Managing Learning Resource Metadata for Secondary Education

Bernhard Göschlberger
Research Studios Austria FG
Department of Telecooperation,
Johannes Kepler University, Linz
Linz, Austria
goeschlberger@researchstudio.at

Fabian Dopler
Research Studios Austria FG
Linz, Austria
fabian.dopler@researchstudio.at

Christoph Brandstetter
Research Studios Austria FG
Linz, Austria
christoph.brandstetter@researchstudio.at

ABSTRACT

Digital learning resources are becoming increasingly important, especially in times of COVID-19 home schooling. To align resources with educational objectives defined in school curricula, we need ways to describe and relate them. In this work, we present a new data model, a data management workflow and a prototypical information system enabling annotators to describe learning resources using the terminology and structure, while maintaining applicability in other contexts through transcription. Preliminary evaluation results with metadata curators show that our model is well-suited to efficiently model secondary education curricula. Provided that future evaluations with content creators, teachers and learners yield positive results, our model may enable curriculum specific learning analytics, search and recommendation without the need to annotate resources for each curriculum.

CCS CONCEPTS

• **Applied computing** → **E-learning**; **Document metadata**; • **Information systems** → *Ontologies.*

KEYWORDS

learning resource metadata, metadata management, transcriptivity, json schema

ACM Reference Format:
Bernhard Göschlberger, Fabian Dopler, and Christoph Brandstetter. 2020. Managing Learning Resource Metadata for Secondary Education. In *The 22nd International Conference on Information Integration and Web-based Applications & Services (iiWAS '20), November 30-December 2, 2020, Chiang Mai, Thailand.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3428757.3429148

1 INTRODUCTION

Over the past decades initiatives, like the Open Educational Resource (OER) movement, have accomplished that a plethora of learning resources is nowadays available on the World Wide Web. With a growing body of resources, researchers and practitioners typically acknowledge the importance of quality learning resource metadata to support search, recommendation and orchestration.

Learning resources can be described in a variety of ways: either through standardized or through unstructured metadata such as tags [9]. Widely accepted standards are important to make OERs comparable and support discovery within general-purpose search engines and they usually represent the least common denominator. While this is a feasible way to provide metadata of learning resources to a general audience, it does not take into account the users' local context, i.e. the terminology and structure they are using in their professional practice, dependent on the educational system, school type and curricula they are familiar with.

Tags and folksonomies, on the other hand result in metadata conflicts that can be subsumed in two main categories [6]: (a) semantic heterogeneity: if a disagreement about the meaning, interpretation or intended use of the same or similar data occurs (b) structural heterogeneity: if the same concepts with different logical structures occur in different systems. In most cases no direct concept to concept mapping is possible. So while the information is addressed at a specific audience with a specific context, it cannot be transferred from one context to another. Instead, it needs to be transcribed for the target audience. Transcriptivity theory refers to this process of contextualizing a medial artifact with an intended recipient in mind as addressing [5]. In this work, we design an information system providing a metadata model as well as a metadata management workflow to address the following challenges:

- allow annotators to use their local context when annotating (e.g. curriculum structure and terminology)
- allow transcription into other local contexts (e.g. different curriculum, new version of an existing curriculum)
- allow transcription into a generalized metadata representation (e.g. metadata standard)

We follow a design science approach to design the information system meeting these fundamental requirements [4]. The remainder of this paper is structured as follows: we present the most important related work we identified in our first rigor cycle iteration in section 2, followed by a description of the design goals that emerged from our first relevance cycle iteration in section 3, before we describe the outcomes of our first design cycle in section 4. Finally we present our first preliminary results in section 5 and conclude by providing outlook on future work in section 6.

