A Holistic Ontology for Digital Libraries

The IFLA-compliant Core

Stefano Ferilli[1,*,†], Eleonora Bernasconi[1,†], Davide Di Pierro[1,†] and Domenico Redavid[1,†]

[1]University of Bari, Bari, Italy

Abstract

The record-based approach to library information organization is obsolete and cannot support advanced opportunities provided by AI. Graph-based representations driven by ontologies are needed, but representation models proposed in this field are not compliant with those adopted by DBs. GraphBRAIN is a framework and technology that applies the ontological approach to the LPG model adopted by graph DBs, taking the best of both worlds: representational power and flexibility of ontologies and efficiency in data handling of DBs. It can be mapped onto the standard Semantic Web representations. The International Federation of Library Associations and Institutions (IFLA) proposed different conceptual models of library data that move from record-based to relational representations, but do not reach the ontological level. Conversely, attempts to tackle the ontological level are often not fully compliant with the IFLA models. This paper reports on developing an ontology based on GraphBRAIN technology that aligns the various models proposed by IFLA, fully representing their elements. It shows how it can be expanded to include elements that are not traditionally represented in library models, but that can unleash new potentiality for library practitioners, researchers and end-users.

Keywords

Digital Libraries, Knowledge Graphs, Ontologies, Graph Databases

1. Introduction

The traditional record-based approaches to library information description are nowadays insufficient and unable to fully grasp and express the complexity of such information [1, 2], both internally and externally. Internally, the main limitations (deriving from legacy paper cards) are that (i) a record is document-centric, reporting a number of information elements all considered

2nd Italian Workshop on Artificial Intelligence for Cultural Heritage (IAI4CH 2023, https://ai4ch.di.unito.it/), co-located with the 22nd International Conference of the Italian Association for Artificial Intelligence (AIxIA 2023). 6-9 November 2023, Rome, Italy

*Corresponding author.

†These authors contributed equally.

stefano.ferilli@uniba.it (S. Ferilli); eleonora.bernasconi@uniba.it (E. Bernasconi); davide.dipierro@uniba.it (D. Di Pierro); domenico.redavid1@uniba.it (D. Redavid)

http://lacam.di.uniba.it/people/ferilli.html (S. Ferilli); https://www.uniba.it/it/docenti/eleonora-bernasconi (E. Bernasconi); http://www.di.uniba.it/~redavid/ (D. Redavid)

0000-0003-1118-0601 (S. Ferilli); 0000-0003-3142-3084 (E. Bernasconi); 0000-0002-8081-3292 (D. Di Pierro); 0000-0003-2196-7598 (D. Redavid)

as ‘belonging’ to the described document and (ii) it includes a small, pre-defined and fixed, set of fields. Externally, the strictly sequential organization of records deriving from legacy files makes no more sense for computer files. Another limitation is the lack of semantics. Even most recent proposals in this direction mostly deal with the description level, without making the leap to semantic technologies, driven by ontologies, that alone can open new landscapes and enable advanced Artificial Intelligence (AI) support to the field.

Concerning internal limitations, it is necessary to distinguish the descriptive elements that are really inherent the document from those that exist on their own, and just ‘happen’ to be associated to the document for some reason. Also, the set of descriptors must become much larger, varied and flexible, allowing to connect the library items to all the wealth of knowledge directly or even indirectly related to them but precious, or sometimes crucial, for practitioners, researchers and end-users to understand the items properly. In our vision, this includes the *content* of the documents, their *materiality*, their *context* and their *lifecycle* (involving their users and their uses). Even information that is not obviously connected to the traditional library descriptions might be used to indirectly connect the documents and, ultimately, provide a deep understanding thereof. We call this a ‘holistic’ approach. As to external limitations, an upgrade to a reticular, graph-based approach, is required, where the different kinds of entities involved in such knowledge live on their own, rather than being just values in record fields, and where relationships among entities are the key feature of descriptions. This is also instrumental to enabling semantic technologies, typically based on graph structures.

