The GraphBRAIN Framework for Knowledge Graph Management and its Applications to Cultural Heritage

Stefano Ferilli[1][0000-0003-1118-0601], Eleonora Bernasconi[1][0000-0003-3142-3084], Davide Di Pierro[1][0000-0002-8081-3292], and Domenico Redavid[1][0000-0003-2196-7598]

University of Bari, Bari, Italy
name.surname@uniba.it

Abstract. The traditional record-based approach to the description of Cultural Heritage is nowadays obsolete. It is unable to properly handle complex descriptions and it cannot support advanced functions provided by Artificial Intelligence techniques for helping practitioners, scholars, researchers and end-users in carrying out their tasks. A graph-based, semantic approach is needed, such as that provided by Semantic Web solutions. Also, a 'holistic' description approach is needed, that includes and inter-connects all branches and types of Cultural Heritage, and that is not limited to describing just the formal metadata of cultural objects, but can deal with their content, physicality, context and lifecycle, as well. The GraphBRAIN framework and technology for Knowledge Graph management enforces all these ideas and enjoys improved efficiency, expressiveness, and flexibility thanks to the the use of the LPG model for knowledge representation. This paper describes GraphBRAIN and its application to several Cultural Heritage-related fields, including digital libraries, archives and museums, history of computing, and tourism as a way to boost fruition of these items.

Keywords: Semantic Web · Knowledge graph · Knowledge representation · Cultural Heritage

1 Introduction & Motivations

For many good reasons, description of Cultural Heritage (CH) items has traditionally been organized in the form of records with a fixed number of pre-defined fields. These fields were organized around a set of metadata mostly oriented to describing the formal aspects of the items. This choice made perfect sense based on the technology available in the past, and to the descriptions being thought specifically for use by experts (practitioners and researchers). The landscape has now changed in both perspectives. Digital technologies provide representational and computational support much more advanced and complex[1] than what was

[1] By 'complex' we mean with a very large number of fields, with variable or repeated fields, and with fields which are in turn record themselves.

possible when using paper cards. Also, there is a trend in the last decades toward opening CH to the wider public, which are often enthusiasts or curious, without any technical skill in cataloging and with an immensely broader range of motivations, interests, goals, backgrounds, preferences, etc.

One of the drawbacks of legacy approaches to cataloging and description of CH is that a different record structure, with different sets of fields, was defined for different kinds of CH. Just to mention the most prominent ones, we may think of Galleries, Libraries, Archives and Museums (the so-called GLAM). First of all, this specialization requires new description standards when new kinds of CH items are approached (e.g., new standards have been issued for the description of scientific instruments, or for electronic instruments, after realizing that the standard records for museum items were totally unable to capture their fundamental peculiarities). The Central Institute for Cataloging and Documentation (ICCU) of the Italian Ministry of Culture currently provides for 4 areas of protection (archaeological, architectural and landscape, demo-ethno-anthropological and historical-artistic), and defined 9 different catalog cards organized according to the different subject areas: archaeological heritage, architectural and landscape goods, demo-ethno-anthropological goods, photographic heritage, musical goods, naturalistic goods, numismatic goods, scientific and technological goods, historical and artistic goods[2]. Still, this is insufficient: the card for scientific and technological goods, while fitting standard equipment, is totally unable to capture the complexity of, e.g., the computing domain, especially from its historical perspective, that is of utmost interest for CH [13]. The landscape, and other kinds of immaterial CH, are progressively attracting attention, both for preservation and for economic exploitment purposes. And new branches and needs will continuously arise as technology and society develop.

Secondly, a consequence of having different description standards means that the various subject areas follow their own way, making it difficult to make them converge, while the CH field is obviously a coherent whole (e.g., archive documents describe the history of museum items, whose interpretation is reported in scholarly books). A solution is needed that can be applied to CH as a whole, while still keeping the specificities of each of its branches, and that is flexible, so as to allow easy extension and variation to cover new or changed needs of the field. Even more, we think that this solution should go beyond the pure and traditional focus on formal CH metadata, also including all knowledge that is non-strictly related to the single items, or even to CH in general, but that can provide precious 'glue' to connect and inter-relate them, opening new possibilities for their understanding, management and exploitation. We strongly believe that such a unified framework will boost the field, providing immense expansion opportunities and unprecedented support to all the stakeholders (practitioners, researchers, scholars, enthusiasts, tourists, end-users), also thanks to the use of state-of-the-art Artificial Intelligence (AI) solutions.

Initially, the legacy approach to description and cataloging has been simply ported to digital as-is. Then, after realizing the flexibility that digital solutions

[2] http://iccd.beniculturali.it/it/settoridisciplinari

could provide, it was expanded, but still centered on the record-based approach, just providing for many more fields, structuring them, and making most of them optional, as in the MARC series of standards. Instead, a much better representational option to overcome the limitations of record-based descriptions is the graph model. Basically, it consists of nodes, typically representing objects, and arcs, representing associations between pairs of objects.

The most outstanding advantage of graphs is their very intuitive interpretation by humans, and ubiquitous use in everyday life, paired with their having a mathematical definition that allows to build a whole formal theory on them. In fact, graphs have been chosen as the basic structure for research in the Knowledge Representation and Reasoning (KRR) branch of AI. KRR investigates how to represent, store and manipulate knowledge, in so-called Knowledge Bases (KBs). More specifically, when KBs are based on graph representations, they are called Knowledge Graphs (KGs). KBs/KGs typically include two components: the ontology (defining what can be represented, how, and what are its properties and behavior) and the instances (i.e., the actual data). The former provides meaning to the latter, and allows different entities or systems to interoperate, by assigning the very same meaning to the same concepts and objects. This is one of the main objectives of the Semantic Web, that indeed adopted this approach and developed its own standards, formalisms and storage solutions for it.