2 RELATED WORK

Standardized metadata, such as LOM[1], specify which aspects of learning objects should be described and which vocabulary should

[1]https://standards.ieee.org/standard/1484_12_1-2002-Cor1-2011.html

iiWAS2020, November 30–December 02, 2020, Chiang Mai, Thailand

ACM ISBN 978-1-4503-8922-8/20/11…$15.00
https://doi.org/10.1145/3428757.3429148

be used [1]. The goal is to obtain well-structured descriptions of learning resources to support the discovery and evaluation of learning resources by teachers, learners and automated software processes. The standard itself defines 80 data fields which are structured in the categories general, life cycle, meta-metadata, technical, educational, rights, relation, annotation and classification. However, the LOM standard has some disadvantages. The exact meaning of some metadata is not exactly defined (e.g. difficulty) and also some important aspects of the learning objects cannot be represented. For example, it is not possible to express whether a learning object is intended for a single individual learner or for group work. Finally, not all defined metadata are fully machine readable: some are defined as free text while others are based on vocabulary that is not precise enough to be processed fully automatically [2].

Schema.org has become a widely agreed format for embedded markups and has recently introduced new terminology to define educational content under the requirements of the Learning Resource Metadata Initiative (LRMI)[2]. Metadata schemes have different descriptions for different purposes. LRMI was created for the publication of structured descriptions on the web with the aim of improving search engine research. The LRMI properties adopted by Schema.org can be divided into educational properties to describe unique features of educational content and general purpose properties to describe features that can be extended to several forms of web content [8]. Some problems, like missing specification and more controlled vocabularies (e.g. language information, topics, type of school, school subjects,...) for the values of metadata elements are still existing.

Another possibility is to use tags without any controlled vocabulary. This allows to design a flexible system and to optimize it for usability [7]. Vuorikari and Koper [10] have compared 3 different types of tags: tags for personal use, tags used when searching for similarly annotated resources, and tags that describe a resource which the user wants to share with others via a tag cloud. The authors found variations depending on the target of each set of tags. Despite these differences, some similarities emerge which allow the annotated earning resources to be reused in multiple systems. For example, learning resources annotated with tags like Mathematics and Linguistics may be repeatedly reused. There are also tags that can be reused in libraries in various countries if written in different languages, such as proper names, regions, and acronyms, similarly. Tags can have some drawbacks, for example, in contrast to other tags and user information a tag context must be interpreted and assumed. A tag may only have value for the person who created it, and could have been annotated with a particular intent in mind.[2]

Generally speaking there are several conflicts that may arise and need to be resolved in the design of a metadata model. Jetinai et al. list the following [6]: (1) *Naming conflicts.* When synonyms are used, which are semantically equivalent concepts or properties defined by different names. (2) *Scaling conflicts.* When different scales or units of measurement are used. (3) *Property value conflicts.* When semantically equivalent properties are defined with different property values. (4) *Generalization conflicts.* When the concepts in one system subsumes the concepts in another system. (5) *Aggregation conflicts.* When a property or definition maps a collection of properties in one system or concepts in another system (6) *Property discrepancies.* When semantically equivalent properties are defined with different property types (datatype). (7) *Concept discrepancies.* When the logical structure of a set of properties and their values that belong to a concept are arranged in one system to create an independent structure in another system.

[2]https://www.dublincore.org/specifications/lrmi/lrmi_1/

3 DESIGN OBJECTIVES

E-Learning is an increasingly important complementary activity to classroom learning in secondary education. Its importance became even more apparent in the COVID-19 pandemic, when many schools where caught by surprise and failed to deliver quality education. The presented work is part of a cooperative research project run by the Research Studios Austria FG and ChabaDoo, a company offering tools and solutions supporting digital learning for secondary schools. The research project itself is focused on search, recommendation, learning path optimization and learning analytics. Its goal is to provide increasing support for individualized learning paths within the rigid boundaries of formal education. Consequently, activities and outcomes need to be linked back to predefined formal curricula.

3.1 Model of System Users

In our first relevance cycle iteration, we identified five distinct stakeholder groups.

3.1.1 Content Creator. People who author learning resources. They use either special authoring tools that are not integrated into the e-learning platform or they can use an simplified content creation tool integrated directly into the platform. In the future content creation will become an more integral platform feature and it will become easier for teachers to create and publish their own content.

