A Holistic Ontology for Digital Libraries

The Educational Extension

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

[1] *University of Bari, Bari, Italy*
[2] *LO*OP Center, CA, USA*

Abstract

Overcoming the limitations of the record-based approach to library information by graph-based representations driven by ontologies may enable advanced support by AI and allow to expand the realm of the descriptions from the traditional formal metadata in several directions. An application domain that heavily relies on libraries and may take enormous advantage from this approach is education. In fact, an educational ontology is at the core of the KEPLAIR project aimed at developing a learner-centered educational environment. This paper describes our effort in this direction, and specifically how we seamlessly extended a core ontology based on the conceptual models of library data proposed by the International Federation of Library Associations and Institutions (IFLA) so that it is aligned with standard educational data models, such as the IEEE LOM, OERschema, and IntelLEO. The alignment resulted in a cross-fertilization of the two domains, and allowed us to better understand them and to find and resolve subtle issues in the aligned resources.

Keywords

Digital Libraries, Knowledge Graphs, Ontologies, Graph Databases

1. Introduction

Adopting a graph-based representation of library information instead of the traditional record-based one (inherited from its original application to paper cards) allows to overcome the limitations of the latter and to easily expand the realm of descriptions. In the graph-based approach, the different kinds of entities involved in such knowledge exist on their own, rather than being just values in record fields, and relationships among entities are the key feature of descriptions.

20th Italian Research Conference on Digital Libraries (IRCDL2024, https://ai4ch.di.unito.it/)). 22-23 February 2024, Bressanone, 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); lizaloop@loopcenter.org (L. Loop)
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); https://loopcenter.org/contact.php (L. Loop)
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)

CEUR Workshop Proceedings (CEUR-WS.org)

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

Graph-based representations are also the basis for semantic technologies, driven by ontologies, that provide formal definitions of the entities, relationships, and attributes in the domain, and of their behavior. Adopting the ontological setting causes a step up from the Data Base (DB) setting to the Knowledge Base (KB) one, enabling interoperability among systems and platforms and advanced exploitation of the stored information by advanced Artificial Intelligence (AI) techniques, such as data mining and automated reasoning, to support all kinds of library stakeholders. KBs organized as graphs are also called Knowledge Graphs (KGs).

In relation to the possibility of expanding the realm of descriptions, we proposed a 'holistic' perspective, that can fully grasp and express the complexity of library information [1, 2]. In this perspective, the set of descriptors becomes 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, the *content* of the documents, their *materiality*, their *context* and their *lifecycle* (involving their users and their uses) should be included. 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. This includes expansions of the descriptions concerning specific fields of interests connected to libraries.

Developing a first version of such a 'holistic' ontology for library information description was the objective of Task 1 for the University of Bari team involved in 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. A first piece of the result of this effort was the specific portion of ontology that refers to standard library descriptors, as defined by leading world library institutions (IFLA), and that represents the core of the complete ontology [3].

In this paper we describe one of the possible extensions of the core, and specifically the one related to the educational domain. In fact, one of the main functions of libraries is supporting students, scholars and researchers in their activities, and thus they must be suitably related to, and integrated in, the overall educational endeavor. This is also in line with the objectives and tasks of another project, named KEPLAIR, aimed at developing an AI-based platform for supporting learners in their educational efforts [4]. The ontology is the heart of the KEPLAIR framework, and the outcome of this work will represent the core portion of that ontology.

As for the previous 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 supports 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: KEPLAIR and GraphBRAIN

KEPLAIR (Knowledge-based Environment for Personalised Learning using an Artificial Intelligence Recommender) is a joint project by the ARA research group at the University of Bari (Italy) and the LO*OP Center (USA). It envisions an online platform mainly designed for learners, and aimed at helping its users find learning opportunities and materials. Using learning goals chosen by the learner, KEPLAIR will harvest relevant materials, filter them and make recommendations that match the learner's cognitive level, pre-existing knowledge about the topic, and preferred (digital, physical and social) environments. KEPLAIR's purpose is to help learners reach their self-chosen goals by highlighting appropriate, attractive, and useful materials so they stand out from the background noise. This will be done in a highly personalised way for each single user, taking into proper account the many aspects involved in recommending, such as needs, background, abilities, aims, interests, tastes, preferences, attitudes, behaviours, motivations, expectations, context, and community. KEPLAIR does not aim at teaching, testing, or promoting a predetermined curriculum, nor does it require learners to be part of any formal school or learning organisations.