In this paper we propose a KG-based approach to CH description and manipulation. Still, we propose to depart from the standard Semantic Web practices, and pursue its same objectives but starting from a database (DB) perspective, for several reasons. First, this is the traditional setting in the field of CH description. Second, this allows us to take advantage of the efficiency and scalability of the latter, while ensuring semantics and interoperability as in the former. In fact, modern DB solutions are available that rely on the graph model. Third, not being bound to the SW representations, we may apply a wider range of tools to our knowledge, and in particular many AI techniques that can support advanced tasks for the final users of the knowledge. To support our vision we developed the GraphBRAIN framework and platform. It uses technology from the DB community for storing instances, and superimposes ontologies that, on the DB side, are interpreted as data schemas, while, on the outer world, enable semantic-based ineroperability. Among others, GraphBRAIN was applied to the field of CH, and to several branches thereof. Here we will provide an overall account of these applications, to show the power of our solution and its potential.

The rest of this paper is organized as follows. After discussing related work in the next section, we summarize the GraphBRAIN framework in Section 3, and review its applications to the CH domain in Section 4. Finally, Section 5 concludes the work and outlines future work issues.

2 Related Work

In this section we will explore two different aspects of related work on KGs for CH, namely the existing KBs and interfaces.

2.1 Knowledge Bases for Cultural Heritage

The development and curation of domain-specific knowledge structures have traditionally been essential in the humanities [18]. Some efforts focused on porting existing information to digital: Pleiades [15], a repository of data pertaining to geographical locations, with relevance to the examination of ancient literature and history; Papyri.info [5], a search engine that seamlessly integrates multiple DBs containing ancient documents; MANTIS [17], the semantically enriched DB maintained by the American Numismatic Society, focused on the comprehensive study of coins from various historical periods and cultures; Open Context[20], a repository encompassing diverse resources, including archaeological reports; Trismegistos [6], a metadata platformcatering to the study of texts from the Ancient World, housing data related to ancient documents, individuals, and locations; EDH [16], the Epigraphic Database Heidelberg, a search tool dedicated to Latin epigraphic data.

In this direction, the interest in semantic annotation through formal languages has also been active in the Semantic Web since its beginning. In [7] the characteristics of CH are analyzed to identify how ontologies could be used to improve CH information management. In particular, this paper analyzes the integration of different schemas using the CIDOC-CRM[3] ontology (ISO 21127 standard) as a reference. This ontology, used by various cultural organizations worldwide, was developed from the bottom up by integrating semantic contents of various DB schemas and documentation structures from all kinds of museum disciplines, archives and libraries. In order to keep it compact, only a part of the initial concepts and properties were used, thus making it unsuitable for handling complex scenarios that combine descriptive and management aspects of CH. To enable semantic interoperability, an OWL version of this ontology[4] is used by the British Museum. It remains, however, difficult to use in scenarios involving the combination with additional aspects that might aid fruition of the annotated material.

Worth mentioning for our purposes are some National projects. A relevant effort was carried out in the Netherlands and Finland for the Dutch national project CLARIAH [25], concerned with Digital Heritage from 3 perspectives: economic and social history, language, and media studies. To tackle the variety of facets, the project used 26 different ontologies, taxonomies, classification systems and lexicons, of which many applicable to broader context (such as SKOS, PROV and Schema.org), and 6 developed from scratch to model new humanities domains. The knowledge representation and storage was based on Semantic Web technology, focusing on Linked Open Data. However, the authors note that this setting was not comfortable for users not familiar with these technologies. In Italy, Cultural-ON (Cultural ONtology) [21] was developed, an ontology aimed at modeling the data regarding cultural institutes or sites, their contact points, all multimedia files which describe them, the agents that play a specific role

[3] https://cidoc-crm.org/

[4] http://erlangen-crm.org/current

in them, events that can take place in them, and any other information useful to the public in order to access them. It is aligned with external ontologies (FOAF, PROV, schema.org, Dublin Core, etc.). ArCo (Architecture of Knowledge) [3], an ontology for, and a KG of, Italian Cultural Heritage, models many types of cultural properties (including technological heritage), for which it allows to capture details such as elements affixed on cultural properties, copies, forgeries and other works related to a cultural property, specific surveys, cadastral information, historical locations, the communication medium of intangible demo-ethno-anthropological heritage, etc. It currently reuses, and is aligned to, CIDOC-CRM, EDM, Cultural-ON, and OntoPiA. These projects were sponsored by the Italian Ministry of Culture. We take inspiration from, and aim at being aligned with, these works, but we propose to use a different technology for handling this information, and envision a much broader context to be described in CH ontologies.

Still from the point of view of using annotations in complex scenarios, [4] presents an ontology-based approach to improve data retrieval by expert users in the CH field, e.g, archaeologists, art historians, geologists, etc. It proposes an advanced architecture with semantic search capability that can transform a vast amount of data into linked concepts for easy information comparison. All concepts are mapped onto elements of the ArCo ontology, and a semantic query layer allows to execute SPARQL queries (based on simple pattern matching, not on Description Logic reasoning). A mapping between ICCU (the Italian Central Institute for a Unique Catalog) and ArCo is proposed so that SPARQL queries can be run on ArCo to search information in ICCU records. However, as reported in the article, the mapping is partial, and therefore some information would not be found unless ArCo is extended. In contrast, our approach can generate the OWL classes and properties needed for a more complete coverage of SPARQL queries, and can also use several types of reasoning, including those provided by standard Semantic Web reasoners.

Linked datasets of libraries or museums have been gaining traction as an interconnecting spine through which community-specific datasets can build outbound links to contribute to a global graph (e.g. the Virtual International Authority File[5]; the Getty Thesaurus of Geographic Names[6]; the Getty Art and Architecture Thesaurus[7]).

