Knowledge Graphs for the Web Economy: The CHiPS&BITS Project on Cultural Heritage

Stefano Ferilli[1,*,†], Eleonora Bernasconi[1,†], Domenico Redavid[1,†] and Giorgio Maria Di Nunzio[2,†]

[1]University of Bari, Bari, Italy
[2]University of Padua, Padua, Italy

Abstract

The growing demand for Cultural Heritage fruition is making it a major economic driver, also in connection with its tourism-related aspects. The current solutions available on the Web are still unable to provide satisfactory support to the various kinds of stakeholders and to the different applications. The History of Computing, as a peculiar, relevant and currently underinvestigated branch of Cultural Heritage, raises additional challenges and provides new opportunities, also in connection with a significant economic flow it generates. The variety and complexity of issues connected to this domain call for even more advanced solutions. In this paper we introduce the CHIPS&BITS project, which tackles these problems using a knowledge-based approach and leverages a novel framework that can meet the needs associated to this specific domain in a better way compared to standard Semantic Web approaches. We describe the contributions of the project to overcome the limitations of the state of the art, and focus on some of its more peculiar and original features, among which the definition of suitable ontologies to describe the complexity of the domain and the use of Artificial Intelligence algorithms for giving the users an advanced and personalized experience.

Keywords

Knowledge Graphs, Semantic Web, Web Economy, Cultural Heritage

1. Introduction

The demand for Cultural Heritage (CH) fruition is becoming larger and larger in the last years, making it a major component in the economic balance of Nations and territories. Not only it generates direct revenues through the fees for accessing cultural items; it also produces wealth at large considering the economic value for associated tourism, thanks to the travel, accommodation and subsistence expenditure the interested people incur. It often has additional connections to the local art, traditions and folklore, that the tourists like to enjoy once they are in the destination places. All this business, each aspect of which poses its own issues and complexities, is nowadays supported by computer applications, but only at a very superficial level: different platforms must be used to book travel, accommodation, meals and tickets, and the various aspects and items are handled independently of each other, while from the tourist's perspective there is one overall and overarching experience in which every piece is strictly connected to all the others, determining them and being in turn affected from them.

So, it is fundamental to tackle this landscape in a holistic way, in order to provide the final user with an approach that resembles more closely his perspective and can help him in a more effective way, also boosting the associated economic value as a side effect. This objective requires connecting, combining and integrating all the pieces of information from the different aspects, which is what we call 'knowledge'. Hence, our knowledge-based approach to the task. Specifically, we propose a Knowledge

IRCDL 2025: 21st Conference on Information and Research Science Connecting to Digital and Library Science, Udine, Italy, 20-21 February 2025

*Corresponding author.
†These authors contributed equally.
stefano.ferilli@uniba.it (S. Ferilli); eleonora.bernasconi@uniba.it (E. Bernasconi); domenico.redavid1@uniba.it (D. Redavid); giorgiomaria.dinunzio@unipd.it (G. M. Di Nunzio)
0000-0003-1118-0601 (S. Ferilli); 0000-0003-3142-3084 (E. Bernasconi); 0000-0003-2196-7598 (D. Redavid); 0000-0001-9709-6392 (G. M. Di Nunzio)

Graph for storing and handling all of the above information, based on semantic technologies that ensure interoperability between different systems and platforms, make the procedures and results personalized and understandable to the final users, and enable the application of advanced Artificial Intelligence (AI) solutions to overcome the limitations of the current solutions. In particular, we propose the use of a framework we are currently developing that can support the needs of this application better with respect to standard Semantic Web solutions. Some examples are the following:

- the full LOD perspective is not very suitable for preserving privacy (e.g., in auction, rental or exchange platforms); in CHIPS&BITS, the items owned by the users, their selling price and other sensitive information is not to be disclosed unless the users give permission.
- Suppose user u owns device d, then sells it and then buys it again. This requires to set two instances of relationship *owned* between u and d, having each the initial and final date of ownership as attributes. This is not immediately feasible using the RDF representation.
- Storing the knowledge on one DB enables the use of analysis, mining, and reasoning algorithms that would not be applicable on the 'unbounded' data organization of the Semantic Web. Still, the knowledge in CHIPS&BITS can interoperate with the knowledge stored in the Semantic Web.

The specific branch of CH we approach in this paper is the history of computing, due to cultural, technological and economic motivations. Our efforts in this direction resulted in the starting of CHIPS&BITS (acronym of 'Computing Heritage Intelligent Platform for Searching & Browsing Information Technology and Science'), a self-funded project aimed at developing a knowledge-based platform for the history of computing, and advanced/innovative Web-based tools to allow its users to effectively use and experience, search and discover its content, and generate value from it, dealing with the complexity of the subject matter. The objective is manyfold: spreading awareness of, and supporting education on, such important and still mostly ignored branch of CH; collecting and preserving knowledge and materials that are at stake of being lost forever, and supporting all the different stakeholders interested in this topic in their activities, including those generating economic value.

CHIPS&BITS proposes innovative solutions and tools for all the above problems. To the best of our knowledge, it would be the first platform on the history of computing:

- where stakeholders can put, interrelate, exchange, search information on the history of computing both from a scholarly perspective and for fostering its fruition in a touristic perspective, possibly generating money exchange and job opportunities as a side effect;
- featuring a unified and coherent knowledge base and repository collecting information and documents on the history of computing;
- featuring a data schema/ontology that can describe this field in all of its complexity, seamlessly dealing with the different facets of computing: hardware, software, documentation, intangible aspects, economic exploitation;
- exploiting advanced AI solutions to ensure more comprehensive, accurate, diverse and personalized results to its users.