This paper proposes the first core of an ontology for library information description, developed as part of the effort for task 1 of Spoke 3 “Digital Library, Archives and Philology” of the CHANGES (Cultural Heritage Active Innovation for Next-Gen Sustainable Society) project, funded by the Italian Ministry of University and Research under the National Recovery and Resilience Plan (NRRP), Enlarged Partnership 5 “Humanistic culture and cultural heritage as laboratories of innovation and creativity”, within the NextGenerationEU programme of the European Union. While the overall objective is building a ‘holistic’ ontology, aimed at representing and inter-relating all the different perspectives mentioned above, here we focus on the specific portion that refers to standard library descriptors, as defined by leading world library institutions, and that represents the core of the complete ontology.

As a peculiarity of this work, the development was carried out within the GraphBRAIN framework, which merges solutions coming from the Knowledge Representation and Reasoning branch of AI, for the ontological part, with advanced platforms developed in the DB field, for data storage. This framework is based on a different graph model than the standard ontological models adopted in the Semantic Web field, which is immediately applicable to the Neo4j graph DB and represents an intermediate format that opens the data to more varied and advanced exploitation possibilities, beyond basic ontological reasoning provided by Semantic Web technologies, including rule-based, constraint-based and MultiStrategy Reasoning, network analysis and graph mining, and interactive knowledge browsing and exploration.

The rest of this paper is organized as follows. After introducing background and related work in the next section, we describe the design process that led to the ontology, and then the structure of the ontology itself. Finally, Sec. 5 concludes the paper and outlines future work.

2. Background & Related Work

After the initial simple transposition to the digital format of library cards, with a pre-defined and fixed set of fields (or descriptive attributes), subsequent approaches to bibliographic records representation and management started leveraging the additional opportunities provided by digital technologies. Perhaps the most obvious of such opportunities is the enormously increased flexibility, that allowed defining a much larger set of fields and using in a record only a subset thereof, or using some of them in many occurrences. Outstanding examples of this are the MARC format and the ISO 2709, currently a standard in the library practice [3]. While fundamental, these are still deeply rooted in the record-based idea of library descriptions, in which the central and aggregating concept is the book or document, and every piece of information is considered as 'belonging' to the book/document.

A step forward toward relational descriptions was taken by the International Federation of Library Associations and Institutions (IFLA) with its Functional Requirements for Bibliographic Records (FRBR) [4], a conceptual model of library information proposed in 1990. It provided an insight into non-obvious technical details of bibliographic description, at the same time widening the scope of descriptions. According to the foundational ideas from relational DB design, it identified a set of entities to be represented and described (e.g., places, persons, organizations) with their own 'dignity', along with their characterizing attributes, and a set of relevant relationships among them that were implicit and hidden in traditional bibliographic records. So, the information that was enclosed in a bibliographic record is now distributed in a reticular information organization. The FRBR model has been refined and expanded in a later recommendation by IFLA, the Library Reference Model (LRM), proposed in 2017 [5, 6].

While important, these proposals still have significant limitations. First, and most importantly, they still focus on the syntactic level of descriptions, without providing formal definitions of the items (entities, relationships, attributes) and of their behavior in the library domain. This requires an upgrade from Data Bases (DBs) to Knowledge Bases (KBs)[1] (or Knowledge Graphs —KGs—, when the KBs are organized as graphs), where ontologies provide such formal definitions, enabling interoperability among systems and platforms and advanced exploitation of that information by AI techniques. Second, the existing proposals still focus on information that is directly related to the documents to be described, and specifically to their formal metadata or properties, while we call for the adoption of a 'holistic' approach.