2.2 Creating and Managing Graph Data Model

In this section, we will explore the current landscape of platforms dedicated to creating, managing, and visualizing the models that make up Knowledge Graphs. These tools, including those for data visualization, schemas, and ontologies, provide users with a comprehensive overview of the underlying data structure [1].

5 https://viaf.org
6 https://www.getty.edu/research/tools/vocabularies/tgn/
7 https://www.getty.edu/research/tools/vocabularies/aat/

TopBraid Composer[8] serves as an ontology editing tool with visualization capabilities as an additional feature. The visualization approach draws inspiration from UML and offers horizontal and vertical tree layouts, accompanied by a traditional indented list view. This visualization represents classes and properties as nodes connected by directional edges labelled with their corresponding predicate names. Notably, this visualization operates at the RDF level, treating owl:Class as a distinct node and linking each class to it through an rdf:type edge.

WebVOWL [22] is an online application designed to offer user-friendly visual representations of ontologies, supporting exploration and allowing its users to engage with and personalize ontology visualizations. It must strictly adhere to the Visual Notation for OWL Ontologies (VOWL) to construct graphical depictions of OWL components. VOWL visualizations are automatically generated from JSON files, requiring the conversion of ontologies into JSON format, carried out by the provided Java-based OWL2VOWL converter. The force-directed graph layout relies on a physics simulation, resulting in dynamic animations that continually adjust node positions.

Ontodia[9] [24] is a web-based tool tailored for visualizing ontologies and semantic datasets. It employs a 2D node-link visualization approach and incorporates UML-inspired techniques to convey supplementary information about nodes. Users can choose between force-directed and grid layouts, and there's a hierarchical relationships view for displaying parent-child connections in a tree format. This tool provides flexibility through drag-and-drop functionality, allowing users to customize views by rearranging elements, removing nodes, and toggling links.

The Arca system [2] seamlessly links unstructured content with concepts within a knowledge graph (KG). This integration empowers users to perform intricate data queries and visualize the rich web of semantic connections that bridge concepts and documents.

2.3 Our proposal

In this context, we introduce an approach that addresses the challenges associated with creating and managing ontologies in diverse domains, ultimately enriching the field of CH. Our proposal revolves around harnessing the power of GraphBRAIN, a versatile framework that transcends domain boundaries. GraphBRAIN is poised to revolutionize the ontology development and maintenance process, offering a robust solution for establishing and nurturing knowledge structures that amplify the preservation and exploration of CH. SKATEBOARD, serving as a GraphBRAIN interface, represents a pivotal advancement in the realm of semantic data visualization and exploration. With its multifaceted visualizations, user-friendly interface, and collaborative features, SKATEBOARD provides added value to researchers, data scientists, and knowledge professionals aiming to unlock the full potential of semantic knowledge graphs and derive actionable insights from complex data structures.

[8] https://www.w3.org/wiki/TopBraid
[9] https://github.com/metaphacts/ontodia

3 The GraphBRAIN Framework

GraphBRAIN [14, 11] is a framework developed to cover all tasks in KG management and exploitation based on the combination of leading graph DB technology for instance storage and ontologies for schema description. From the former it draws efficiency and a wide library of data analysis tools; from the latter it draws semantic power, interoperability and the possibility of plugging automated reasoning facilities. Differently from standard Semantic Web approaches, based on the simple atomic triples $< subject, predicate, object >$ provided by the RDF model, it is based on the Labeled Property Graph (LPG) model. LPGs allow to add labels to nodes and arcs, and to specify attributes with their value for both nodes and arcs. Moreover, each node and arc gets a unique identifier, allowing to have different nodes with the same content and different arcs of the same type between the same pair of nodes. This enhances their expressiveness, readability and compactness (it is estimated that an LPG takes one order of magnitude less nodes to store the same information as an RDF graph). As typical in traditional relational DBs, and differently from the Semantic Web approach, GraphBRAIN keeps apart the schema/ontology, described in a GBS file, from the data/instances, stored in the DB.

GraphBRAIN ontologies can be defined using an XML-based formalism specifically designed to match the features of LPGs. It is organized in different sections that allow to: import existing ontologies in order to expand them; define new datatypes in the form of lists or trees of values; define a hierarchy of entities with their attributes; define a hierarchy of relationships with properties (simmetricity, transitivity, functionality, etc.) and their attributes; define axioms in the form of logic formulas (typically rules or constraints) that must be verified by the instances in the KG. The basic datatypes provided by GraphBRAIN are: boolean, integer, real, string, text. Ontologies can be combined using the import section provided that they are compliant to each other, i.e., basically, that their hierarchies of entities are not inconsistent (a class C' is a superclass of class C'' in one ontology, while class C'' is a superclass of C' in the other) and that their attributes are, too (the same attribute in different ontologies must be of the same type). Two ontological components are considered as the same if they have the same name.

The instances handled by one GraphBRAIN installation are stored in a single graph, using the Neo4j graph DB [27]. In GraphBRAIN's use of Neo4j, nodes are used to represent class instances, arcs are used to represent relationship instances (i.e., object properties in the Semantic Web); node labels are used to specify the specific class of the instance represented by the node, and all domains that are relevant to that instance; arc labels are used to specify the relationship expressed by the arc; node attributes are used to represent datatype properties in the Semantic Web; arc attributes have no counterpart in the Semantic Web (they cannot be expressed). Since Neo4j is schemaless, the ontology acts as the schema to determine what information can be stored in the graph, and how. Still, different ontologies may be applied to the same graph, providing different views on the data. The single-graph approach is fundamental for our purposes: even if not visible when using an ontology for accessing the graph, the information associated to other ontologies is still there and may allow indirect connections among the items of the current ontology, that can be explored by the end user while browsing the graph or might be used by graph-based algorithm during their execution.

