling errors, *Communications of the ACM* **7**(3): 171–176.

Doerr, M. (2003). The CIDOC conceptual reference module: An ontological approach to semantic interoperability of metadata, *AI Magazine* **24**(3): 75.

Doerr, M., Gradmann, S., Hennicke, S., Isaac, A., Meghini, C. and van de Sompel, H. (2010). The Europeana Data Model (EDM), *World Library and Information Congress: 76th IFLA General Conference and Assembly, Gothenburg, Sweden*, pp. 10–15.

Europeana (2014). Results of the Europeana user survey 2014, *Technical report*, Europeana, `https://pro.europeana.eu/post/results-of-the-europeana-user-survey-2014`.

Fernie, K., Griffiths, J., Stevenson, M., Clough, P., Goodale, P., Hall, M., Archer, P., Chandrinos, K., Agirre, E., de Lacalle, O.L., de Polo, A. and Bergheim, R. (2012). PATHS: Personalising access to cultural heritage spaces, *18th International Conference on Virtual Systems and Multimedia (VSMM), Milan, Italy*, pp. 469–474.

Herman, D. (2011). *Basic Elements of Narrative*, John Wiley & Sons, Oxford.

Levenshtein, V.I. (1966). Binary codes capable of correcting deletions, insertions, and reversals, *Soviet Physics Doklady* **10**(8): 707–710.

Mannocci, A., Casarosa, V., Manghi, P. and Zoppi, F. (2014). The Europeana network of ancient Greek and Latin epigraphy data infrastructure, *Research Conference on Metadata and Semantics Research, Karlsruhe, Germany*, pp. 286–300.

McGuinness, D.L. and Van Harmelen, F. (2004). OWL web ontology language overview, *W3C Recommendation* **10**(10): 2004.

Meghini, C., Bartalesi, V., Metilli, D. and Benedetti, F. (2017). Introducing narratives in Europeana: Preliminary steps, *in* T. Morzy *et al.* (Eds.), *Advances in Databases and Information Systems*, Springer, Berlin/Heidelberg, pp. 333–342.

Petras, V., Hill, T., Stiller, J. and Gade, M. (2017). Europeana, a search engine for digitised cultural heritage material, *Datenbank Spektrum* **17**(1): 41–46.

Ruotsalo, T. and Hyvönen, E. (2007). An event-based approach for semantic metadata interoperability, *in* K. Aberer *et al.* (Eds.), *The Semantic Web*, Springer, Berlin/Heidelberg, pp. 409–422.

Shaw, R., Troncy, R. and Hardman, L. (2009). LODE: Linking open descriptions of events, *Asian Semantic Web Conference, Shanghai, China*, pp. 153–167.

Van Hage, W.R., Malaisé, V., Segers, R., Hollink, L. and Schreiber, G. (2011). Design and use of the simple event model (SEM), *Web Semantics: Science, Services and Agents on the World Wide Web* **9**(2): 128–136.

Vrandečić, D. and Krötzsch, M. (2014). Wikidata: A free collaborative knowledge base, *Communications of the ACM* **57**(10): 78–85.

Carlo Meghini is a prime researcher at ISTI-CNR in the area of conceptual modelling, digital libraries and digital preservation. He graduated in computer science from the University of Pisa, Italy, in 1979 with a research thesis on distributed databases. He has participated in several EC funded research actions in the areas of information systems, digital libraries and digital preservation. He has been involved in European projects since 1988, in the areas of digital libraries (DELOS Network of Excellence, BRICKS, EDLNet, Europeana version 1.0, Europeana version 2.0, ASSETS) and digital preservation (CASPAR). Since 2007 he has been involved in the making of the European digital library (Europeana, `www.europeana.eu`), taking care of the scientific aspects of the project. He has published about 100 scientific papers in international journals, books and conferences.

Valentina Bartalesi received her PhD in information engineering at the University of Pisa. She also holds a Master's degree in digital humanities from the University of Pisa. She had worked for seven years at CELCT, the Italian Center for the Evaluation of Language and Communication Technologies. Since 2011, she has been with the Institute of Information Science and Technologies of CNR in Pisa, and currently she is a researcher at the Networked Multimedia Information Systems Laboratory. Her research interests are mainly related to semantic Web technologies and the development of formal ontologies to represent textual content to be accessed through Web applications. She is a co-author of 28 research papers presented at national and international conferences and in journals.

Daniele Metilli is a research associate at the Networked Multimedia Information Systems Laboratory of ISTI-CNR and a PhD candidate in computer science at the University of Pisa. He holds a Master's degree in digital humanities from the University of Pisa and a Bachelor's degree in computer engineering from the Polytechnic University of Milan. His research interests revolve around the semantic Web, digital libraries, natural language processing, and online collaborative projects. Currently he is working on the formal representation of narratives and on narrative extraction from natural language text.

Filippo Benedetti is a PhD student of information engineering at the University of Pisa. He has obtained a Master's degree in computer science at the University of Pisa. Since 2015 he has been working for ISTI-CNR as a graduate fellow at the Networked Multimedia Information Systems Laboratory. Initially he contributed to the Smart Campus project and currently his research focuses on semantic Web technologies, ontology reasoning and other techniques to infer knowledge.

Received: 6 February 2018
Revised: 3 October 2018
Re-revised: 20 October 2018
Accepted: 16 November 2018