A step by IFLA toward Semantic Web technologies is the Resource Description and Access (RDA) standard, whose main features are the extensibility of the descriptors, and the use of a formal representation that is closer to the ontological standards. Still, it cannot be considered an ontology. In this direction: [7] proposes one based on FRBR, ingesting the data from MARC sources; [8] fosters a switch to open data, based on adapting to RDF the FRBR and RDA; [9, 10] focus on FRBRized entities, and specifically on the challenges related to metadata migration and semantic enrichment of bibliographic data.

Many works in the literature highlight the advantages of moving to a semantic-based approach to Digital Libraries (DLs). Semantic DLs (SDLs) [11] extensively use metadata to support

[1] A traditional distinction is made in AI between 'data' (values), 'information' (interpreted data), and 'knowledge' (networked information, where the value of the whole is more than simply the sum of the parts' values).

information retrieval and classification tasks. They can use ontologies to organize bibliographic descriptions, represent and expose document contents, and share knowledge among users [12]. Moving from traditional descriptions, [11] investigates how the Knowledge Organization Systems (KOS) of early DLs can be integrated with the DL architectures using semantic technologies and data. Noting that the standard RiC-CM and CIDOC-CRM models used in library practice derive from the traditional approach used for the construction of catalogues, [13] proposes to connect the entities identified in LAM logical models to the much larger, richer and more numerous ontologies of the semantic web, represented by the Linked Open Data (LOD) Cloud. However, it still focuses on the elements in traditional schemes. [14] depicts the situation of Linked Data technologies applied to libraries, outlining best practices, gaps and future trends based on the information found in authoritative library websites from the last 5 years and selected national libraries worldwide. Albeit often using RDF, most works do not use fully-fledged ontologies, just controlled vocabularies and thesauri, such as: OAI-ORE (Open Archives Object Reuse and Exchange), SKOS[2], Dublin Core, Newman's[3], MOAT [15], etc. Even those that use ontologies, never propose an overall ontology from the specific DL perspective, except perhaps [12] that, by analyzing some SDL projects, identifies three application areas of ontologies, and mentions the MarcOnt Ontology [16] as a candidate for combining different metadata standards that can describe various concepts at different levels of granularity. However, the project does not exist anymore.

Some works highlight the need for an overall ontology that accommodates and 'coordinates' all the different perspectives: [17] aims at overcoming the limited semantics of thesauri and similar knowledge models proposing an automatic process to convert a knowledge model into a domain ontology through alignments with the lexical DB Wordnet [18] and the upper-level ontology DOLCE. To overcome the current limitations to the reuse of bibliographic data in the Semantic Web and of RDF languages, due to the lack of a common conceptual framework, [19] proposes a reference model and a super-ontology to overcome the misalignments between existing bibliographic ontologies and the principles and techniques of LOD. Other works propose ontologies that tackle specific aspects of digital library and bibliographic data management: [20] for information commonly found in survey and review articles; [21] (inspired by Schema.org) to describe educational resources; [22] to support free and semantic annotation (tags, notes, comments, errata, etc.) over the paper, also providing facilities for curation, provenance, authoring and versioning; [2] to deal with Open Science.

Ontologies are important not only because they allow a more structured and semantic description of DL knowledge. They also enable automated reasoning that can support researchers and practitioners. E.g., [23] shows how a Case-Based Reasoning system based on an ontology can enhance the effectiveness of information retrieval. The Cogito Intelligence API also conducts semantic reasoning [24].

Concerning the extent of descriptions, some attempts to widen it have been carried out mainly in the wider context of Cultural Heritage. Worth mentioning are CIDOC-CRM [25] and ArCo (Architecture of Knowledge) [26], an ontology and KG of Italian Cultural Heritage[4].

[2]http://www.w3.org/2004/02/skos/
[3]http://www.holygoat.co.uk/projects/tags/
[4]http://dati.beniculturali.it/

It 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[5], and OntoPiA. 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 and networked, 'holistic', set of descriptors.