3.1.2 Teacher. The teachers are responsible for accompanying and supporting the learning process of the students. A teacher can teach several main focuses (subjects) including in different school systems for various age groups. They use tools to find relevant learning content and then prepare it as recommendations for their learners. Teachers can also become "content creators" when they create their own learning resources. They also want to know what the learners have learnt in order to have a constant overview of their development.

3.1.3 Metadata Curator. Metadata curators are responsible for building and populating the metadata catalog. These are specialized persons who mostly have a connection to the teacher profession and possess pedagogical knowledge. They also know the specific school system requirements that are necessary to guarantee educational standards. They are thus able to establish a structure for the semantic relationship between educational standards and learning resources.

3.1.4 Learner. The learning resources themselves are used by the learners. Above all, the learners need a way to find suitable learning content during independent learning. This learning content should correspond to the learner's abilities and be thematically relevant. It is important for learners to receive feedback on their learning activities in order to assess their own learning success.

3.1.5 Public Domain User. Learning resources may be published as OERs to increase the visibility of the product. Public domain users search the internet for learning resources. They are either teachers looking for supplemental resources for their students or self-determined learners.

3.2 Model of System User Requirements

Each of the stakeholder groups has their own specific view on learning resource metadata and on a desirable learning workflow. Through iterative inquiry we identified, organized and refined a set of requirements for each of the identified stakeholder groups.

As a content creator…

- I want to add metadata to newly created learning content.
- I want to situate newly created learning content in the context and structure of a curriculum.
- I want to use terminology and educational frameworks that I'm familiar with.
- I want to specify for a learning resource which topics or competencies of a curriculum are covered.

As a teacher…

- I want to find learning contents that are appropriate for the current level of knowledge of my students.
- I am looking for teaching contents that correspond to the age of the students.
- I want to monitor my students progress along a defined curriculum.
- I teach the same subject at different types of schools.
- I want to use learning resources from or designed for different curricula.
- I want to use learning resources that were created in another school system.
- I want to use a learning resource in different subjects (e.g. mathematics/music lessons).

As a metadata curator…

- I want to create a taxonomic structure reflecting existing curricula.
- I want to create relationships between similar curricula and similar parts of curricula.
- I want to update curriculum information and account for changes over time.
- I don't want to update individual annotations on learning resources.

As a learner…

- I want to be able to search for learning content according to my level of knowledge and competencies.
- I want to know about my current learning progress.
- I want to know which areas of knowledge I have already learned and which I have not yet learned.

As a public domain user…

- As a public domain user I want to be able to find learning resources using third party search engines.
- As a public domain user I want to be able to compare learning resources to other OERs.

Based on the identified roles and their specific requirements we designed a workflow and a metadata model. As the interests of different roles may conflict, it is important to state that the needs of teachers and content creators have higher priority in the projects context. For instance, the teachers' requirement to monitor students progress along a defined curriculum makes it necessary to link the metadata to the respective curriculum, which in turn adds complexity for metadata curators.

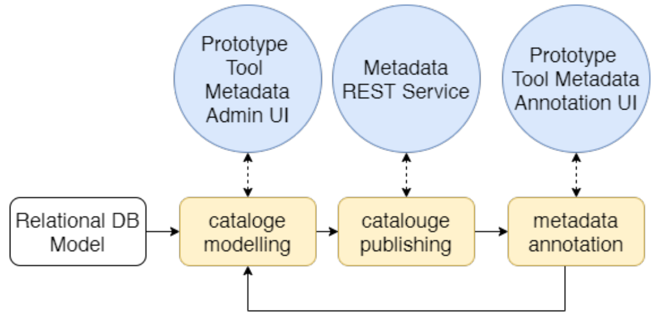

Figure 1: Metadata Management Workflow

4 DESIGN AND IMPLEMENTATION

Based on our model of system users and their requirements we designed a data model covering the derived use cases. We implemented it as a relational model that preserves relationships between the different curricula, stores similar terms and meanings, and allows ongoing changes and further development of the metadata catalog.