Since the proposed approach is general, it may act as a pilot for other branches of CH.

This paper is organized as follows. After discussing the motivations and contributions of this work (Section 2), the ontology for the history of computing is presented in Section 3. Then, Section 4 briefly describes the features and interactive interface of GraphBRAIN, and its current content. Finally, after discussing some related work (Section 5), Section 6 concludes the paper and outlines future work issues.

2. Motivations & Contribution

The CHIPS&BITS project aims at contributing under several aspects, as described in the following sections. We summarized the motivations and the specific key points for each aspect in Table 1.

Table 1
Summary of Motivations & Contributions of CHIPS&BITS

Aspect	Key Points
Cultural Heritage	• Computing's rapid advancement makes it a significant cultural heritage item. • Urgency to document and preserve vintage computing knowledge while contributors are still active. • Collaborative tools to build a knowledge base which align with Open/Citizen Science principles.
Domain Complexity	• Interdependency of hardware, software, documentation, and intangible factors. • Need for a holistic ontology to capture and manage the complexity of computing heritage. • Current cataloging standards inadequate for the field's uniqueness.
Web Economy	• Increasing vintage computing market with high-value exchanges and auctions. • Crowdfunding and events generating economic flow and public interest. • Enhancing awareness, connections, and personalized stakeholder support.
Technology	• Ontology-driven knowledge representation to manage domain complexity. • Proprietary database (GraphBRAIN) ensures data protection and flexible access control. • Advanced AI techniques enable user-friendly and diverse knowledge delivery.

2.1. Cultural Heritage

While Computer Science and Engineering are quite young areas in the landscape of human knowledge, their incredibly rapid advances in the last decades, and the relevance computers gained in every aspect of our lives, recently raised significant interest in the study of the history of computing and in the preservation of knowledge and artifacts related to it. So, computing is now not just a means to support and foster all activities related to CH (known as *Digital Cultural Heritage) [1]*, but also it is becoming the very object of study and fruition as CH items, as well. Museums and private collectors started popping up all around the world. Such an ongoing excitement for vintage computing is recently pushing the discovery, retrieval and preservation of huge amounts of precious information on this topic. Many people that by first-hand contributed to the development and to the milestones of the field are still alive, and can be included in the loop, which makes this is an unprecedented opportunity, that probably will soon vanish and never happen again, to collect, safely store and sensibly organize all this wealth of information, so that it may be made available to future generations for research, study or education purposes. Unfortunately, there are some obstacles that may tamper this objective. First, the field is so new that there is no established research and scholarship yet. Second, knowledge in this field is spread across many people, each of which has just a partial (incomplete and biased) view of the whole story. What is worse, technology in this field suffers from extremely rapid obsolescence, and thus new practitioners tend to ignore important technological information needed to understand and handle items of just a few years earlier. Hence, the need for education about the history of computing and the way in which vintage hardware and software worked.

Leveraging the enthusiasm of practitioners in this field, CHIPS&BITS provides a (set of) tool(s) for collaborative building and enrichment of a knowledge base covering all aspects of the history of computing from personal, scientific, industrial sources. In line with the EU idea of Open/Citizen Science[1], the idea is that putting together all the 'partial' pieces, the final picture will be more complete, more understandable and, ultimately, more valuable than the simple sum of the pieces. Every new piece of information opens new knowledge paths, supporting new ways for discovering information and completeness, accuracy and diversity of the search results. Also, comparing different sources it can spot controversial cases and support trustworthiness.

[1]https://www.fosteropenscience.eu/foster-taxonomy/open-science-definition

2.2. Domain Complexity

The production of computing devices is much more complex than other kinds of manufactures. First of all, it inextricably connects to each other aspects that are usually in the realm of different communities, using different platforms and standards: hardware (museums), software (Internet repositories), documentation (libraries and archives), and intangible aspects (people). Indeed, hardware is just dead matter if software does not give it life, and software is meaningless and useless if detached from the hardware platform it is intended for; both require documentation to understand and properly run them; finally, why hardware, software and documents are as they are can often be understood only in connection with events, customs, anecdotes, etc. that (sometimes fortuitously) determined a course of history rather than another. The development of these objects involves many contributors, at different levels and with different roles. Even considering hardware and software only, complexity is taken to the extreme, and stands apart from all other CH items (see [2] for a more comprehensive discussion of this). Things get even worse if additional perspectives are to be considered. For example, in order to support research, education, fruition, preservation/restoration, commercial and touristic exploitation, and, more in general, spread of awareness about the topic, one cannot ignore all the context-related information (people, companies, intellectual works, records and firsts, historical events, etc.) that provide background to the items, often explain and justify them, or connect them to the more general knowledge. This additional knowledge should be connected to the strictly technical and historical part of the representation.

The current cataloguing and description standards proposed for CH items, even those specifically developed for technological items, are totally unable to handle all this complexity. Hence, a pressing need for the definition of a specific new scheme, to be shared and reused by all the stakeholders involved in this area of interest. However, this is a challenging task, because there is no standard, nor precise categorization taxonomy for computing elements, which makes it difficult to define the classes, assign them stable attributes, and organizing them into a hierarchy.

In this landscape, CHIPS&BITS contributes by adopting a knowledge-based approach and defining a 'holistic' ontology, to be used as a data schema, bringing together all these aspects.