Crucial to KEPLAIR's implementation is an ontology that includes four major learning classes: Goal/Pathway; Learner Profile; Social, Physical, & Digital Environment; and Learning Resource. Based on such an ontology, the AI will draw on semantic analysis of online materials from formal educational institutions, open educational resources (OER), and pre-existing pathways, environments and learning objects. The ontology informs a KG that is the basis for all of KEPLAIR's behavior and functionality, implemented using the GraphBRAIN technology.

GraphBRAIN [5, 6] 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 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. Another difference from the Semantic Web approach, but similarity with traditional relational DBs, is that GraphBRAIN keeps apart the schema/ontology, described in a GBS file using an XML-based formalism, from the data/instances, stored in the DB. The ontology acts as the schema to determine what information can be stored in the graph, and how.

GraphBRAIN ontologies are organized into different sections: import of existing ontologies to be reused; definition of new enumeration or tree-based datatypes; definition of a hierarchy of entities with their attributes; definition of a hierarchy of relationships with their properties (symmetry, transitivity, functionality, etc.) and attributes; definition of axioms in the form of logic formulas to be applied to the instances in the KG. The basic datatypes provided by GraphBRAIN are: boolean, integer, real, string, and text. Ontologies can be combined provided that they are compliant with each other. Two ontological components are considered as the same if they have the same name. A relevant feature in GraphBRAIN ontologies is the possibility of defining *abstract* classes. Abstract classes cannot have direct instances (only their specializations can). While any class in the taxonomy may be defined as 'abstract', an additional kind of abstract classes is provided for, in the form of a union of other existing classes in the taxonomy. This is just ‘syntactic sugar’ to collect under a single name several different entities, allowing to partly recover some features of multiple inheritance (which is not available in GraphBRAIN).

GraphBRAIN stores all instances in a single graph, using the Neo4j graph DB [7]. Different ontologies may be applied to the same graph, providing different views on the data. When selecting an ontology to access the graph, the information associated with other ontologies, albeit not visible, may provide indirect connections among the items of the current ontology. 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.

GraphBRAIN is interoperable with Semantic Web repositories, thanks to a mapping by which ontologies and instances can be exported or imported to or from the Semantic Web standard OWL. This allows to immediately use ontological reasoners on the knowledge handled by GraphBRAIN. Additionally, a large set of network analysis and graph mining functions can be applied to the data, provided by the Neo4j libraries and tools or by other AI tools (we are currently working on the MultiStrategy Reasoning engine GEAR [8]).

A GraphBRAIN API is available that can be used by any third-party application. It ensures that, given an ontology and a DB, all interactions with the DB happen according to the schema. Basic functionality includes standard CRUD (Create, Read, Update, Delete) operations. For queries, it wraps the Neo4j language Cypher. Advanced functionality include analysis, mining and reasoning functions.

3. Related Work

3.1. Library Description

Current standards for data description used in the library practice, such as the MARC format and the ISO 2709 [9], are still deeply rooted in the record-based idea of library descriptions. The International Federation of Library Associations and Institutions (IFLA) opened to relational descriptions with its FRBR (Functional Requirements for Bibliographic Records) [10] and LRM (Library Reference Model) [11] conceptual models, that identified a set of entities to be represented and described (e.g., places, persons, organizations), along with their characterizing attributes and relationships that were implicit in traditional bibliographic records. These proposals still focus on the syntactic level of descriptions, and on formal metadata or properties that are directly related to the documents to be described.

As a step up, Semantic DLs (SDLs) [12] can use ontologies to organize bibliographic descriptions, represent and expose document contents, and share knowledge among users [13]. The IFLA effort in this direction resulted in the Resource Description and Access (RDA) standard. Other works, not originated from IFLA, are based on FRBR, aimed at ingesting data from MARC sources [14], fostering a switch to open data [15], or focusing onmetadata migration and semantic enrichment of bibliographic data [16].

Moving from traditional descriptions, [12] 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, [17] 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. [18] depicts the situation of Linked Data technologies applied to libraries, outlining best practices, gaps and future trends. Most works do not use fully-fledged ontologies, just controlled vocabularies and thesauri. Even those that use ontologies, never propose an overall ontology from the specific DL perspective. An exception is the MarcOnt Ontology [19] as a candidate for combining different metadata standards that can describe various concepts at different levels of granularity. Unfortunately, the project does not exist anymore.

