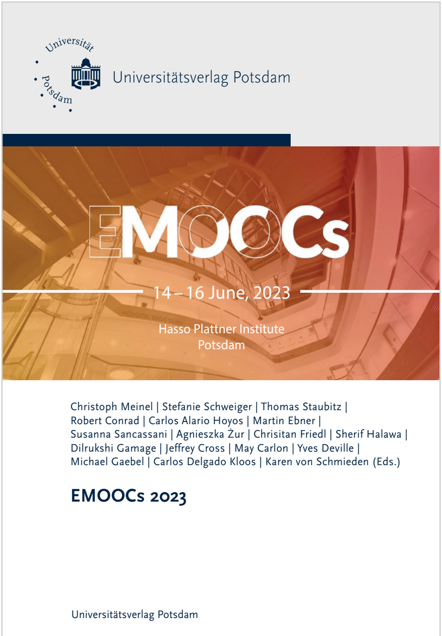
Universitätsverlag Potsdam

Article published in:

Christoph Meinel, Stefanie Schweiger, Thomas Staubitz, Robert Conrad, Carlos Alario Hoyos, Martin Ebner, Susanna Sancassani, Agnieszka Żur, Christian Friedl, Sherif Halawa, Dilrukshi Gamage, Jeffrey Cross, May Kristine Jonson Carlon, Yves Deville, Michael Gaebel, Carlos Delgado Kloos, Karen von Schmieden (Eds.)

EMOOCs 2023

2023 – vii, 350 p.
DOI https://doi.org/10.25932/publishup-57645



Suggested citation:
Max Thomas; Thomas Staubitz; Christoph Meinel: Preparing MOOChub metadata for the future of online learning, In: Christoph Meinel, Stefanie Schweiger, Thomas Staubitz, Robert Conrad, Carlos Alario Hoyos, Martin Ebner, Susanna Sancassani, Agnieszka Żur, Christian Friedl, Sherif Halawa, Dilrukshi Gamage, Jeffrey Cross, May Kristine Jonson Carlon, Yves Deville, Michael Gaebel, Carlos Delgado Kloos, Karen von Schmieden (Eds.): EMOOCs 2023 : Post-Covid Prospects for Massive Open Online Courses - Boost or Backlash?, Potsdam, Universitätsverlag Potsdam, 2023, S. 329–338.
DOI https://doi.org/10.25932/publishup-62483

This work is licensed under a Creative Commons License: Attribution 4.0
This does not apply to quoted content from other authors. To view a copy of this license visit:
https://creativecommons.org/licenses/by/4.0/

Preparing MOOChub Metadata for the Future of Online Learning

Optimizing for AI Recommendation Services

Max Thomas, Thomas Staubitz, and Christoph Meinel

openHPI
Hasso Plattner Institute, Potsdam, Germany
max.thomas@hpi.de | thomas.staubitz@hpi.de | christoph.meinel@hpi.de

With the growing number of online learning resources, it becomes increasingly difficult and overwhelming to keep track of the latest developments and to find orientation in the plethora of offers. AI-driven services to recommend standalone learning resources or even complete learning paths are discussed as a possible solution for this challenge. To function properly, such services require a well-defined set of metadata provided by the learning resource. During the last few years, the so-called MOOChub metadata format has been established as a de-facto standard by a group of MOOC providers in German-speaking countries. This format, which is based on schema.org, already delivers a quite comprehensive set of metadata. So far, this set has been sufficient to list, display, sort, filter, and search for courses on several MOOC and open educational resources (OER) aggregators. AI recommendation services and further automated integration, beyond a plain listing, have special requirements, however. To optimize the format for proper support of such systems, several extensions and modifications have to be applied. We herein report on a set of suggested changes to prepare the format for this task.

1 Introduction

In recent years, the number of Massive Open Online Courses (MOOCs) and MOOC platforms has constantly increased. So-called MOOC aggregators – tools that list the courses of multiple MOOC platforms on the basis of course metadata – such as Class Central [2] or MOOC List [14] are almost as old as the first MOOC platforms. The benefit of these aggregators for the learners is the wider range of offers on similar topics on different platforms. The aggregator serves as a one-stop shop. The benefit for the platforms is the extension of their marketing range to attract new learners.

Metadata is defined as *"data about data"*[1]. Hence, the metadata of a course provides information about the course itself, such as its title, instructor(s), a description, etc. Usually, the aggregators fetch the metadata of the courses from the respective platforms, which are mostly provided via an API. Generally, this process is fully automated to reduce the workload on both sides and make sure the offers are reliably up to date.