2.3. Web Economy

There is a lively market for vintage computing elements and memorabilia, with exchanges, sales and auctions ongoing all over the world. Some items were sold for values similar to those of works of art by famous artists (e.g., one of the surviving units of Apple I, the first computer developed by Apple, was sold for $ 375,000; one of the prototypes of Commodore 65 — a computer which was never released —, endowed with a — probably unique — expansion board, was sold for about $ 95,000; some of the first models of personal computers, produced in thousands units, are sold for several thousand dollars, and so on). This market relies on standard, non-domain-aware platforms, providing the users (sellers and buyers) absolutely no support to understand and put in perspective the items they are exchanging.

Dozens crowdfunding campaigns have been successfully run aimed at reproducing old computers (or building new ones based on obsolete technology), developing new software for vintage machines, writing books on the history of computing, producing films and documentaries, establishing museums, etc. The number of supporters is in the order of thousand people, and the money collected for each of these campaigns is typically in the order of tens of thousands of euros, often ending up in the hundreds of thousands. Also, many kinds of events (conferences, seminars, shows) on 'vintage computing' have been organized, attracting scholars, researchers, amateurs, collectors or other kinds of enthusiasts. This generates tourism, and associated economic flow, as a side effect. Finally, additional economic flow is associated to job opportunities in the form of rental of items, service or restoration, consultancies, and education.

CHIPS&BITS aims at supporting all these activities by providing the stored knowledge to the various kinds of stakeholders involved, making them more aware, boosting the connections and exchanges, and helping to match requests and offers, all in a personalized way.

2.4. Technology

Supporting the previous needs also poses technological challenges that require new solutions to overcome the limitations of the current state-of-the-art.

A collaborative approach in which many people, with different expertise, culture, background and perspective contribute small pieces that together make up the big picture, requires powerful and shared representation and management solutions to organize the knowledge.

CHIPS&BITS adopts a knowledge-based approach based on ontologies as schemas for representing and interconnecting all the different kinds of data. The integration of different ontologies describing the different aspects allows to deal with the complexity and peculiarities of the field. A valuable result by itself, it enforces qualitative approaches, more focused on concepts than on statistics. The ontology itself will be searchable knowledge, useful for understanding the domain. Since CH data may have Intellectual Property-related restrictions, and their economic exploitation may involve private or sensitive information, CHIPS&BITS guarantees data protection, overcoming some limitations of both the collaborative approach and the LOD perspective, while still supporting systems' interoperability. This is obtained by the use of GraphBRAIN, a framework that is based on a proprietary DB to store the data, so that the data owners may decide what, when, how and to whom can be disclosed. The portion meant to be public can be published as LOD, and transparently be perceived as such by standard Semantic Web technologies. In particular, GraphBRAIN was extended to add flexibility to the data visibility settings, by which the data owner may decide at several levels of granularity who is guaranteed access to each kind of information stored in the knowledge base. Finally, the fruition of such a varied and complex knowledge to serve the needs of different kinds of stakeholders and applications, including research, hobby, trade, and education, requires the system to ensure user-friendliness, personalization, diversity, accuracy, and comprehensiveness in knowledge delivery.

CHIPS&BITS uses advanced AI solutions based on Natural Language Processing (e.g., for expressing in natural language the formal knowledge extracted from the graph), Document Image Analysis (to auatomatically extract knowledge from existing documents), Knowledge Representation and Reasoning (to obtain implicit information in an explainable way), Network Analysis (to understand the relevance and possible connection of knowledge items), and Machine Learning/Data Mining (to learn new knowledge from the knowledge base or profiles from users' behaviors) to extract and deliver relevant knowledge by fulfilling all of these requirements. The use of qualitative approaches (not affected by frequency and explainable), in addition to quantitative ones, enforces diversity in the outcomes and trustworthiness of the system and of its outcomes.

3. Technical Details

This section provides details about some of the technical solutions embedded in CHIPS&BITS. A graphical overview of the framework is shown in Figure 1.

3.1. GraphBRAIN Framework

GraphBRAIN is a general-purpose knowledge base management system aimed at covering all stages and tasks in the lifecycle of a knowledge base, including knowledge acquisition, organization, and (personalized) fruition[2]. It adopts the Labeled Property Graph (LPG) data model, where nodes (representing individuals) and arcs (representing relationships) may have labels (usually expressing their type) and associated attribute-value pairs, and uses the Neo4j [3] DBMS. Neo4j is schema-less, which ensures great flexibility but does not allow to associate a clear semantics to the graph items. For this reason, GraphBRAIN requires its users to work according to pre-specified data schemes, expressed in the form of ontologies. Thus, a characterizing feature of GraphBRAIN is its bringing to cooperation a database management system for efficiently handling, mining and browsing the individuals, with an ontology level that allows it to carry out formal reasoning on the knowledge. GraphBRAIN can apply several