Some works highlight the need for an overall ontology that accommodates and ‘coordinates’ all the different perspectives: [20] 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. [21] 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: survey and review articles [22], educational resources [23], free and semantic annotation (tags, notes, comments, errata, etc.) over the paper, also providing facilities for curation, provenance, authoring and versioning [24], Open Science [2]. CIDOC-CRM [25] widens the extent of descriptions in the direction of Cultural Heritage.

Summing up, while it is widely acknowledged that ontologies are a need and an opportunity for modern and advanced DL and bibliographic data management, the existing data representations are almost always stuck to standard metadata, and only occasionally perceive the need to expand them.

Ontologies are important also because they enable automated reasoning. E.g., [26] 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 [27].

3.2. Educational Description

Metadata schemes in education were strongly influenced by Bloom’s taxonomic systematisation work [28] prior to the development of the Internet and thus the knowledge-sharing tools available today. Over time, various revisions have been applied in order to include the technological developments that have taken place in recent decades [29].

First, an attempt was made to create a set of standardised metadata to have a common and shared nomenclature of the various categories identified. These led to two main proposals, both proposed to facilitate cataloguing, searching, and reuse of such resources [30]: IEEE Learning Object Model (LOM) [31] and Dublin Core (DC) [32]. While the LOM specification starts from the analysis of educational aspects, the DC starts from the analysis of a structure capable of specifying metadata of any type and then converges, in a sub-set, towards educational ones. There is thus a difference in the specified metadata sets that overlap for a sub-set but are not exactly the same. Furthermore, at the level of representation LOM presents a record organisation while DC uses the RDF data model. This is reflected in the sharing modes: LOM uses the Sharable Content Object Reference Model (SCORM) [33], while DC uses the Semantic Web.

From these metadata standards, ontologies were developed for practical use at the application level. For instance, an ontology based on RDF Schema [1] was developed for LOM in order to expose content as Linked Open Data (LOD). DC is instead used as a vocabulary in many educational ontologies developed within the Semantic Web [34]. A more recent initiative is the Open Educational Resource (OER) [2] Schema, also supported by the W3C, whose aim is the development of a universal RDF vocabulary to enhance open educational resources throughout the Internet.

The development of ontologies is, however, dependent on the starting educational models. Indeed, it is necessary to specify the concepts and properties that define such a model and enable its use at application level, particularly in the field of e-learning [35]. One example is The Intelligent Learning Extended Organisation (IntelLEO) [36] is a service-oriented framework for collaborative learning and knowledge construction between organisations of different natures (corporate, academic or other interested communities). The framework specifies an ontology based on concepts and properties useful for the realisation of services to discover, access, manage and reuse learning resources in the organisations involved.

4. Educational Extension of the Library Ontology

This section describes how we developed the core of a general ontology for education, that merges several existing but independent schemes. While the result can be valuable by itself, since it allows communication and interoperability of resources described according to the different schemas, we specifically aim at using it within the KEPLAIR project, and thus at exploiting it as a schema for the GraphBRAIN technology. 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, but no implementation details or examples, again due to space constraints.

Our educational ontology relies on a previous ontology for library concepts, fully aligned with the conceptual models proposed by IFLA. In facts, libraries are essential as learning object repositories, and a fundamental component of any educational system. Concerning the educational section proper, we decided to include the following schemes, because they are complementary to each other and derive from different approaches (more schemas can be aligned and added in the future):

IntelLEO a comprehensive formal ontology, developed in a research project, describing the dynamic aspects of education and the organizational structure of learning objects;

LOM a standard IEEE metadata schema for learning objects, still rooted in the record-based perspective;

OERschema a less widespread schema more focused on the concepts related to course organization and delivery.

Our design decision was to include all the elements in these schemes, so as to ensure interoperability with the existing resources based on them. This required us to identify an alignment