Summing up, while there is a widespread feeling that ontologies are a need and an opportunity for modern and advanced DL and bibliographic data management, the existing works often propose tools, mostly aimed at improving IR. The underlying data representations are almost always stuck to standard metadata (LAM or other), with those more oriented towards relational descriptions focusing on FRBR, sometimes proposing migration of data from the former to the latter, and only occasionally perceiving the need to expand them. FRBR is only a starting point for us. We have worked to build 'holistic' ontologies that include content and context as fundamental components of the descriptions, in order to support advanced and intelligent exploitation of DL data [28, 2]. Here we work on the library-related section specifically, to obtain an ontology that expresses all the IFLA models and is aligned to our previous proposal.

3. The GraphBRAIN Framework and Technology

GraphBRAIN [29, 30] 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 provided by the RDF model, it is based on the LPG model, allowing labels and attributes on both graph nodes and arcs. This enhances its expressiveness, readability and compactness. 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 based on the features of LPGs. It is organized into 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 (symmetry, transitivity, functionality, etc.) and their attributes; define axioms in the form of logic formulas that must be verified by the instances in the KG. The basic datatypes provided by GraphBRAIN are: boolean, integer, real, string, and text. Ontologies can be combined using the import section provided that they are compliant with 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

[5] Cultural-ON (Cultural ONtology) [27] is an ontology aimed at modelling the data regarding cultural institutes or sites and their environment. It is aligned with external ontologies (FOAF, PROV, schema.org, Dublin Core, etc.).

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 GraphBRAIN are stored in a single graph, using the Neo4j graph DB [31]. 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 would otherwise be unknown. The system also acts as a repository, allowing to add files as attachments to the instances, and allows the users to approve, disapprove or comment on any information item stored.

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 the interconnection of ontologies and instances alike, and the 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 to 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 [32]).

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 with the ontology before running the query. Advanced functionalities 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 the 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 collaborative interactions to populate the knowledge, etc. Note that the form-based interface is automatically generated based on the ontology, ensuring seamless integration and full compliance of the data with the schema. Figure 1 shows the form-based interface for entity book (top) and for the relationship expressing that a company published a book (bottom), also showing instances of these elements with the associated attribute values taken from the IFLA-aligned ontology. A graph-based visualization is also provided, where the user can browse the knowledge, reshape and expand 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.

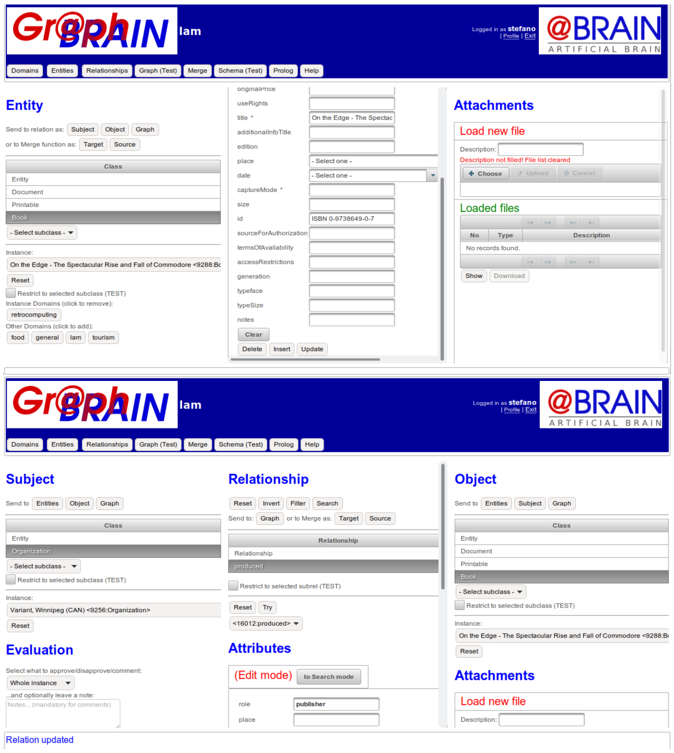

Figure 1: Form-based Interface in GraphBRAIN Web Application

For end-users, a separate interface called SKATEBOARD is also provided as a Web Application. It is mainly based on knowledge browsing and exploration, allowing one to visualize, expand or compress a portion of the graph, look into the single nodes or arcs, and 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 with standard SPARQL endpoints as well.