While mainly designed to allow semantic-based processing on a single KG, GraphBRAIN is open to integration with other resources, especially those available in Semantic Web repositories. In fact, a mapping between the GraphBRAIN formalism and standard Semantic Web is available, allowing interconnection of ontologies and instances alike, and interoperability of systems. As a first advantage, this allows to immediately use ontological reasoners on the knowledge handled by GraphBRAIN. On the other side, a large set of network analysis and graph mining functions can be applied on the data, inherited by the Neo4j libraries and tools. Additionally, not being tightly bound to the standard RDF format, the information in the KG can also be sent to other AI tools, such as rule-based or constraint-based reasoners. We are currently working on the MultiStrategy Reasoning engine GEAR [12], providing a combination of deduction, abduction, abstraction, induction, argumentation, probabilistic reasoning, and abstraction.

A GraphBRAIN API is provided, ensuring that all interactions with the DB happen according to the schema. Given an ontology and a DB, the API provides both basic and advanced functionality on the KG. Basic functionality includes standard CRUD (Create, Read, Update. Delete) operations. For queries, it wraps the Neo4j language Cypher, checking that the specified information is compliant to the ontology before running the query. Advanced functionality include analysis, mining and reasoning functions. E.g.: computing the centrality of an entity instance in the graph according to different algorithms; extracting a relevant portion of the graph starting from given nodes, possibly considering the user profile to obtain a personalized result; finding all possible paths in the graph between given pairs of nodes; checking consistency of the available knowledge; deducing or abducing knowledge that is not explicitly present in the graph; etc.

The API can be used by any third-party application. GraphBRAIN natively provides a Java-based Web Application implemented in JSF technology that allows ontology browsing and development, form-based CRUD operations on the single nodes (entity instances) or arcs (relationship instances), management of attachments and of collaborative interactions to populate the knowledge, etc. A graph-based visualization is also provided, where the user can browse the knowledge, reshaping and expanding the visible portion of the graph, and can apply the various advanced tools provided by the API. Through this interface, ontologies and instances can also be exported or imported to or from other formalisms, including the Semantic Web standard OWL.

Figure 1 shows an ontology as seen in the GraphBRAIN Web Application interface. On the left, the hierarchy of classes can be browsed; selecting a class its attributes are displayed; selecting an enumeration attribute its values are displayed. Classes, attributes, and their values can be added, deleted or renamed

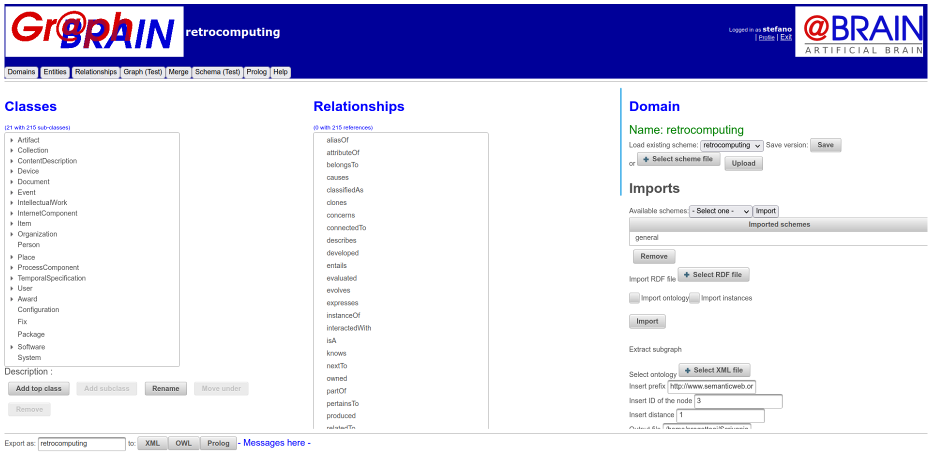

Fig. 1. Ontology in GraphBRAIN KG

from this section. In the center section, the same information and controls as for the classes are available, plus information and tools concerning the inverse relationships and the subject and object classes available for the relationships. On the right, controls to upload ontologies, or to import ontologies in the current one, are provided. On the bottom, controls to save local copies of the current ontology in several formats, including standard OWL, are available.

Figures 2 and 3 show the form-based visualization tabs for entities and relationships, respectively, in GraphBRAIN's Web Application. They allow to select a type of class or relationship and to perform CRUD operations on their instances. The entity or relationship attributes are shown in the middle. Attachments can also be managed on the right, and various kinds of filters and of tools for moving information across different sections are provided. Users can also provide feedback on the available knowledge items using the section on the bottom-left.

Figures 4 and 5 show a portion of the instances in the KG at various levels of zoom, as rendered by the graph-based visualization section of GraphBRAIN's Web Application. The starting nodes to extract the subgraph are those listed in the table on the top-left of the interface and highlighted with a thick border in the graph on the right. Different node colors denote different classes. In Fig. 5 it is better visible that the node content is a summary of the entity instance's data, and that arcs are labeled with the corresponding relationship name. Automatically formed aggregates of strictly related nodes are clearly visible, especially in Fig. 4. For each selected node, its detailed information can be displayed in a table on the bottom of the left column, in a pop-up window, or in a dedicated area on the right column (see Fig. 4). Controls to compute indicators about the node (centrality, etc.), or to run several node-based analysis and mining algorithms on the graph, are available on the node-related areas in the right column or by left clicking on any node in the central area. By right clicking on a blank