---

[1] https://lov.linkeddata.es/dataset/lov/vocabs/lom
[2] https://oerschema.org/

GraphBRAIN	IFLA	IntelLEO	LOM	OERschema
Entity	Res	Thing (owl)	—	—
+ Category	—	—	✓	—
+ + Concept	Concept	Concept		
+ + + Topic	—	Topic	—	Topic
+ + Object	Object	—		
+ Agent	Agent / Responsible Entity	Agent	✓	
+ + Person	Person	Person		
+ + Organization	Corporate Body	Organization		
+ IntellectualWork	—	—	—	✓
+ + WorkOfArt	Work	Work		
+ + + MusicalWork	Work (Musical attributes)	—		
+ + Expression	Expression	—		
+ + + SoundExpression	Musical or Sound	Audio		
+ + + + MusicExpression	Musical	Music		
+ + Taxonomy	Scheme	Scheme	✓	—
+ Document	Manifestation	—	—	AssociatedMaterial
+ + Printable	—	—		
+ + Audio	Sound Recording	Audio		
+ + Visual	—	—		
+ + + Image	Image	—		
+ + + Video	—	—		
+ + + + Movie	—	—	—	Movie
+ + ComputerFile	Electronic Resource	—		
+ + + RemoteComputerFile	Remote Access Electr. Res.	Document (foaf)	LOM	—
+ Collection	—	—		
+ + Collection(Word)	Scheme	—	✓	—
+ Place	Place	—	—	—
+ Event	Event	—		
+ TemporalSpecification	—	—		
+ + TimeInterval	TimeSpan	TimeSpan		
+ Software		—	—	—
+ + OperatingSystem		Environment>OS	OperatingSystem	—
+ + Browser		Website.browser	Browser	—
+ Goal		—	—	—
+ + LearningGoal		LearningGoal	—	LearningObjective
+ LearningObject		—	Educational	Resource

Table 1
Entity alignment for the general GraphBRAIN ontology, the library standards and the educational ontologies

between the names (and underlying concepts) of the elements in the different schemes, and between them and the corresponding items already present in GraphBRAIN's extant ontologies (the *general* one and the library one). In turn, the alignment sometimes required us to split, merge or rename classes in the existing schemes, or to add new classes (usually generalizations). The resulting alignment of entities is reported in Table 1.

In order to provide a slightly broader view of the ontology, the table also includes generalizations that have no counterparts in the aforementioned schemas. These generalizations were present in GraphBRAIN's extant ontologies.

General considerations on the alignment in Table 1 are as follows.

- There are a number of shared classes, that are fundamental to act as bridges connecting the information associated to the two domains (Libraries and Education). As expected, being IntelLEO quite large an ontology, it shares many classes, especially top-level ones, with IFLA. On the other hand, LOM and OERschema have very few classes in common with both IFLA and IntelLEO. For LOM, this can be explained by its being a metadata schema, and thus more record-oriented: indeed, while not explicitly defined in the standard, many classes (marked as ✓in Table 1) are implicit in the type of LOM fields. However, the shared classes are fundamental and ubiquitous (Topic and Document for OERschema, Document for LOM), ensuring a tight integration among the different knowledge sections.

- In addition to the shared ones, IntelLEO provides lots of other classes, mainly connected to the educational process, many of which already included in GraphBRAIN's existing ontologies. There are classes related to Users (especially concerning their account data and their profiles), Projects and their components (Tasks, Milestones, Effort, etc.), Environments, Learning Styles, Goals and Competences, Activities, Annotations and, perhaps strangely enough, Movies. On the other hand, compared to IFLA, it lacks descriptive depth on the different kinds of documents and on their features: no Cartography, Printable, Remote Sensing, Microform/Projection, 'Graphic or Projected Image', Serial, nor any distinction between Expression, Manifestation, and Item, are present.
- As said, if IntelLEO is more focused on the dynamic aspects of education, LOM and OERschema are more related to the static description of learning objects and of the educational domain, respectively. The additional classes in OERschema express educational containers (Class, Course, etc.), learning materials (for which some overlapping with LOM is present) and assessments.
- The educational-specific schemes have no Place class, in spite of its being fundamental in any domain. This is probably due to their focusing specifically on the on-line setting.

Let us now discuss in more details some cases of alignment and the underlying rationale. We first consider the alignment to IFLA classes:

- Both LOM and OERschema deal with digital documents available on-line, and thus their Document and LOM classes correspond to the RemoteComputerFile specialization of Document in GraphBRAIN. Specifically, LOM is the root of the LOM schema, aimed at describing one educational resource (i.e., a Document in the interpretation of the other schemes).
- IntelLEO provides for a Concept class, analogous to those in IFLA and GraphBRAIN (where it is a specialization of Category). On the other hand, no Object class, expressing in IFLA the concrete counterpart of Concept, is present in any educational-related scheme/ontology. Moreover, both IntelLEO and OERschema provide for a Topic class, which is not available in IFLA, but is a specialization of Concept in GraphBRAIN.
- Albeit not defining a formal class for it, OERschema mentions "creative works" in its definitions, which can be associated to the IntellectualWork class in GraphBRAIN (including any kind of product of intellect: works of art, algorithms, taxonomies, etc.).
- Due to its record-based nature, often in LOM some pieces of information (e.g., Categories, Agents, Schemes) are considered as simple string fields to describe the learning objects, without recognizing their relevance as separate entities.
- In contrast to the previous case, LOM provides for a fully-fledged class "Classification", described by several attributes, which cannot be associated with a simple Category in GraphBRAIN, and thus required the addition of a new class in our educational ontology.