4. IFLA-based Ontology

In this section we describe the development of the ontology, that is to act as a schema for the GraphBRAIN technology-based KG. In particular, we will focus on the portion of holistic ontology that corresponds (i.e., is aligned and compliant) to the IFLA reference models FRBR and LRM. More specifically, due to space constraints, we focus here on the entities, that are central and preliminary to determining the relationships and attributes. We will provide information on the general design strategy and on relevant specific design decisions. The resulting alignment of entities is reported in Table 1. In order to provide a slightly broader view of the ontology, it also reports GraphBRAIN entities that are superclasses of classes aligned with IFLA entities but have no counterpart in IFLA standards.

The core of the ontology must concern established, necessary or useful descriptors provided

GraphBRAIN	FRBR	LRM
Entity	—	Res
+ Category	—	—
+ + Concept	Concept	—
+ + Object	Object	—
+ Agent	Responsible Entity	Agent
+ + Person	Person	Person
+ + Organization	Corporate Body	Corporate Body
+ IntellectualWork	—	—
+ + WorkOfArt	Work	Work
+ + + MusicalWork	Work (Musical attributes only)	—
+ + + CartographicWork	Work (Cartographic attributes only)	—
+ + Expression	Expression	Expression
+ + + SoundExpression	Musical or Sound	—
+ + + + MusicExpression	Musical	—
+ + + CartographicExpression	Cartographic	—
+ + + RemoteSensingImage	Remote Sensing Image	—
+ + + VisualExpression	Graphic or Projected Image	—
+ Document	Manifestation	Manifestation
+ + Printable	—	—
+ + + PrintedBook	Printed Book	—
+ + + + HandPrintedBook	Hand Printed Book	—
+ + Audio	Sound Recording	—
+ + Visual	—	—
+ + + Image	Image	—
+ + + Microform, Projection	Microform or Visual Projection	—
+ + ComputerFile	Electronic Resource	—
+ + + RemoteComputerFile	Remote Access Electronic Resource	—
+ Item	Item	Item
+ Collection	—	—
+ + Series	Expression>Serial, Manifestation>Serial	—
+ Nomen	Nomen	Nomen
+ Collection(Word), Taxonomy	—	Scheme
+ Place	Place	Place
+ Event	Event	—
+ TemporalSpecification	—	—
+ + TimeInterval	—	TimeSpan

Table 1
Entity alignment

for by library and archival theory, standards or practice. For entities, we consider as mandatory that the ontology includes all the items in the FRBR and LRM conceptual models proposed by IFLA to make up the core portion of the class hierarchy. A preliminary alignment was required between the IFLA standards, FRBR and LRM, as well, since they sometimes use different names for the same concepts or gave different interpretations to concepts with the same name. E.g., Agent in LRM corresponds to Responsible Entity in FRBR (where the name reflects a specific focus on the role), which were mapped onto the Agent entity in GraphBRAIN. Another alignment was needed between some IFLA standards names and the corresponding items already present in GraphBRAIN's *general* ontology. We reused this ontology because it supports the description of basic items.