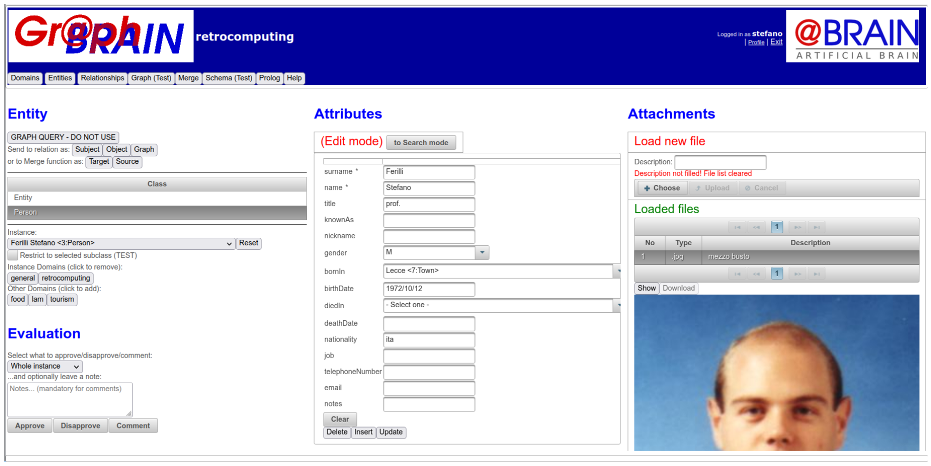

Fig. 2. GraphBRAIN Web Application: Entities form-based management tab

zone in the central area, other controls are available to run various graph-based algorithms (clustering, path finding, personalized subgraph extraction, etc.) and for rendering in natural language selected portions of the graph.

For end-users, a separate interface called SKATEBOARD is also provided as a Web Application (shown in Figure 6). It is mainly based on knowledge browsing and exploration, allowing to visualize, expand or compress a portion of the graph, to look into the single nodes or arcs, and to apply a number of semantic filters that can support the needs of the different users. Since these functions can be applied also to standard Semantic Web KGs, this interface is separate from the previous one, and designed to work also with standard SPARQL endpoints.

4 Cultural Heritage Applications of GraphBRAIN

The GraphBRAIN framework and technology have been used to power several projects, many of which in the field of Cultural Heritage. Here we mention LAM (Libraries/Archives/Museums) [9, 8], Open Science [10], Linguistics [23] and Retrocomputing [13].

As to the LAM domain, it was investigated after the consideration that traditional record-based approaches are obsolete and insufficient to support modern exploitation of, and research on, library, archive and museum items. We termed our graph-based approach a ‘holistic’ one, since it aimed at representing all possible aspects of LAM, not just those related to the formal metadata traditionally used to describe Cultural Heritage items. While starting from a core ontology that is fully aligned with the IFLA proposals for library description, FRBR [19] and LRM [26, 28], we expanded it to make it able to capture the *content* of cultural objects (text, images, concepts expressed therein), their *physicity* (materials, manufacture, shape, structure), their *context* (the periods,

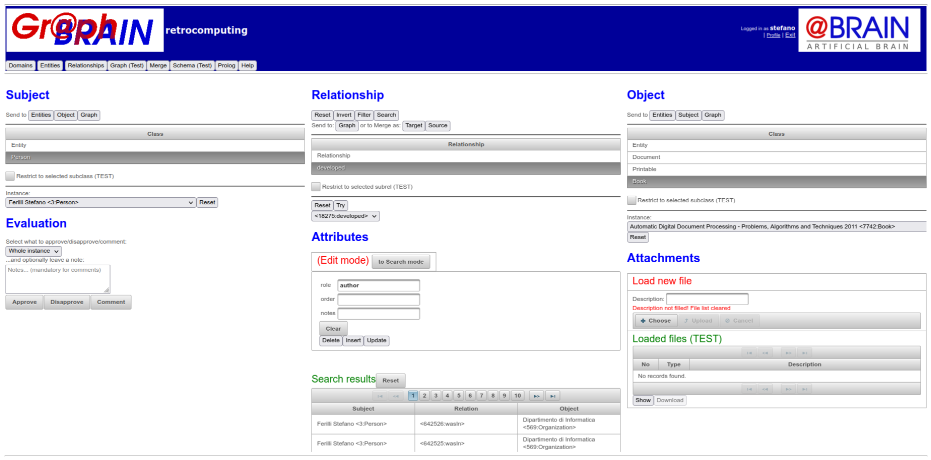

Fig. 3. GraphBRAIN Web Application: Relationships form-based management tab

people, organizations, events, places, etc. to which it has some relationships, and the relationships among these contextual items, even independently of the reference cultural object, that are crucial to find unknown and/or unexpected indirect connections that may suggest, prove or support investigation hypotheses and directions), and even their *lifecycle* (involving all the history of the objects, of its uses and of its users). For specific educational applications, the ontology was also extended and aligned with IEEE's LOM schema for describing Learning Object Metadata[10] and with OERschema for Open Educational Resources[11].

The Open Science domain was an almost straightforward extension of the LAM one. In fact, Digital Libraries are the obvious candidate infrastructure to support it, given that scientific publications are the core of open science. Still, it had to be expanded to describe the context and environment in which scientific development takes place. This involves processes and projects, datasets and corpora, scientific groups and communities, hardware and tools, software and storage facilities, etc. The ontology is currently compliant with OpenAIRE[12] and OAI-ORE[13].

Connected to the LAM domain are also the investigations carried out on linguistics, due to their being based on the sources available in library and archive documents, or on museum items such as epigraphs. Here, the integration of semantic information into language resources was key to open up new avenues of enquiry into the mechanisms of language change. Experiments were run in integrating data from Latin textual corpora and language resources, and showed

[10] https://ieeexplore.ieee.org/document/9262118
[11] https://oerschema.org/docs/schema.html
[12] https://www.openaire.eu/
[13] https://www.openarchives.org/ore/