- IntelLEO has an entire section (sub-ontology) devoted to Movies, which are not present in the other (library or educational) schemes, but corresponds to a deep specialization of Document in GraphBRAIN. Sounds are also considered in IntelLEO, but only in connection with movies' soundtracks and narratives. In our alignment, we considered these elements as Manifestations (in IFLA's terminology), because the actual embodiment, not the idea, is involved.
- IFLA entity Scheme is another case worth discussing. 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 a product of intellect). LOM considers both, while IntelLEO provides only for the Taxonomy case.
- IntelLEO defines a class Event which has a different meaning than the class with the same name defined in IFLA and GraphBRAIN. The latter is concerned with historical or life events, the former with events having to do with the digital interaction between a user and the learning environment (like clicks, link following, etc.).

Although not about classes, another connection between IFLA and IntelLEO concerns the 'periodicity' datatype, used in IFLA for describing serials, and in IntelLEO for describing the frequency of recommendations for users.

Then, concerning the alignment among classes in the educational schemes that are not present in the library core:

- As a consequence of their focusing on on-line resources, LOM and OERschema define specific classes for browsers and operating systems, which are not available in IFLA (and in IntelLEO). These classes are present in GraphBRAIN extant ontologies, as specializations of Software. In particular, "browser" appears just as a string field, not as a fully-fledged class, in IntelLEO.
- LearningObject is an abstract class introduced in GraphBRAIN to collect several other classes located in different branches of the class hierarchy. It includes EducationalContainers (courses, modules, classes, etc.), multimedia Documents, EducationalTools (for practice or assessment), etc. In particular, it has an alignment to OERSchema's Resource, but with two exceptions. One is that LearningObjective and Topic, specializations of Resource in OERschema, have not been considered as specializations of LearningObject in GraphBRAIN, not being indeed 'objects'. The other is that OERschema's classes Course and CourseSection are considered in GraphBRAIN not as direct specializations of LearningObject, but of EducationalContainer (corresponding to InstructionalPattern in OERschema, which is a specialization of Resource).
- Goal is another example of abstract class in GraphBRAIN, because it is composed of several quite different entities. In the educational domain, its subclass LearningGoal may include subjects, accomplishment evidences, skills, jobs, documents and many others that can all be considered as the objective of a learning activity. It includes IntelLEO's LearningGoal and OERSchema's LearningObjective.
- Also, IntelLEO provides for two classes: ContinuousContentFragment (with specializations Animation, Audio, and Video) and DiscreteContentFragment (with specializations

Graphics, Image, and Text). Since their subclasses have been placed in different paths in GraphBRAIN (e.g., Graphics and Image are under Visual, just like Animation and Video, but unlike Audio and Text), these classes have been added for alignment purposes as abstract classes in GraphBRAIN.

A few final notes about inconsistencies in IntelLEO that were found during the alignment:

- Class ProgressScale defined/used in subsections ‘competencies’, ‘userModel’ and ‘workflow’ is exactly the same as class NumericScale, subclass of Scale, defined in subsection ‘annotations’. They were merged in our alignment, preferring the latter name that is more generic (no specific information about the use of the scale for describing ‘progress’ was present in the other concept).
- Class TaskPriority defined in subsection ‘userModel’, with values {low, medium, high} expresses exactly the same meaning as class GoalPriority defined in subsection ‘annotations’, with values {LowPriority, IntermediatePriority, HighPriority}. In our alignment they were merged and mapped onto a class named 3LevelHeightScale, that has no reference to specific uses of the scale (for priority, or for tasks/goals), having values with generic names {low, medium, high} (not connected to the specific priority use).

5. Conclusions & Future Work

Libraries are an important component of an educational setting. Thus, a system aimed at providing advanced support to education stakeholders must rely on, and suitably extend, effective and efficient handling of library information. However, the traditional 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. This is the background on which the KEPLAIR initiative to develop a learner-centered, AI-based educational platform was built.