[2]A demo of the system can be found at `http://193.204.187.73:8088/GraphBRAIN/`

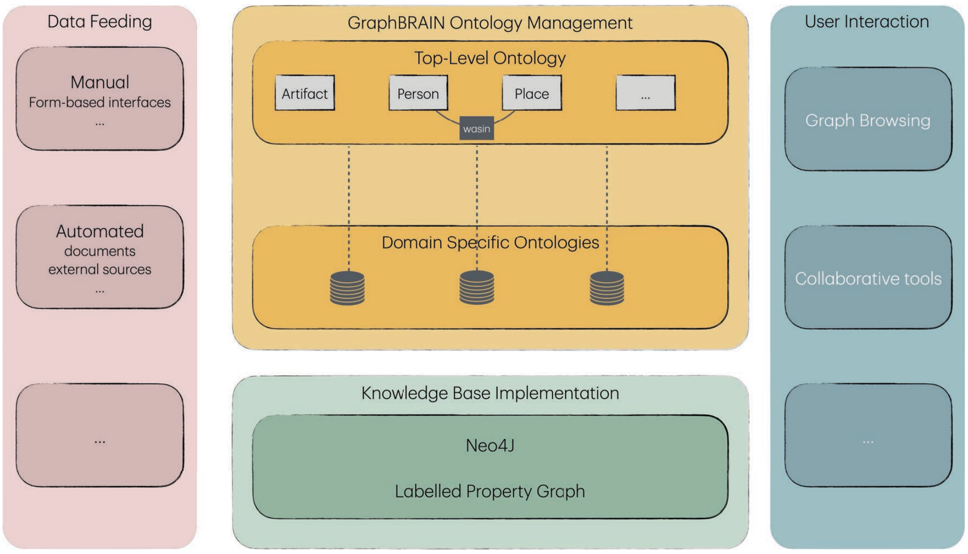

Figure 1: Overview of the GraphBRAIN framework

ontologies on the same graph, representing different perspectives on the same knowledge. The classes shared by different ontologies allow the system to connect knowledge across domains: their individuals act as bridges, allowing the users of a domain to reach information coming from other domains. The ontologies are expressed in a proprietary format, specifically tailored for LPGs [4]. Both ontologies and instances may be imported from, or exported to, the standard Semantic Web format Ontology Web Language (OWL)[3], in order to support their interoperability and reuse as linked open data (LOD) [5].

GraphBRAIN provides a top-level ontology defining very general and highly reusable concepts and relationships (e.g., Person, Place; Person.wasIn.Place; etc.). This top-level ontology plays a crucial role to interconnect the domain-specific ontologies, ensuring an overall connected knowledge graph. Using a suitable tool, GraphBRAIN administrators may create, build and maintain additional ontologies.

After setting up the ontologies, information can be fed into the knowledge base. A form-based interface allows users to manually insert/update/remove instances or to query the knowledge base for instances of entities and relationships: they must select one of the available domains/schemas, and the forms are automatically generated by the system starting from the corresponding ontologies. The knowledge base can also be fed by automatic knowledge extraction from documents and other kinds of resources (e.g., books or the Internet). Instances may also have attachments, making GraphBRAIN a digital library, whose content is organized according to formal ontologies, fostering interoperability with other systems. Users may add, show, or delete attachments.

In a collaborative spirit, users may add comments on (to provide suggestions or add information), or approve/disapprove, each entity or relationship instance, each single attribute value thereof, and even the ontological items. This feedback is used to assign a trust value to the users.

Another interface allows users to display a portion of the graph, browse it interactively and display detailed information about entity and relationship instances. This allows the user to continue his search in a less structured way, by exploring the graph (expanding or compressing node neighbors) with a predefined goal in mind, or letting the data themselves drive the search, possibly finding relevant information in a serendipitous way.

The users can also run on the available knowledge several algorithms for graph mining, network analysis, information extraction, automated reasoning, natural language processing, etc. Some of these algorithms are reused from the literature; others are original. Currently included functions are:

- assess relevance of nodes and arcs in the graph, and extract the most relevant ones;
- extract a portion of the graph that is relevant to some specified starting nodes;

[3]http://www.w3c.org/owl

- extract frequent patterns and associated sub-graphs;
- predict possible links between nodes;
- recommend relevant knowledge items;
- translate into natural language the information content of a portion of the graph.

If available, a user profile can be used to personalize the results of all these algorithms. This would ensure that each user obtains tailored information, which is another novelty introduced by GraphBRAIN.

While a proprietary technology, all the functions of the GraphBRAIN framework are exposed as services of an API whose code will be released for free use, so that any third-party application and organization will be able to exploit it.

3.2. Holistic Ontology

Currently the CHIPS&BITS knowledge graph is organized according to a merger of the following ontologies:

general dealing with very general concepts and relationships that are expected to be present in almost all domains; of interest to CHIPS&BITS are, e.g., entities `Agent` (including persons, organizations and users), `Place` (geographical, administrative, locations, etc.), `Event` (both historical ones and conferences, fairs, shows, etc.), `Collection` (any kind of grouping), `Document` (in its most general definition as "something that serves as evidence or proof", and thus not limited to printed (or printable) documents), `IntellectualWork` (the original result of an intellectual effort, such as inventions, subjects, works of art, etc.), `Award` (any kind of recognition that can be awarded to, or record that can be marked by, persons, companies, devices, documents, or components: e.g., Education achievements, Prizes, Records), `Item` (a specific, identifiable specimen of a serially produced object).

lam concerning libraries, archives and museums; it includes and is aligned to the standard data models from IFLA (the International Federation of Library Associations) [6], is connected to the general ontology via concepts such as `Document`, `Collection` (for series and libraries), `Item`, and `IntellectualWork`, and provides CHIPS&BITS with many representational components for documentation.

education describing educational aspects such as subjects, learning resources, learning paths, learning accomplishments, etc.; it includes and is aligned to standard data models such as the IEEE LOM (Learning Object Metadata), IntelLEO (an ontology on educational procedures developed within a European research project), and OERschema (an ontology on educational organization at its early stages of development) [7]; it is also connected to the lam ontology via the concepts concerning documents, and obviously to the general ontology via various high-level concepts.

computing The core of CHIPS&BITS, for which more details will be provided below.

tourism concerning touristic-related information, including history, cultural heritage items, points of interest, transportation, hospitality, logistics and services; it was borrowed from another ongoing effort [8] in which GraphBRAIN will underlie an integrated system aimed at supporting all stakeholders involved in touristic activities (tourists, entrepreneurs and institutions).

food Complements the tourism section by adding information about food&beverage, especially from the perspective of typical dishes and beverages from specific regions of touristic interest.