The strategy for identifying subclasses, specifically under Expression and Document, was to define specific subclasses for those attributes in which the IFLA proposals specified applicability to some specific kinds of documents only. In this respect, a note is worth it for the "Series" class. In the IFLA proposals, "Serial" is a subclass of (i.e., specific attributes are defined for it under) both Expression and Manifestation. In our taxonomy, Series is a separate class, which is a subclass of "Collection". While as a subclass of Expression and Document it would inherit many attributes (author, publisher, title, etc.), in principle we decided not to consider it as an "expression" or "manifestation" of an intellectual work. Rather, it is a collection of Documents, because a series is made up of specific editions of expressions. Two classes in IFLA's conceptual models have been mapped onto classes with different names in GraphBRAIN, since they were already provided for by the general ontology:

- IFLA Work has been mapped onto WorkOfArt, which has a broader meaning by itself and is also a subclass of a more general IntellectualWork class, involving other kinds of works (e.g., algorithms, etc.). Indeed, literary works are a kind of art. This allows the connection of library materials to a broader range of intellectual works, that may be interconnected to each other.
- IFLA Manifestation has been mapped onto Document, which is intended in its etymological definition of "an instrument to teach". So, again, the scope of this class is broader than that of the IFLA class, and allows to express and include many other kinds of objects. As a minimum, it can include archival materials, in addition to bibliographic ones.

Some IFLA classes correspond to one GraphBRAIN entity. E.g., IFLA Concept and Object classes have been mapped as subclasses of a more general class Category, that also includes other kinds of elements in addition to ideas (corresponding to Concept) and physical objects (corresponding to Object): e.g., subjects, periods, etc. Items in Category can be inter-related, structured and organized in taxonomies, defining in some sense an ontology (e.g., WordNet synsets can be expressed as Category instances in our ontology). Differently from the classes, attributes and relationships in the schema ontology, this solution allows to represent ontological elements as instances in the knowledge graph, thus providing several advantages:

- allowing to connect them to other instances of the ontology (e.g., for use as descriptors);
- allowing to carry out meta-reasoning on them;
- allowing to continuously expand the set of ontological items of which the system can "talk" without having to change the schema ontology.

Also IFLA entity Serial, which in FRBR is a specialization of both Expression and Manifestation, was mapped onto the single GraphBRAIN entity Series, which is a specialization of Collection. The distinction can be recovered in an attribute, or as subclasses, of this class.

Conversely, other IFLA entities were split. E.g., 'Microform or Visual Projection' is clearly a class obtained by merging two well-distinguishable classes. So we provided for two distinct entities Microform and VisualProjection in our ontology, both specializations of a Visual entity.

In another case, the design decision was different. FRBR entities ‘Musical or Sound’ and ‘Graphic or Projected Image’ were mapped onto single entities SoundExpression and VisualExpression, respectively, removing the ugly disjunction in their name and assuming that music is also a kind of sound, as well as graphic and projected images are all cases of visual expressions.

Another design decision was to define two entities that were not explicitly present in FRBR, but most likely due to a design error. FRBR entity Work has some attributes that can be applied only to Musical or Cartographic works, and this is the typical case in which a good design approach would define two specializations of Work, that inherit the common attributes and define each of its specific attribute extensions.

LRM entity Scheme is another peculiar case. It accounts for thesauri (collections of words) and taxonomies, which in the general ontology of GraphBRAIN are described by two distinct entities, belonging to different branches of the hierarchy: the former is under top-level entity Collection, the latter is under IntellectualWork (the development of a taxonomy is the product of intelligence).

5. Conclusions & Future Work

The record-based approach to library information organization is obsolete and cannot support advanced opportunities provided by AI. Graph-based representations driven by ontologies are needed, as proposed by research in Knowledge Representation and Reasoning. Still, such research adopts representation models that are not compliant with those adopted by research on DBs. The GraphBRAIN technology proposes a framework to apply the ontological approach to the representation models adopted by graph DBs, taking the best of both: representational power and flexibility of ontologies and efficiency in data handling of DBs.