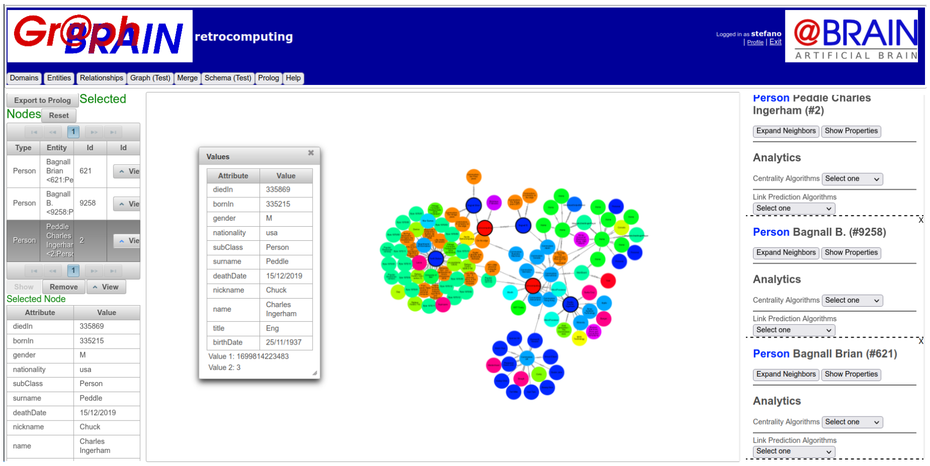

Fig. 4. GraphBRAIN Web Application: Graph-based visualization and management tab (zoomed out)

the potential of the GraphBRAIN framework for research into the mechanisms of semantic change in Latin.

While being one of the oldest applications of GraphBRAIN, the Retrocomputing domain (concerning the history of computing) took great advantage from the subsequent development of the LAM and Open Science perspectives. In fact, the history of computing heavily relies on the LAM perspective because it encompasses books and manuals for the machines and software (concerning libraries), archival documents concerning the persons, organizations and events that took place (concerning archives), and the hardware components (concerning museums). On the other hand, it also relies on the Open Science perspective because the scientific research in Computer Science is primarily concerned with scientific papers, experiments, datasets, tools and, as a consequence, with the hardware and software used in the research and experiments.

Even on its own, the Retrocomputing domain is representative of an extremely complex domain to represent. It involves, and inextricably interconnects, documentation, hardware, software and even immaterial heritage (e.g., the anecdotes that can still be known from the pioneers and central players in the history of computing, that are in many cases still alive and willing to tell behind-the-scenes information that is lost for most of the other, much older disciplines). Hardware and software cannot be understood without their associated documentation; the software is nothing without the hardware to run it; the hardware is dead without software to run; the archival documents and immaterial knowledge are often key to properly understanding all the other items. Also, the traditional fields defined for other types of CH (even those for scientific instruments and electronic equipments) do not fit at all the needs for the description of computing hardware, where nearly each single unit is unique, for several rea-

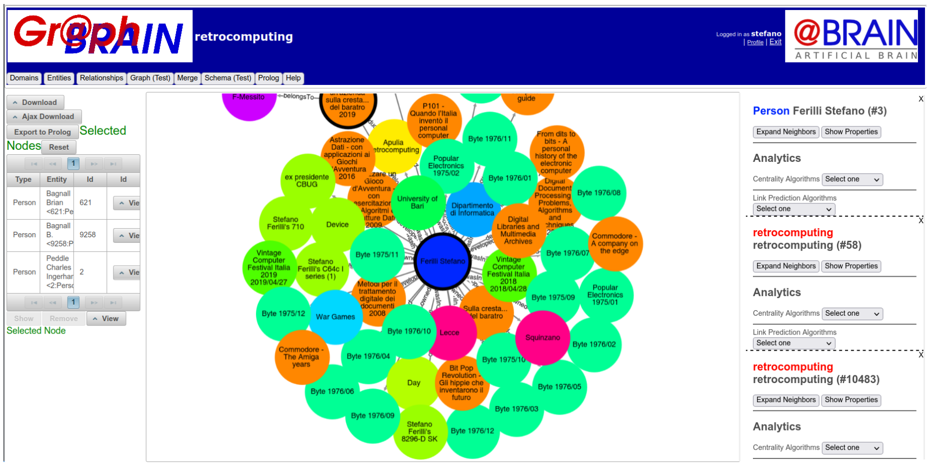

Fig. 5. GraphBRAIN Web Application: Graph-based visualization and management tab (zoomed in)

sons: several versions of the apparently same model may exist, units are highly configurable and expandable, many components are interchangeable while apparently not changing the perceived behavior, several tweaks or modifications can be needed to restore and fix some units, especially old ones for which the original parts are not available anymore.

We include in this list also the Food and Tourism domains. While not CH domains by themselves (albeit, if considering traditional dishes and recipes, or landscapes and folklore, they may well be considered immaterial CH), together with the CH proper sections of the KG, they contribute to make up an ecosystem aimed at enhancement and exploitation of the CH items by final users. This again falls in our holistic perspective, and provides a clear example of how it can open up new possibilities with respect to traditional approaches to CH.

While these ontologies can be connected to each other via a few common entities, that act as bridges between the different domains and allow reuse of knowledge across them, the most relevant opportunity for their interconnection comes from a *general* top-level ontology, defined in GraphBRAIN independently of the various specific domains, and including ubiquitous and highly reusable concepts that can be reused (and specialized, if needed) by the domain-specific ontologies: Person, Organization, Event, Place, Collection, IntellectualWork, Item are just a few prominent examples.