These shared elements among these ontologies allow CHIPS&BITS to relate knowledge items from different domains, extending in this way the available scope of search beyond the single perspectives. For example, a tourist interested in the history of computing, while in Bari, might be spotted the chance to visit the collection at the Department of Computer Science, in order to see a specimen of the Olivetti Programma 101 computer.

A sample of domain-specific classes from the computing ontology is the following:

- `ElectronicComponent`: a part, useful or needed to build a *Device* but not operable directly by the final user (e.g., Printed Circuit Boards, Integrated Circuits).
- `Device`: a manufact that can be operated directly from the final user (it includes Calculators, several kinds of Computers, Input and Output Devices, StorageDevices, etc.).
- `Software` (of various kinds: Development, Educational, Embedded, Office Automation, Operating Systems, Videogames, ...).
- `Package`: a specific packaging of a *Device* or of a set of devices sold together (often very relevant for collectors);
- `System`: a group of *Device*s that is functional only as a whole; it differs from a *Configuration* in that, in a *Configuration*, at least one of the *Device*s would be functional if taken alone.
- `Configuration`: a relevant group of *Device*s, relevant because typical or determined in order to satisfy specific needs (e.g., a configuration of devices for desktop publishing);

Also useful are the concepts of `Award` (that may be useful to note the firsts in the history of computing), `IntellectualWork` (extended to account for algorithms, approaches, inventions, programming languages, technologies, theorems, theoretical models), `Item` (extended from just documents to components, devices, software, etc.), borrowed from the other ontologies.

A sample of domain-specific relationships are `Agent.acquired.Item` (including donations and purchases), `clones` `compatibleWith` and `mayReplace` (between pairs of components, devices or software), owned (for agents who owned organizations, but also to record notable previous owners of some items) `Component.partOf.{Item,Device}` (some rare parts may dramatically change the value of an item or device), `evolves` (for hardware and software versions), `requires` (software may require another software or a device to work properly), `repaired` (to record who is able to repair what). Some of these are relevant for the economic aspects, others are relevant for repairing and restoration, which is extremely important because computing devices has peculiarities that cannot be expressed in existing ontologies designed for other kinds of CH.

The proposed conceptualization may be a starting point for the definition of cataloguing standards for cultural heritage material related to the history of computing. Indeed, existing standards for cultural heritage, even those developed for technological and scientific elements (e.g., [9]), are totally unable to express the complexity and subtleties of this specific field.

3.3. Personalization and Privacy

Personalization and Privacy are handled in CHIPS&BITS with a system of registered users and with the help of a (relational) database, which was inherited and extended from the GraphBRAIN framework.

The user profile takes the form of a set of weights, associated to the ontological elements (entity, relationship or attribute) or to specific nodes or arcs (i.e., entity or relationship instances). The weight is formed and continuously updated by taking into account both explicit preference indications by the users and implicit preferences computed on usage. Many kinds of user interaction contribute to the latter part: selection of an entity or relationship, access to a specific node, arc or attribute (to display its value or to add, modify or delete it), approval/disapproval or comment on a node, arc or attribute value. These parameters are stored in 3 tables of the relational DB: Preferences for explicit preferences, Interactions for visualization and CRUD operations, Evaluations for approvals/disapprovals and comments. The profile is used to highlight the user changes in the knowledge associated to his preferred elements, or to drive knowledge extraction and graph mining algorithms so that they can focus more on his preferred elements and less on the others.

While designed according to the linked data perspective, the GraphBRAIN knowledge graph is not available in its entirety as Linked Open Data. Indeed, it is not directly accessible to the public. Access is available only through the query, graph browsing and knowledge extraction facilities provided by an API, that any interested application must use. Each function in the API, exposed as service, returns relevant portions of the graph based on the input parameters, on the user's profile (if available), and on the privacy specifications for the various knowledge items. Privacy is obtained by adding to each

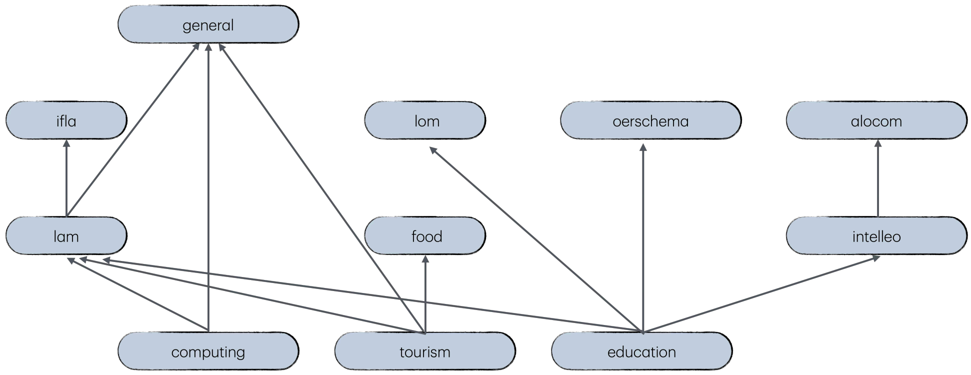

Figure 2: Dependencies among ontologies involved in CHIPS&BITS

ontological element a *privacy* attribute, which can be set to True or False. When True, the owner of a knowledge item (an entity or relationship instance, i.e., a node or arc in the graph) can set its privacy value to one of the values Private (visible only to its owner), Restricted (only selected users can access it), or Public (visible to all users and publishable as open data). This information is stored in a Privacy table in the relational DB. When Restricted, another table Permission in the DB reports, for each node or arc as a whole, or for its labels and attributes, which users may access it, specifying for each of them if they can see, add, modify or delete it (in case of a node or of an arc) or its value (in the case of an attribute).