The IFLA proposed different conceptual models of library data that go beyond record-based representation, moving to relational representations, but still not reaching the ontological level. On the other hand, attempts to tackle the ontological level resulted in solutions that are not, or only partially, compliant with the IFLA models. This paper proposed an ontological solution that aligns the various models proposed by IFLA, fully representing their elements and is based on the GraphBRAIN technology. It can be translated into the standard representations adopted by the Semantic Web community, this way ensuring interoperability with the LOD Cloud.

The ontology is currently being aligned to other widely used representations, such as the Dublin Core Metadata Initiative and MARC, and can be expanded to include elements that are not traditionally represented in library models, but that can unleash new potentiality for library practitioners, researchers and end-users. In the perspective of the holistic approach, future work will expand the ontology to include descriptors that are not part of traditional bibliographic descriptions, but that can support the objectives of the CHANGES project: archaeology (especially related to books and other bibliographic materials, such as microfilms), history of books, law (for intellectual property management) and economics (for commercial exploitation of the items).

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] S. Ferilli, Holistic graph-based representation and ai for digital library management, in: Linking Theory and Practice of Digital Libraries, volume 13541 of *LNCS*, Springer, 2022, pp. 485–489.

[2] S. Ferilli, D. Redavid, D. D. Pierro, Holistic graph-based document representation and management for open science, International Journal on Digital Libraries (IJDL) 11:389 (2022) 23.

[3] S. Galabova, T. Trencheva, I. Trenchev, Standardized structure of electronic records for information exchange, in: FMNS2013: 3 International Scientific Conference, 2009.

[4] IFLA Study Group on the FRBR, Functional Requirements for Bibliographic Records - Final Report, Technical Report, International Federation of Library Associations and Institutions (IFçA), 2009.

[5] P. Riva, P. Le Boeuf, M. vzumer, IFLA Library Reference Model – A Conceptual Model for Bibliographic Information, Technical Report, International Federation of Library Associations and Institutions (IFCA), 2017.

[6] M. Žumer, P. Riva, Ifla lrm - finally here, in: Proceedings of the 2017 International Conference on Dublin Core and Metadata Applications, DCMI'17, Dublin Core Metadata Initiative, 2017, p. 13–23.

[7] A. Hinze, G. Buchanan, D. Bainbridge, I. Witten, Greenstone: a platform for semantic digital libraries, 2008.

[8] G. Alemu, B. Stevens, P. Ross, J. Chandler, Linked data for libraries: Benefits of a conceptual shift from library-specific record structures to rdf-based data models, in: Proceedings of the World Library and Information Congress: 78th IFLA General Conference and Assembly, 2012.

[9] J. Decourselle, Case-oriented semantic enrichment of bibliographic entities, 2016.

[10] J. Decourselle, Towards a pattern-based semantic enrichment of bibliographic entities, Bull. IEEE Tech. Comm. Digit. Libr. 12 (2016).

[11] P. Babu, A. Sarangi, D. Madalli, Knowledge organization systems for semantic digital libraries, in: International Conference on Trends in Knowledge and Information Dynamics: Vol. II. Paper presented at the ICTK 2012, DRTC, Bangalore: Documentation Research & Training Centre, 2012, pp. 988–1007.

[12] S. Kruk, B. Haslhofer, P. Piotrowski, A. Westerski, G. Poland, The role of ontologies in semantic digital libraries, European Networked Knowledge Organization Systems (NKOS) Workshop (2006).