To facilitate the automated collection of course metadata, MOOC platforms, and MOOC aggregators had to agree on common metadata formats. These formats are not only useful to deliver the metadata from the platform to the aggregator but also allow the aggregators to implement search algorithms that enable the learners to quickly find courses matching their search criteria. Next to searching, learners can filter the courses efficiently using metadata.

As there are multiple aggregators, there are also many different metadata formats, often developed in parallel with little or no knowledge of each other. This led to inconsistencies among the formats and incompatible systems. Fixing this, and, in the long run, establishing a standardized format, not only in the MOOC domain but more generally for any type of online course, is the aim of several initiatives.

As of today, several standards for metadata formats are on the market. While this does not solve the problem of different formats yet, it helps to reduce their number. Most of the formats in question feature open documentation, which makes it – at least technically – easy to achieve compatibility among the different formats. It is important to note that in this context, we follow the definition of the Cambridge dictionary for the term *standard*: "a pattern or model that is generally accepted"[2], which includes formal standards (standards issued by an organization following a formal standardization process) as well as de-facto standards (a model or pattern becoming a standard by wide usage).

Examples of metadata format standards according to our definition above are DublinCore [6] and schema.org [22]. They provide a very general vocabulary and basic structure for any purpose without a particular focus. As these standards are so general, they are a perfect starting point for the standardization of a metadata format with a more specific purpose, such as the exchange of information regarding educational resources. Namely, the ISO 19788 series of standards (*Information technology – Learning, education, and training – Metadata for learning resources*) directly implements [11] and the *Learning Object Metadata* (LOM) published by IEEE [9] is based on DublinCore. It is noteworthy that LOM is also an underlying standard for SCORM (*Shareable Content Object Reference Model*) and its successor xAPI/Project TinCan [24].

[1] https://www.merriam-webster.com/dictionary/metadata

[2] https://dictionary.cambridge.org/dictionary/english/standard

In the realm of the *schema.org* based metadata standards, two derivates are of particular interest due to their importance in German-speaking countries. First, there is the *Allgemeines Metadatenprofil für Bildungsresourcen* (AMB) [20], which again describes a more general format for educational resources. It serves as the basis for the *Open Educational Resources Search Index* (OERSI) [19]. The second format is the MOOChub format, whose name is derived from the aggregator platform MOOChub [15]. This format has to be supported by all platforms that are interested to be listed in the MOOChub catalog. In contrast to the previously mentioned standards, the MOOChub format is specifically designed to represent MOOC metadata. Additionally, the *Digitale Vernetzungsinfrastruktur Bildung (DVIB)*[3] (formerly known as *Nationale Bildungsplattform* (NBP)) [4] and the Digital.Campus Bayern [3] have committed to support the MOOChub format. Having been involved in the development of this format [7] and due to its proven usefulness and widespread acceptance, we decided to enhance it and prepare it to also support the needs of AI-driven recommendation engines and learning path assistants.

2 Development of an AI-ready course metadata format

As previously mentioned, different standards for the exchange of metadata are available and in use. Furthermore, these standards range from very general to very specific formats. While the general standards often only set a basic framework, the specific standards build on top of the general ones, modifying them for their special purpose. This, on the one hand, guarantees compatibility at least on the very fundamental level of the general standard, and, on the other hand, allows to have a format that exactly fits the needs of its application.

The MOOChub format is a good starting point for the next steps in the evolution of metadata formats for online courses. It already covers the very specific needs of platforms and aggregators and specifically implements many features that are needed for a comprehensive description of a course. Furthermore, the changes and modifications do not touch the underlying standards and, hence, do not affect the basic level of compatibility.

So, why is it necessary to modify and change things, if the format exists and is widely accepted in its target area? The answer is that many of the surrounding conditions have changed. During and after the pandemic the need for (high-quality) online educational resources has increased massively [21]. Not only the demand for these resources has increased but also the supply. For example, the offer of available MOOCs has increased again after a period of moderate decline [23]. Furthermore,

[3] only available in German

the awareness of the importance of life-long learning is spreading among the responsible parties in industry and academia [8]. Particularly, *re-skilling* and *up-skilling* are important topics, particularly for the industry [17]. In this context, recommendation services and learning path assistants are proposed as valuable support tools for the learners to help them to navigate through the jungle of offers [12]. Particularly, the huge improvements in the area of artificial intelligence (AI) allow new approaches here [1]. Several publicly funded research projects are working on smart solutions for recommendation services and assistants. For the German-speaking countries, we need to name here the DVIB and *MarkEtplace foR LifelOng educaTional dataspaces and smart services provisioning* (MERLOT) [13]. These AI-driven systems, however, have different requirements and need to be supplied with more detailed metadata than it was required for simple search and filter algorithms, as we learned from the projects above.