Figure 1 shows the workflow which is supported by suitable prototypical tools. The catalog modeling process is supported by a prototypical user interface implemented in Microsoft Access (using ODBC), which then writes the data to our relational data model. These are then published with a JSON Schema Generator Service to create a machine- and human-readable taxonomy, depending on the users local context (i.e. educational region, school type, curriculum, subject). In order to be able to subsequently annotate learning resources according to the schema, a prototype rendering a GUI according to the JSON schema was developed. Figure 2 illustrates the different components of the designed system.

4.1 Metadata model

The metadata model consists of seven concepts that allow the description of learning resources with respect to specific local contexts, and the modeling of relations across them.

4.1.1 Educational Region. represents areas with the same education system (law). For example Austria is a single region, whereas Germany is divided into its federal states, each being a distinct educational region.

4.1.2 School Type. A school type (e.g. primary school, middle school, middle school for music, middle school for sports, ...) has a link to a specific educational region. For example in Germany there would be multiple entries for the school type "Realschule" with different links to the relevant educational regions.

4.1.3 Curriculum. Typically a school curriculum defining a whole program for specific school types and educational regions.

4.1.4 Syllabus. Represents a set of coherent learning content. In most school systems a syllabus typically represents a single subject (e.g. Mathematics). A syllabus can be linked to one or more curricula—which is mainly used to keep an unchanged syllabus throughout multiple revisions of a curriculum.

4.1.5 Label Class. Defining the structure of a syllabus. For example mathematics is structured hierarchically in grade, topic, and competence. To allow this hierarchical structure, a label class can have another label class as parent, where one parent can have multiple children. If necessary label classes defining a specific structure can be used in different syllabuses. It is also possible to define different label classes for syllabuses which have the same name and learning objectives but are structured in a different way. This would be necessary if two school types have both English in their curriculum, but the structure is completely different.

4.1.6 Label Value. The actual labels a user can use in the annotation process. A label value is always tied to a specific label class, as illustrated in table 1. Additionally there is a link between syllabuses and label values defining the set of allowed values. For example, a new syllabus would typically allow all label values depending on the referenced label class structure. If, however, a new version of a syllabus is released, that is essentially a subset of the former version, the same overall structure of labels can be reused by removing the links between the new syllabus version and the label values are no longer part of it. For hierarchical structures of labels it is necessary to provide information about dependencies of label values on *parent* label values. Therefore a $n : m$ relation on the label value entity itself exists. For example for the label class *Topic* a value *fractions* is only allowed if the parent label is *2nd grade*, but not if it is *1st grade*.

4.1.7 Label Value Group. To model semantic similarity and group similar label values and synonyms the *Label Value Group* can be used. It aggregates label values across curricula.

4.2 Metadata Workflow and Prototypical Information System

The workflow consists of three steps:

(1) *Modeling the Metadata Catalog.* First, a catalog of metadata labels must be created, which in the case of the school system is based on the respective legal requirements, in order to ensure a later formal accreditation. This catalog ensures that the individual curricula are truthfully represented.
(2) *Publishing the Metadata Catalog.* The metadata catalog must be made easily available to the teachers who use it. Furthermore, it must be able to interact with existing systems using suitable interfaces.
(3) *Annotating Learning Resources.* The labels from the metadata catalog can now be used to enrich learning resources with meta information. This requires tools optimized for usability to support this process.

LabelClass	**LabelValues**
Grade	1st grade 2nd grade
Topic	fractions percentages
Competence	add fractions with common denominator expand fractions to common denominators

Table 1: Label examples for mathematics

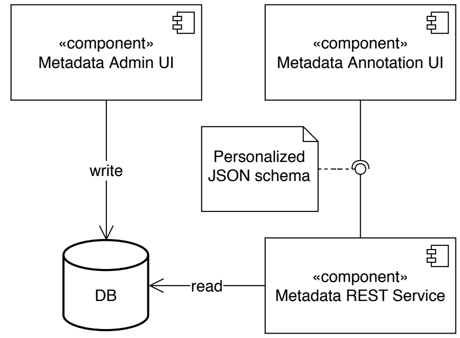

Figure 2: Component Diagram.