4. Current Status of the Project

CHIPS&BITS currently involves 11 ontologies, describing various contextual or application domains related to the history of computing (general, libraries/archives/museums, education, tourism). The dependencies among such ontologies are graphically shown in Figure 2, where lower-level ones import the concepts from the higher-level ones. The figure shows that ontologies purposely defined in GraphBRAIN all import the 'general' ontology, while ontologies obtained by translating existing schemes are stand-alone. However, a connection still exists between these classes and the other ones, because in the translation an alignment was carried out, and the same names were used for the classes referring to the same concept.

Table 2 reports statistics on the number of ontological elements (classes, relationships, attributes) in these ontologies. The number of top classes is reported separately from the number of their subclasses, in order to distinguish the specializations from the core concepts. Since the model underlying GraphBRAIN is an LPG, also relationships may have attributes. Note that in GraphBRAIN one relationship name may represent many specific relationships, distinguished by different subject-object pairs (which would require different names in OWL). So, the 'Relationships' column reports the number of different relationship names followed, in parentheses, by the number of actual relationships. Considering both its own elements and those inherited by the imported ontologies, the core ontology for the history of computing currently includes 267 classes and 540 relationships.

Population of the CHIPS&BITS knowledge graph has already started, based on the current version of the ontology. Concerning the specific domain, the catalogues of a private collection and of the Museum of Computer Science at UniBA are being used as sources of knowledge, entered partly manually and partly using an upload procedure from their catalogues. These two collections were chosen because their owners and maintainers share our holistic view, according to which a seamless integration among hardware, software, documentation and intangible aspects. The private collection includes about 500 units between devices and rare components, plus hundreds of books, magazine issues, videos, software packages and other memorabilia. The Museum of Computer Science has on show a mainframe, several minicomputers and terminals, many personal computers and workstations, plus many items in the deposits and a huge collection of several thousand historical books, magazines and software, a large

Table 2
Statistics on the elements in the ontologies involved in CHIPS&BITS

Domain	Ontology	Classes+Subclasses	Attributes	Relationships	Attributes
Top-level	general	13+128	107	33 (157)	43
Libraries, Archives & Museums	ifla	12+60	128	23 (77)	22
	lam	7+45	88	32 (143)	26
Education	lom	12+43	88	14 (22)	11
	oerschema	6+14	15	11 (31)	2
	alocom	5+35	6	5 (8)	0
	intelleo	33+102	156	33 (101)	6
	education	4+16	6	17 (35)	26
Tourism & Traditions	food	8+21	13	13 (34)	3
	tourism	4+118	29	7 (104)	12
Core	computing	7+101	66	22 (81)	33

part of which is still to be catalogued.

Other instances uploaded in the knowledge base concern the contextual part, expressed by the general ontology. For example, instances of class **Place** are being uploaded from an atlas. Classes **Category** and *Word* were populated from WordNet [10, 11], together with the associated relationship instances, from the standard part of the Dewey Decimal Classification (DDC) system [12], and from the ACM Computing Taxonomy. These instances can be linked to individuals of other classes (e.g., documents, persons, places) and used as semantic or lexical tags (respectively) to express information about them (e.g., 'Alan Turing' might be linked with 'Computer Science', 'World War II', etc.). This will be another important way to contextualize the knowledge items and support personalization of behavior and outcomes.

An automated tool is also under development that can fill entity, relationship and attribute instances by extracting them from the text of books, leveraging Large Language Models technology.

The current content of the CHIPS&BITS knowledge base consists of 347295 entity instances, of which 1858 related to the computing domain and 4822 from the libraries/archives/museum domain, and 641293 relationship instances. The education and tourism sections were started recently, so they include just a few hundred instances. Conversely, the general ontology contributes with a few hundred thousand instances, thanks to the batch ingestion of various resources: the atlas for places, WordNet for words and concepts, and some subject taxonomies.

Let us now report some use cases we ran on the knowledge base. Figure 3 shows a subgraph extracted from the overall knowledge based on starting nodes that are relevant to the user's information need (shown with thicker borders in the graph and listed on the left-hand side of the interface). The subgraph extraction algorithm is aware of the user's profile and uses it to drive the extraction. Most of the functions used in the following use cases rely on a preliminary subgraph extraction like this, which is the starting point for carrying out the different tasks described.

Use Case 1 User X is a collector, with broad interests but a special focus on personal computing and specifically the Commodore brand. He misses a few models from Commodore, and the system knows that another user, Y, owns two units of one of these models. The system prompts Y in case he would like to sell one of his units to X, or exchange it with one of X's units. He accepts, and the system establishes a contact between X and Y, still hiding their real identities for privacy reasons, until they decide they can disclose their identities to each other.

Use Case 2 Institution X owns a unit of a computer model, but knows almost nothing about it. The system extracts relevant knowledge in the form of a subgraph centered in that model, and returns a natural language description of the model, its history and relevance, and its connections to other models, hardware and software. The system also retrieves other users who own(ed) or use(d) that model, so that they may be contacted for more information. Since the unit is not working, the system also provides