When we talk about recommendation systems, we mean a solution that suggests courses to a learner with comparable content. This can include courses on the same level but also more or less difficult ones. More sophisticated recommendation systems might also take the learner's data into account. E.g. a learner dropped out of a certain course as it was too difficult. The system will suggest courses on a lower level. Or if the learner provides information about his or her linguistic proficiency, only courses in the respective language(s) will be suggested.

Learning path assistants (LPA) go even further. Not only a set of similar courses will be recommended but a whole series of courses of different content and difficulty levels to enable learners to go on a personalized education journey. Advanced training and retraining for new challenges in the job or a complete career change can be represented with a learning path provided by the assistant.

The MOOChub format currently lacks this particular information for recommendation services and learning path assistants. We identified a particular need in the areas: *field of study/topic*, *competency level*, *required competencies*, and *model interests of learners*.

We created a draft [5] of a revised MOOChub format on the basis of our observations. The revised format will be discussed in the near future to become the successor of the current format. The draft is published and open for comments. The above-mentioned missing information can be given in our proposed format. We will give insights into how we designed the data fields in the following. The fields are:

- Field of study (subsection 2.1)
- Competency level (subsection 2.2)
- Required competencies (subsection 2.3)
- Model of interests of learners (subsection 2.4)

2.1 Field of Study

The recommendation systems and the LPA both depend on reliable information about the field of study because it provides information about the course content. With this information, similar courses can be identified by recommendation systems. The LPA, on the other hand, will need this data to choose courses with topics lying on the learner's path.

For all fields, it is crucial to be machine-readable, because of their intended usage in AI applications. Therefore, the vocabulary must be standardized. And there are many standards describing fields of study (e.g. *International Standard for Classification of Education – Fields*: ISCED-F [10] or *Fields of Science and Technology*: FOS [18]) – maybe even too many. Hence, one of the challenges is to agree on a set of standards to be used. To that end, we will take a flexible approach that is first limited to only two standards and allows us to add further standards when needed by our partners.

For the implementation, it will be mandatory to provide the name of the field of study and the name of the standard or framework used. There will also be the possibility to add additional data, like a shortcode or alternative names. We used schema.org *EducationalAlignment* objects with slight modifications (a shortCode field was added) to facilitate that.

As an example, an excerpt from a JSON file is given below (based on schema.org/EducationalAlignment):

Listing 1: JSON example for educationalAlignment

```
"educationalAlignment": [
    {
        "alignmentType": "educationalSubject",
        "educationalFramework": "ISCED-F",
        "url": "http://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-fields-of-education-and-training-2013-detailed-field-descriptions-2015-en.pdf",
        "name": [
            {
                "inLanguage": "en",
                "name": "Computer use"
            }
        ],
        "alternateName": [
            "use of computers",
```

```
            "working with computers"
        ],
        "shortCode": "0611",
        "targetUrl": null,
        "type": "EducationalAlignment",
        "description": "Computer use is the study of using
    computers and computer software and applications for different
    purposes. These programs are generally of short duration.
    Programs and qualifications with the following main content are
     classified here: Computer use Use of software for calculating
    (spreadsheets); Use of software for data processing; Use of
    software for desktop publishing Use of software for word
    processing Use of Internet"
    }
],
```

2.2 Competency level

Another information of tremendous impact on the choice of a course or its selection by an AI is the competency level or difficulty of a course. A course might be too easy or too difficult for certain people depending on their background and education. So this information can be used by the recommendation systems and LPAs. Recommendation systems can react to the learner's data (if provided) and LPAs can arrange courses about a certain topic by increasing difficulty. The recommended courses shall meet the learner's level and course series have to be ordered correctly regarding their difficulty. A rudimentary orientation like "beginner", "advanced", and "expert" (or similar) is good enough for manual scanning of the courses by learners but is certainly not sufficient for AI-driven services according to our partners designing such systems. AI will need a much more detailed set of competency levels following clear and standardized definitions of the levels.