4.2.1 Catalogue Modeling. The first step of the metadata workflow is to define the desired metadata and come up with a supporting structure. All relevant data will be stored in a relational database according to the model described in section 4.1. A simple GUI was implemented using Microsoft Access in order to administrate the needed metadata. The basic process for a user when it comes to creating a new syllabus looks like the following:

(1) Create a new curriculum or select an existing one.
(2) Create a new syllabus with the desired name.
(3) Declare a structure by creating/reusing one or more label classes and define relations between them.
(4) Create label values for the relevant label classes and define relations between them.
(5) Optionally the user can exclude individual label classes or label values from the syllabus, especially useful when a new version of the same syllabus doesn't include the exact same data.

4.2.2 Catalogue Publishing. Depending on the preferences of individual users, only the relevant labels should be available for annotation. This will be achieved by tracking a user's behavior and maintaining a usage profile. For publishing the metadata to the annotation GUI, a JSON schema [3] of the relevant content will be generated via a REST service. The advantage of using a JSON schema is that it not only provides a clear human- and machine-readable representation of the data, but also can be used for validation and form rendering. For now, the prototype doesn't use a user's profile to determine the relevant metadata automatically. Instead all desired syllabuses can be selected at the annotation GUI from different curricula. A schema contains definitions for the needed label classes, label values, the dependencies between label values and the syllabuses. The REST service will query all necessary data from the database and return the personalized JSON schema as described in figure 2. The responsible endpoint awaits a list of syllabus IDs in order to build the schema with all it's entities and relations in order to return the result. The generated schema will be used to render the actual annotation form by using an existing open source library [3], containing only the metadata previously selected.

4.2.3 Metadata Annotation. The main part of the annotation GUI is the form generated from the previously explained JSON schema. The annotation form provides the user with the possibility to only select labels from the relevant catalogue. The root select box contains all syllabuses which are part of the generated schema. Depending on the structure/hierarchy of the currently selected syllabus, new selection boxes will be displayed providing the user with further labels for annotating. The result of the annotation is stored in a JSON format and can be used for further processing. The root elements in the result JSON is a list of the annotated syllabuses. Every syllabus contains a list of label values which can contain a list of nested label values themselves. Every label value also has a reference to the belonging label class element.

5 RESULTS AND DISCUSSION

While this paper is set out to position our design science project, we can only present episodic qualitative observation-based evaluation results. These results are mainly based on a workshop, held to implement the designed workflow with nine metadata curators, and a three week participant observation of a single metadata curator modeling a curriculum for Austrian Middle Schools. Thus our results only cover the metadata curator perspective and only the catalogue modeling process.

The workshop consisted of a theoretical introduction to the data model, a practical exercise to model a real-world syllabus on paper, and introduction to the Metadata Admin UI. All participants were able to model a syllabus on paper, and felt confident about being able to use the provided user interface. We subsequently spent three weeks observing and supporting a metadata curator with no prior experience (except for the workshop). This was important to ensure that metadata annotators were able to understand the workflow, the data model and the provided tools. The metadata curator modeled a curriculum of the Austrian NMS (middle school) consisting of 19 syllabuses, 48 label classes and 1,164 label values. During this process we collected data to inform our second design iteration through observation and interviews.

The participant found the designed data model to be appropriate and useful. A perceived limitation of the model was the lack of support for multi-dimensional syllabus structures. For example if topics reoccur for each grade and the associated competencies are dependent on grade and topic. We found this to be a common misconception, as the topics in the example were named equivalent while differing in their semantic. In other words, the topic *functions* in the 1st grade and the topic *functions* in the fourth grade are fundamentally different.