In this context, this paper proposed an ontology that aligns several, different but complementary, schemes to describe the educational domain, both among themselves and with the models for library description proposed by IFLA. The alignment effort allowed us to enrich the IFLA-based library ontology with additional descriptors that are relevant to the library domain, and to tightly and seamlessly connect the library core with the educational extension. A number of concepts definitions were refined and a few inconsistencies were fixed during the work.

The ontology is currently being aligned to other widely used representations, both in the library direction (e.g., aligning the Dublin Core Metadata Initiative and MARC) and in the educational direction. Specifically concerning the latter, we think that all existing schemes to describe educational knowledge are biased toward the teacher’s perspective (in spite of prefixing the various class names by ‘Learning’), and that an extremely relevant contribution of AI to education may come from its supporting the learners, and thus adopting their perspective. For this reason, after securing the alignment with existing schemes recognized and/or adopted by the research or practice communities, in future work we will expand the educational section of the ontology in the learner’s direction. Also, in the perspective of the holistic approach, we will expand the ontology in other directions, to include the content, context and use of the educational resources, usually neglected by the available descriptions.

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.

The authors would like to thank Avital Drucker and Michael Leaf for the useful discussions on the ontology and manuscript.

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. 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.

[4] S. Ferilli, D. Redavid, D. Di Pierro, L. Loop, An ontology-driven architecture for intelligent tutoring systems - with an application to learning object recommendation, International Journal of Computer Information Systems and Industrial Management Applications (IJCISIM) 14 (2022) 297–312.

[5] 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.

[6] 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.

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

[8] S. Ferilli, Gear: A general inference engine for automated multistrategy reasoning, Electronics 12 (2023) 256.

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

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

[11] 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.

[12] 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.

[13] 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).

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

[15] 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.

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

[17] C. Bianchini, The entities of the ifla-lrm, ric-cm and cidoc-crm models in the semantic web, JLIS.it 13 (2022).

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] 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).

[24] 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.

[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] A. Martín, C. León, Expert knowledge management based on ontology in a digital library,

volume 2, 2010, pp. 291–298.

[27] M. Zeng, Semantic enrichment for enhancing lam data and supporting digital humanities. review article, El Profesional de la Información 28 (2019).

[28] B. Bloom, D. Krathwohl, Taxonomy of Educational Objectives: The Classification of Educational Goals, number v. 1 in Taxonomy of Educational Objectives: The Classification of Educational Goals, Longmans, Green, 1956.

[29] L. W. Anderson, D. R. Krathwohl, P. W. Airasian, K. A. Cruikshank, R. E. Mayer, P. R. Pintrich, J. Raths, M. C. Wittrock, A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives, abridged edition, 2000.

[30] M. Mcclelland, Metadata standards for educational resources, Computer 36 (2003) 107 – 109. doi:10.1109/MC.2003.1244540.

[31] Ieee standard for learning object metadata, IEEE Std 1484.12.1-2020 (2020) 1–50. doi:10.1109/IEEESTD.2020.9262118.

[32] B. Eversberg, Dublin Core, Technical Report, 2008. URL: http://www.allegro-c.de/formate/kap107.htm.

[33] B. Furht (Ed.), Sharable Content Object Reference Model (SCORM), Springer US, Boston, MA, 2008, pp. 816–817. doi:10.1007/978-0-387-78414-4_219.

[34] S. M. Rashid, D. L. McGuinness, Creating and using an education standards ontology to improve education, in: K. K. Waterman (Ed.), Proceedings of the Workshop on Semantic Web for Social Good co-located with 17th International Semantic Web Conference, SW4SG@ISWC 2018, Monterey, California, USA, October 9, 2018., Monterey, California, USA, October 9, 2018, volume 2182 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2018. URL: https://ceur-ws.org/Vol-2182/paper_5.pdf.

[35] K. Stancin, P. Poscic, D. Jaksic, Ontologies in education - state of the art, Educ. Inf. Technol. 25 (2020) 5301–5320. URL: https://doi.org/10.1007/s10639-020-10226-z. doi:10.1007/S10639-020-10226-Z.

[36] D. Stokić, K. Pata, V. Devedžić, J. Jovanović, L. Urošević, D. Gašević, B. Kieslinger, J. Wild, Intelligent learning extended organizations, Proceedings of TELearn2008 (2008).