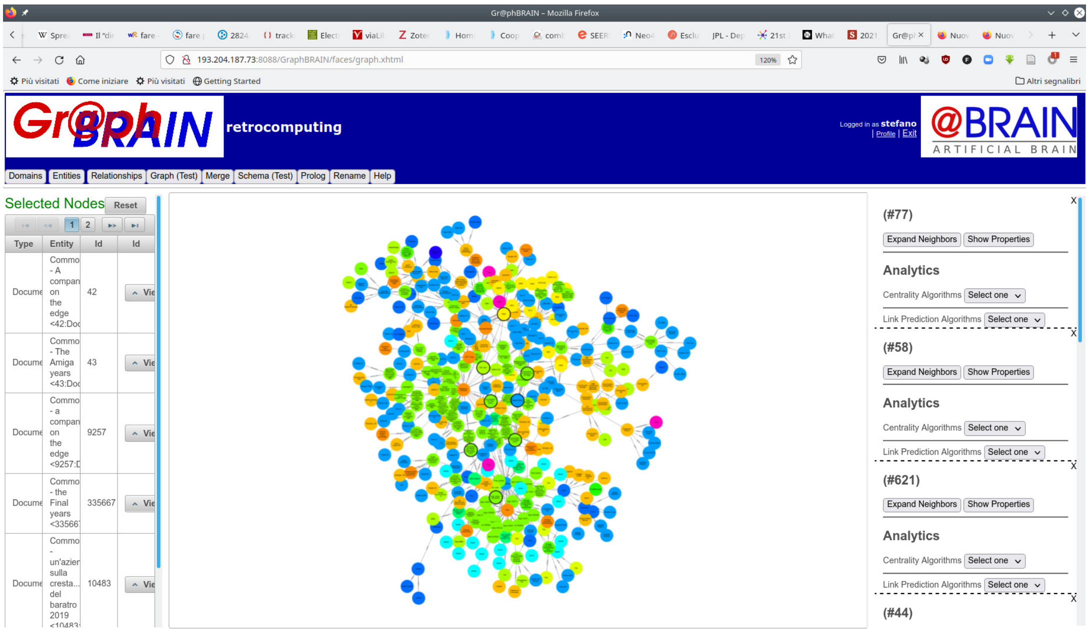

Figure 3: A sample subgraph from the computing knowledge graph

information on how to fix it, and points to users who can repair or restore it. It also determines the value of the unit based on previous transactions, and retrieves users potentially interested in purchasing the unit in case the institution is not interested in it.

Use Case 3 Researcher X, interested in peripherals, comes to know about a company he was not aware of, that had a role in the history of computing and is relevant to his research. Actually, the system itself pointed X to that company, based on X's interests. The system delivers a report on that company, describing its history, products and achievements, the people who worked for it and their roles and tasks, all from the specific perspective of peripherals (albeit the company may also have produced computers and software).

Use Case 4 Association X is organizing an exhibition about a pioneer of computer science and engineering, Y. The system finds the set of devices and software he developed, the companies he worked for, his history and some anecdotes that might make the exhibition more interesting. It also points X to users who own and might rent units of the devices and software developed by Y, to documents, audios and videos involving Y (with information on where they deal with Y), to other relevant memorabilia, and to previous events about Y.

Use Case 5 User X is interested in the Olivetti brand, and specifically in the firsts that it marked in the computing history. The system identifies the models developed by Olivetti which are landmarks, provides an explanation for the choice, identifies museums or exhibitions where units of these models are on show, and provides a description of touristic points of interests, typical food and folklore of the regions where such units are hosted, taking into account X's interests and preferences.

5. Related Work

Work on Cultural Heritage knowledge representation is more oriented toward data schemas and vocabularies rather than ontologies. However, there are some noteworthy initiatives also about the development of ontologies, and associated knowledge graphs, for CH. A relevant initiative in this direction is supported by the Italian Ministry of Culture: ArCo (Architecture of Knowledge)[4], is an ontology for, and a knowledge graph of, Italian CH. It currently reuses, and is aligned to standard CH-related schemes and ontologies:, CIDOC-CRM, EDM, Cultural-ON, and OntoPiA. The resulting knowledge graph includes, and provides as LOD, data from records of the General Catalogue of Cultural Heritage, a database of Italian cultural heritage entities. Like ArCo, CHIPS&BITS organizes several ontologies in a network for modeling different kinds of cultural properties and their corresponding catalog records. Some branches of knowledge described by ArCo overlap with those of CHIPS&BITS (e.g., Archive, Catalogue & catalogue records, Cultural events & exhibitions). In principle, any part of ArCo (and of the ontologies it includes) can be reused in CHIPS&BITS, as well as content might be included from many resources such as Geonames, Wikidata, VIAF, etc. Describing in detail what is or might be reused from these and other existing resources in CHIPS&BITS is out of the scope of this paper, where the aim is just to say that this can be done and that interoperability with them is supported.

Related specifically to the computing domain, [13] applied knowledge-based approaches to an open dataset with persons and relationships extracted from the official biography of Steve Jobs and the 1999 film Pirates of Silicon Valley. CHIPS&BITS aims at being much wider and more comprehensive than this initiative, and at having impact on practical aspects of the discipline.

From the technological viewpoint, some works exist that analyze the possibilities for cooperation between ontologies and graph DBs. [14] discusses technical issues that might limit the impact of symbolic Knowledge Representation on the Knowledge Graph area, and summarizes some developments towards addressing them in various logics. Most works tried to merge research on RDF and LPG knowledge representations, but always giving the RDF perspective priority and predominance. GraphBRAIN was the first to push for a native LPG-oriented approach [4]. After the publication of GraphBRAIN, other initiatives investigated the possibility of developing suitable schemas for LPGs specifically [15].