Table 1 reports figures on the current content of the KG that can be freely consulted from the GraphBRAIN's demo prototype available (upon registration) at `http://digitalmind.di.uniba.it:8088/GraphBRAIN/`. Note that the overall number of items is much larger than the number of items labeled with a domain. This is important, because unlabeled items are not part of any specific domain, but allow to indirectly link and inter-relate items otherwise disconnected

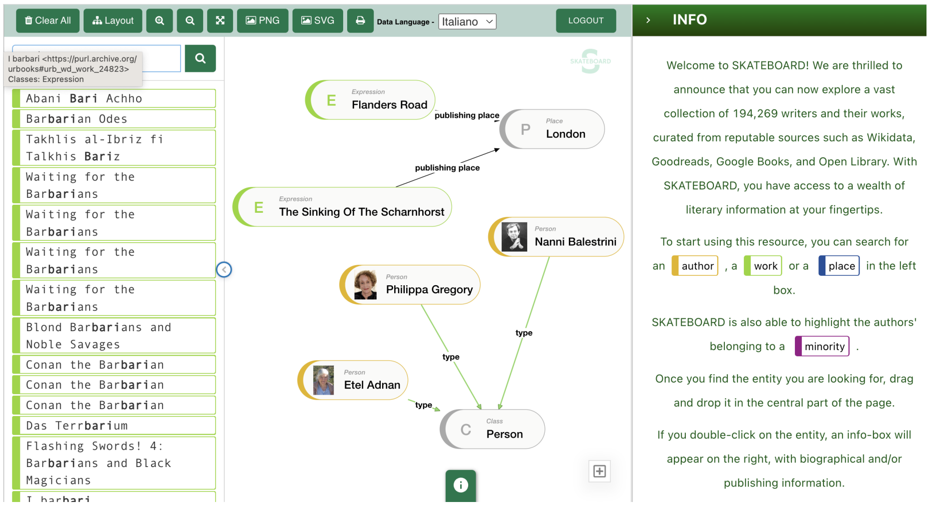

Fig. 6. Instance browsing in GraphBRAIN KG

across domains or even within single domains. Not all domains have figures in Table 1, because some of them are still under investigations and their data were not yet uploaded in the prototype. The most populated domain is Retrocomputing, which could be expected since it is the oldest one. The less populated are Food and Tourism, which were most recently introduced. The number of entity attribute values is much larger than that of relationship attribute values, which again could be expected, since relationships are meaningful by themselves, while entity instances can be identified and distinguished only based on their attributes. On the other hand, the number of relationship instances is typically larger than that of entity instances, because many different relationships can be established among the same set of objects[14].

5 Conclusions & Future Work

Since the traditional record-based approach to the description of Cultural Heritage is nowadays unable to properly handle complex descriptions, or to support advanced functions provided by Artificial Intelligence techniques for helping practitioners, scholars, researchers and end-users in carrying out their tasks, this paper focused on a graph-based, semantic approach, such as that provided by Semantic Web solutions. Also, a 'holistic' description approach is needed,

[14] For each type of relationship, given n objects the number of possible relationship instances is 2^n if only one such instance can be set between a given pair of objects. Since in LPGs many instances of the same relationship may be set between the same pair of objects (distinguished by unique ids associated to each instance), in our case this number is theoretically unbound.

Table 1. Statistics on the content of the current GraphBRAIN prototype's KG

Domain	entity inst.	entity attr.	relationship inst.	relationship attr.
Overall (unlabeled)	337287	2089580	496839	41594
Overall (labeled)	2038	8069	2512	1958
General	102	573	222	132
LAM	63	294	93	69
OpenScience	—	—	—	—
Linguistics	—	—	—	—
Retrocomputing	1688	6801	2142	1757
Food	169	338	47	0
Tourism	14	56	8	0

that includes and interconnects all branches and types of Cultural Heritage, and that is not limited to describing just the formal metadata of cultural objects, but can deal with their content, physicality, context and lifecycle, as well. The GraphBRAIN framework and technology for Knowledge Graph management enforces all these ideas and enjoys improved efficiency, expressiveness, and flexibility thanks to the the use of the LPG model for knowledge representation. This paper described GraphBRAIN and its application to several Cultural Heritage-related fields, including digital libraries, archives and museums, history of computing, and tourism as a way to boost fruition of these items.

Future work is ongoing in several directions, to extend and refine the GraphBRAIN framework, its API and interfaces, and its Cultural Heritage-related KG. In particular, within the effort for Spoke 3 "Digital Libraries, Archives and Philology" of project CHANGES "Cultural Heritage Active innovation for NextGEn Sustainable society", winner of the NRRP program of the Italian Ministry of University and Research, funded by the NextGenerationEU, GraphBRAIN plays a key role in the development and exploitation of a 'holistic' ontology that, starting from the aforementioned descriptions for digital libraries and archives, expands them to support the fields of history and archaeology of books, intellectual property law enforcement, and economic exploitation of library and archive materials. Also, extensions to make the CH ontology compliant with, or aligned to, the CIDOC-CRM and ArCo initiatives are foreseen. Future work will also include releasing an open source version of the GraphBRAIN API.

Acknowledgments

This research was partially supported by projects FAIR – Future AI Research (PE00000013), spoke 6 – Symbiotic AI, and CHANGES – Cultural Heritage Active innovation for Next-GEn Sustainable society (PE00000020), Spoke 3 – Digital Libraries, Archives and Philology, under the NRRP MUR program funded by the NextGenerationEU.