[13] C. Bianchini, The entities of the ifla-lrm, ric-cm and cidoc-crm models in the semantic web, JLIS.it 13 (2022).
[14] M. Hallo, S. Luján-Mora, A. Maté, J. Trujillo, Current state of linked data in digital libraries, Journal of Information Science 42 (2016) 117–127.
[15] A. Passant, P. Laublet, Meaning of a tag: A collaborative approach to bridge the gap between tagging and linked data, Proceedings of the WWW 2008 Workshop Linked Data on the Web (LDOW2008), Beijing, China, Apr (2008). URL: http://ftp.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-369/paper22.pdf.
[16] S. R. Kruk, M. Synak, K. Zimmermann, Marcont: Integration ontology for bibliographic description formats, in: Proceedings of the 2005 International Conference on Dublin Core and Metadata Applications: Vocabularies in Practice, DCMI '05, Dublin Core Metadata Initiative, 2005.
[17] J. Lacasta, J. Nogueras-Iso, G. Falquet, J. Teller, F. J. Zarazaga-Soria, Design and evaluation of a semantic enrichment process for bibliographic databases, Data & Knowledge Engineering 88 (2013) 94–107.
[18] G. Miller, Wordnet: A lexical database for English, Communications of the ACM 38 (1995) 39–41.
[19] H. S. Patrício, M. I. Cordeiro, P. N. Ramos, Formalizing enrichment mechanisms for bibliographic ontologies in the semantic web, in: Metadata and Semantic Research, Springer International Publishing, Cham, 2019, pp. 147–158.
[20] S. Fathalla, S. Vahdati, S. Auer, C. Lange, Semsur: A core ontology for the semantic representation of research findings, Procedia Computer Science 137 (2018) 151–162. Proceedings of the 14th International Conference on Semantic Systems 10th – 13th of September 2018 Vienna, Austria.
[21] S. Sadhu, P. Anthony, P. K. Bhowmick, D. K. Sanyal, Towards development of ontology for the national digital library of india, ???? Available on-line at https://save-sd.github.io/2018/submission/sadhu/index.pdf (consulted 6 September 2023).
[22] L. J. G. Castro, O. X. Giraldo, A. G. Castro, Using the annotation ontology in semantic digital libraries, in: Proceedings of the 2010 International Conference on Posters & Demonstrations Track - Volume 658, ISWC-PD'10, CEUR-WS.org, Aachen, DEU, 2010, p. 153–156.
[23] A. Martín, C. León, Expert knowledge management based on ontology in a digital library, volume 2, 2010, pp. 291–298.
[24] M. Zeng, Semantic enrichment for enhancing lam data and supporting digital humanities. review article, El Profesional de la Información 28 (2019).
[25] C. Bekiari, G. Bruseker, M. Doerr, C.-E. Ore, S. Stead, A. Velios, Definition of the CIDOC Conceptual Reference Model v7.1.1 (Version v7.1.1), Technical Report, The CIDOC Conceptual Reference Model Special Interest Group, 2021. doi:https://doi.org/10.26225/FDZH-X261.
[26] V. A. Carriero, A. Gangemi, M. L. Mancinelli, L. Marinucci, A. G. Nuzzolese, V. Presutti, C. Veninata, 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, volume 11779 of *Lecture Notes in Computer Science*, Springer, 2019, pp. 36–52. doi:10.1007/978-3-030-30796-7_3.

[27] G. Lodi, L. Asprino, A. G. Nuzzolese, V. Presutti, A. Gangemi, D. R. Recupero, C. Veninata, A. Orsini, Semantic web for cultural heritage valorisation, in: Data Analytics in Digital Humanities, Springer International Publishing, Cham, 2017, pp. 3–37.
[28] S. Ferilli, D. Redavid, An ontology and knowledge graph infrastructure for digital library knowledge representation, in: Digital Libraries: The Era of Big Data and Data Science, volume 1177 of *Communications in Computer and Information Science*, Springer, Berlin, Heidelberg, 2020, pp. 47–61.
[29] S. Ferilli, D. Redavid, The GraphBRAIN system for knowledge graph management and advanced fruition, in: Foundations of Intelligent Systems, volume 12117 of *LNAI*, Springer, Berlin, Heidelberg, 2020, pp. 308–317.
[30] 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.
[31] I. Robinson, J. Webber, E. Eifrem, Graph Databases, 2nd ed., O'Reilly Media, 2015.
[32] S. Ferilli, Gear: A general inference engine for automated multistrategy reasoning, Electronics 12 (2023) 256.