6. Conclusions and Future Work

Computing is starting to be considered par of Cultural Heritage, not only a means, due to the rapid obsolescence and advancements in technology calling for preservation. Luckily, we still have the opportunity to get knowledge from the pioneers who are still alive. The CHIPS&BITS project's aim is to collect, interrelate, preserve and deliver knowledge on the history of computing. The peculiarities and complexity of the domain are such that all available data models are inadequate for it. We proposed a new data schema in the form of an ontology, that enables the use of advanced AI solutions, and might be the core of a new cataloguing standard. This domain is also generating significant economic flow (due to publishing and educational initiatives, collectors' exchanges, event organization and job opportunities): CHIPS&BITS aims at leveraging this in the perspective of the Web economy. CHIPS&BITS proposes flexible solutions to provide tailored behaviors for the variety of potential users and stakeholders involved, and guarantees privacy for sensitive information associated to the economic exploitation. Most of the proposed solutions are general and applicable to any kind of CH.

Population of the knowledge base has started, and other work is currently undergoing or planned in several directions: refining and extending the ontologies, populating the knowledge base, developing systems for automated population of the knowledge base from existing information sources, and the development of tools that can serve the needs of the various kinds of stakeholders. In particular, the design of a platform for providing professional support to enthusiasts and collectors (e.g., e-commerce and auction functions, matching of offers and requests for items and for competences, etc.).

Acknowledgments

This research was partially supported by projects CHANGES "Cultural Heritage Active Innovation for Sustainable Society" (PE00000020), Spoke 3 "Digital Libraries, Archives and Philology" and FAIR

[4]http://wit.istc.cnr.it/arco/?lang=en

"Future AI Research" (PE00000013), spoke 6 "Symbiotic AI", funded by the Italian Ministry of University and Research NRRP initiatives under the NextGenerationEU program.

References

[1] N. Barbuti, G. D. Felice, A. D. Zanni, P. Russo, A. Valentini, Creating Digital Culture by co-creation of Digital Cultural Heritage: the Crowddreaming living lab method, Umanistica Digitale (2020) 19–34. URL: https://umanisticadigitale.unibo.it/article/view/9956. doi:10.6092/issn.2532-8816/9956, number: 9.

[2] S. Ferilli, D. Redavid, 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, volume 2375 of *Central Europe (CEUR) Workshop Proceedings*, 2019, pp. 49–60.

[3] I. Robinson, J. Webber, E. Eifrem, Graph Databases, 2nd ed., O'Reilly Media, 2015.

[4] S. Ferilli, Integration strategy and tool between formal ontology and graph database technology, Electronics 10 (2021). URL: https://www.mdpi.com/2079-9292/10/21/2616.

[5] T. Heath, C. Bizer, Linked Data: Evolving the Web into a Global Data Space, Synthesis Lectures on the Semantic Web, Morgan & Claypool Publishers, 2011.

[6] S. Ferilli, E. Bernasconi, D. Di Pierro, D. Redavid, A holistic ontology for digital libraries - the ifla-compliant core, in: Proceedings of the 2nd Workshop on Artificial Intelligence for Cultural Heritage (IAI4CH@AIIA2023), co-located with the XXII International Conference of the Italian Association for Artificial Intelligence (AI*IA 2023), volume 3536 of *CEUR Workshop Proceedings*, 2023, pp. 39–52.

[7] S. Ferilli, E. Bernasconi, D. D. Pierro, D. Redavid, L. Loop, A holistic ontology for digital libraries, in: Proceedings of the 20th Conference on Information and Research science Connecting to Digital and Library science (formerly the Italian Research Conference on Digital Libraries), Bressanone, Brixen, Italy - 22-23 February 2024, volume 3643 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2024, pp. 74–86.

[8] S. Ferilli, B. De Carolis, P. Buono, N. Di Mauro, S. Angelastro, D. Redavid, Una piattaforma intelligente per la gestione integrata del settore turistico, in: Primo Convegno Nazionale CINI sull'Intelligenza Artificiale – Workshop on AI for Cultural Heritage, 2019, p. 2 pages. In Italian.

[9] F. Vannozzi, Catalogare il patrimonio scientifico e tecnologico: da sic a sts a pst, storia di un percorso (e di una collaborazione), in: G. Pratesi, F. Vannozzi (Eds.), I valori del museo. Politiche di indirizzo e strategie di gestione, Economia e management della cultura e della creatività, Franco Angeli Edizioni, 2015, pp. 98–101. URL: https://core.ac.uk/download/pdf/53672194.pdf, in Italian.

[10] G. A. Miller, Wordnet: A lexical database for english, Communications of the ACM 38 (1995) 39–41.

[11] C. Fellbaum (Ed.), WordNet: An Electronic Lexical Database, MIT Press, 1998.

[12] M. Dewey, A classification and subject index for cataloguing and arranging the books and pamphlets of a library, Amherst, Mass., 1876.

[13] G. Drakopoulos, A. Kanavos, P. Mylonas, S. Sioutas, D. Tsolis, Towards a framework for tensor ontologies over neo4j: Representations and operations, in: 8th International Conference on Information, Intelligence, Systems & Applications, IISA 2017, Larnaca, Cyprus, August 27-30, 2017, IEEE, 2017, pp. 1–6.

[14] M. Krötzsch, Ontologies for knowledge graphs?, in: A. Artale, B. Glimm, R. Kontchakov (Eds.), Proceedings of the 30th International Workshop on Description Logics, Montpellier, France, July 18-21, 2017., volume 1879 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2017.

[15] R. Angles, A. Bonifati, S. Dumbrava, G. Fletcher, A. Green, J. Hidders, B. Li, L. Libkin, V. Marsault, W. Martens, F. Murlak, S. Plantikow, O. Savkovic, M. Schmidt, J. Sequeda, S. Staworko, D. Tomaszuk, H. Voigt, D. Vrgoc, M. Wu, D. Zivkovic, Pg-schema: Schemas for property graphs, Proc. ACM Manag. Data 1 (2023).