References

1. Bernasconi, E., Ceriani, M., Di Pierro, D., Ferilli, S., Redavid, D.: Linked data interfaces: a survey. Information **14**(9), 483 (2023)
2. Bernasconi, E., Ceriani, M., Mecella, M., Catarci, T.: Design, realization, and user evaluation of the arca system for exploring a digital library. International Journal on Digital Libraries **24**(1), 1–22 (2023)
3. Carriero, V.A., Gangemi, A., Mancinelli, M.L., Marinucci, L., Nuzzolese, A.G., Presutti, V., Veninata, C.: Arco: The italian cultural heritage knowledge graph. In: The Semantic Web - ISWC 2019 - 18th International Semantic Web Conference, Auckland, New Zealand, October 26-30, 2019, Proceedings, Part II. Lecture Notes in Computer Science, vol. 11779, pp. 36–52. Springer (2019). https://doi.org/10.1007/978-3-030-30796-7_3
4. Casillo, M., Santo, M.D., Mosca, R., Santaniello, D.: Sharing the knowledge: exploring cultural heritage through an ontology-based platform. J. Ambient Intell. Humaniz. Comput. **14**(9), 12317–12327 (2023). https://doi.org/10.1007/s12652-023-04652-3
5. Celano, G.G.: An automatic morphological annotation and lemmatization for the idp papyri. Digital Papyrology II. Berlin and Boston: De Gruyter pp. 139–148 (2018)
6. Depauw, M., Gheldof, T.: Trismegistos: An interdisciplinary platform for ancient world texts and related information. In: Theory and Practice of Digital Libraries–TPDL 2013 Selected Workshops: LCPD 2013, SUEDL 2013, DataCur 2013, Held in Valletta, Malta, September 22-26, 2013. Revised Selected Papers 3. pp. 40–52. Springer (2014)
7. Doerr, M.: Ontologies for cultural heritage. In: Staab, S., Studer, R. (eds.) Handbook on Ontologies, pp. 463–486. International Handbooks on Information Systems, Springer (2009). https://doi.org/10.1007/978-3-540-92673-3_21
8. Ferilli, S.: Holistic graph-based representation and ai for digital library management. In: Linking Theory and Practice of Digital Libraries, LNCS, vol. 13541, pp. 485–489. Springer (2022)
9. Ferilli, S., Redavid, D.: An ontology and knowledge graph infrastructure for digital library knowledge representation. In: Digital Libraries: The Era of Big Data and Data Science, Communications in Computer and Information Science, vol. 1177, pp. 47–61. Springer, Berlin, Heidelberg (2020)
10. Ferilli, S., Redavid, D., Pierro, D.D.: Holistic graph-based document representation and management for open science. International Journal on Digital Libraries (IJDL) **11:389**, 23 (2022)
11. Ferilli, S.: Integration strategy and tool between formal ontology and graph database technology. Electronics **10**(21) (2021), https://www.mdpi.com/2079-9292/10/21/2616
12. Ferilli, S.: Gear: A general inference engine for automated multistrategy reasoning. Electronics **12**(2), 256 (2023)
13. Ferilli, S., Redavid, D.: An ontology and a collaborative knowledge base for history of computing. In: 1st International Workshop on Open Data and Ontologies for Cultural Heritage (ODOCH-2019), at the 31st International Conference on Advanced Information Systems Engineering (CAiSE 2016), Rome, Italy. Central Europe (CEUR) Workshop Proceedings, vol. 2375, pp. 49–60 (2019)
14. Ferilli, S., Redavid, D.: The GraphBRAIN system for knowledge graph management and advanced fruition. In: Foundations of Intelligent Systems. LNAI, vol. 12117, pp. 308–317. Springer, Berlin, Heidelberg (2020)

15. Gleyzes, M.A., Perret, L., Kubik, P.: Pleiades system architecture and main performances. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences **39**, 537–542 (2012)
16. Grieshaber, F.: Epigraphic database heidelberg–data reuse options (2019)
17. Gruber, E., Matsumoto, M., Uleberg, E.: Linked open data for numismatic library, archive and museum integration. In: CAA2016. Oceans of Data. Proceedings of the 44th Conference on Computer Applications and Quantitative Methods in Archaeology (Oxford 2018). pp. 55–62 (2018)
18. Haslhofer, B., Isaac, A., Simon, R.: Knowledge graphs in the libraries and digital humanities domain. In: Sakr, S., Zomaya, A.Y. (eds.) Encyclopedia of Big Data Technologies. Springer (2019). https://doi.org/10.1007/978-3-319-63962-8_291-1
19. IFLA Study Group on the FRBR: Functional requirements for bibliographic records - final report. Tech. rep., International Federation of Library Associations and Institutions (IFLA) (February 2009)
20. Kansa, E., Kansa, S.W.: Digital data and data literacy in archaeology now and in the new decade. Advances in Archaeological Practice **9**(1), 81–85 (2021)
21. Lodi, G., Asprino, L., Nuzzolese, A.G., Presutti, V., Gangemi, A., Recupero, D.R., Veninata, C., Orsini, A.: Semantic web for cultural heritage valorisation. In: Data Analytics in Digital Humanities, pp. 3–37. Springer International Publishing, Cham (2017)
22. Lohmann, S., Link, V., Marbach, E., Negru, S.: Webvowl: Web-based visualization of ontologies. In: Knowledge Engineering and Knowledge Management: EKAW 2014 Satellite Events, VISUAL, EKM1, and ARCOE-Logic, Linköping, Sweden, November 24-28, 2014. Revised Selected Papers. 19. pp. 154–158. Springer (2015)
23. McGillivray, B., Cassotti, P., Basile, P., Di Pierro, D., Ferilli, S.: Using graph databases for historical language data: Challenges and opportunities. vol. 3365, p. 88 – 96. CEUR-WS (2023)
24. Pavlov, D., Emelyanov, Y., Mouromtsev, D., Morozov, A., Razdyakonov, D., Belyaeva, O.: Ontodia. org-a simple cognitive service to fill the gap in linked open data management tools. ISWC 2016 (2016)
25. Meroño Peñuela, A., Boer, V., Erp, M., Zijdeman, R., Mourits, R., Melder, W., Rijpma, A., Schalk, R.: Clariah: Enabling interoperability between humanities disciplines with ontologies. In: Applications and Practices in Ontology Design, Extraction, and Reasoning, Studies on the Semantic Web, vol. 49, pp. 73–90. IOS Press (11 2020)
26. Riva, P., Le Boeuf, P., Žumer, M.: Ifla library reference model – a conceptual model for bibliographic information. Tech. rep., International Federation of Library Associations and Institutions (IFLA) (December 2017)
27. Robinson, I., Webber, J., Eifrem, E.: Graph Databases. O'Reilly Media, 2nd edn. (2015)
28. Žumer, M., Riva, P.: Ifla lrm - finally here. In: Proceedings of the 2017 International Conference on Dublin Core and Metadata Applications. p. 13–23. DCMI'17, Dublin Core Metadata Initiative (2017)