From a usability standpoint the participant stated to be able to efficiently model the diverse structures of the different syllabuses in the curriculum. After modeling the curriculum, the participant estimated that it would take about 2 hours to add a new syllabus to the catalogue, of which about one and a half hour were attributed to analysis and conceptualization and 30 minutes to creating the database entries. This also aligns with our observations.

[3]https://github.com/rjsf-team/react-jsonschema-form

6 CONCLUSION AND FUTURE WORK

In this work we presented our metadata model, and a workflow that allows (1) annotators to use their local context when annotating, (2) transcription into other local contexts, and (3) transcription into a generalized metadata representation.

We conducted preliminary evaluations with metadata curators, and plan to thoroughly evaluate our model in future work. So far, we feel that the JSON schema representation that allows to generate personalized metadata schemas depending on the local context was the right technological choice for publishing the metadata catalog. It will be interesting to investigate this solution further in the future with respect to performance in larger-scale environments.

We intend to especially contribute to transferability of metadata from a specific local context (e.g. curriculum) to other local contexts or a more global context. Consequently, we strive to demonstrate the capabilities of our model in that regard, and plan to evaluate its feasibility for cross-curriculum usage in the wild. We plan to include representatives of all identified stakeholder groups in these future evaluations.

Finally, we want to demonstrate how our metadata model can support the transformation of curriculum specific metadata into an LMRI standard representation to enable a broad support for search and recommendation outside the ChabaDoo ecosystem.

REFERENCES

[1] Phil Barker. 2005. What is IEEE learning object metadata/IMS learning resource metadata. *CETIS Standards Briefing Series, JISC (Joint Information Systems Committee of the Universities' Funding Councils)* (2005).

[2] Elisabetta Di Nitto, Luca Mainetti, Mattia Monga, Licia Sbattella, and Roberto Tedesco. 2006. Supporting interoperability and reusability of learning objects: The virtual campus approach. *Journal of Educational Technology & Society* 9, 2 (2006), 33–50.

[3] Michael Droettboom et al. 2015. Understanding JSON Schema. *Available on: http://spacetelescope. github. io/understanding-jsonschema/UnderstandingJSONSchema. pdf (accessed on 14 April 2014)* (2015).

[4] Alan R. Hevner, Salvatore T. March, Jinsoo Park, and Sudha Ram. 2004. Design science in information systems research. *MIS quarterly* 28, 1 (2004), 75–105.

[5] Ludwig Jäger, Matthias Jarke, Ralf Klamma, and Marc T. Spaniol. 2008. Transkriptivität: Operative Medientheorien als Grundlage von Informationssystemen für die Kulturwissenschaften. *Informatik-Spektrum* 31, 1 (2008), 21–29. https://doi.org/10.1007/s00287-007-0218-9

[6] Kotchakorn Jetinai, Ngamnij Arch-int, and Somjit Arch-int. 2016. Ontology mapping and rule-based inference for learning resource integration. *Journal of information and communication convergence engineering* 14, 2 (2016), 97–105.

[7] Ralf Krestel and Peter Fankhauser. 2012. Personalized topic-based tag recommendation. *Neurocomputing* 76, 1 (2012), 61 – 70. https://doi.org/10.1016/j.neucom.2011.04.034 Seventh International Symposium on Neural Networks (ISNN 2010) Advances in Web Intelligence.

[8] R. Navarrete and S. Luján-Mora. 2017. Microdata for Semantic Annotations: Quantitative Analysis of the Deployment of Educational Properties. In *2017 International Conference on Computational Science and Computational Intelligence (CSCI)*. 1830–1831.

[9] M. Peralta, R. Alarcon, K. Pichara, T. Mery, F. Cano, and J. Bozo. 2018. Understanding Learning Resources Metadata for Primary and Secondary Education. *IEEE Transactions on Learning Technologies* 11, 4 (2018), 456–467.

[10] Riina Vuorikari and Rob Koper. 2009. Evidence of Cross-boundary Use and Reuse of Digital Educational Resources. *International Journal of Emerging Technologies in Learning (iJET)* 4, 4 (December 2009), 51–56. https://www.learntechlib.org/p/45147