As for the fields of study, it is necessary to implement the competency level attribute in a way that allows giving information about the level name and the used framework together with additional data. In contrast to the field of study, there is no final decision on which frameworks to implement. It is subject to current investigations. The *educationalLevel* from schema.org gave us the chance to use a self-defined term. Our DefinedTerm object provides the following fields to fit the requirements of an AI-based service:

- name: The name of the educational level according to the framework used (can be given in several languages)

- educationalFramework: The educational framework used
- shortCode: A shortcode for the educational level as provided by the framework
- alternateName: A list of alternative names for the educational level
- description: A description of the educational level
- url: An iri (internationalized form of uri, allows Unicode char set) pointing at the document describing the framework
- targetUrl: An iri pointing at a web node representing the educational level
- type: Labels the object as an "EducationalLevel"

2.3 Required competencies

Learning paths and recommendations of courses strongly rely on what a learner already knows and which competencies he or she has. It does not make sense to recommend a course about AI, which expects fundamental programming skills, to a learner without any experience in writing computer programs. A recommendation system can take this into account and filter only for courses, in which the prerequisites are fulfilled by the learner. An LPA, on the other hand, uses this information to build a learning path, since it needs to know, which other courses a learner has to take beforehand to another course. In other words, creating useful learning paths is only possible for an LPA, if the course prerequisites are known.

To implement this, schema.org delivers a *competencyRequired* field. As for the competency level, schema.org allows for a DefinedTerm here, too. The fields of our proposed DefinedTerm are the same as for the *educationalLevel* except that the *educationalLevel* itself can be given according to our definition above.

2.4 Model of interests of learners

Besides these hard-defined data, there is also the question of what the learner wants and what the learner's interests are. One approach to get this information is to ask the learners, but sometimes the learner him- or herself does not know, either.

Different projects trying to create recommendation services and learning path assistants use surveys, questionnaires, or conversations with chatbots to find out the interests of a learner. To score the interests some projects and organizations use the RIASEC model (also known as Holland Code) [16].

schema.org does not provide any suitable fields or attributes to represent the RIASEC model. That is why for this case we added an extension to the original schema.org. This does not cause any harm since an extension does not interfere with the original standard.

We propose a simple array of strings here. The input values to the array are limited by an enum to the respective letter representation of the category within the model. With this field, we can now tag courses to map them better to the learner's interests.

3 Conclusion

To sum up, we have shown that the MOOChub format is a suitable starting point but has still some drawbacks when it comes to providing a data model for training AI-based recommendation services and LPAs. We proposed the development of a metadata format for courses that fills the gaps and allows better consumption of the provided data by AI-driven services on this basis.

We have demonstrated how the addition of a manageable number of further fields to the original format can enhance machine-readability and thus the usage in AI-based services. It is our strong belief that our metadata format will greatly support recommendation systems and LPAs. However, we are aware that the evolution of education is not finished by any means. We will have a close eye on current developments to further enhance our format. Thus, the development of our new format is not the end but rather the next iteration in the development of AI-ready metadata formats for MOOCs.

Acknowledgment

Funding by the BMBF (German Federal Ministry of Education and Research) via the project HPI4NBP (16INB2029) is gratefully acknowledged.

References

[1] W. Bagunaid, N. Chilamkurti, and P. Veeraraghavan. "AISAR: Artificial Intelligence-Based Student Assessment and Recommendation System for E-Learning in Big Data". In: *Sustainability* 14.17 (Jan. 2022), page 10551. ISSN: 2071-1050. DOI: 10.3390/su141710551.

[2] *classcentral*. 2011. URL: https://www.classcentral.com/ (last accessed 2016-10-01).

[3] *Digital.Campus Bayern*. URL: https://www.stmd.bayern.de/760-000-euro-fuer-qualifizierungsplattform-digital-campus-bayern-gerlach-mit-online-fortbildungen-den-digitalen-buergerservice-verbessern/ (last accessed 2016-10-01).

[4] *Digitale Vernetzungsinfrastruktur Bildung (DVIB)*. URL: https://www.bildungsraum.de/ (last accessed 2016-10-01).

[5] *Draft MOOChub API metadata schema V3*. URL: https://openhpi.stoplight.io/docs/draft-proposal-moochub-api-v3/ufqbpe1u9ywag-course (last accessed 2016-10-01).

[6] *DublinCore*. URL: https://www.dublincore.org/ (last accessed 2016-10-01).

[7] M. Ebner, M. Koschutnig-Ebner, F. Rampelt, S. Serth, T. Staubitz, A. von Stetten, M. Thomas, and A. Wittke. "Metastandard für den internationalen Austausch von MOOCs – der MOOChub als erster Prototyp". In: *Zeitschrift für Hochschulentwicklung*. Massive Open Online Courses und ihre Rolle in der digitalen (Hochschul-)Lehre 18.1 (Mar. 2023). Edited by M. Ebner, T. Staubitz, M. Koschutnig-Ebner, and S. Serth, pages 17–35. DOI: 10.3217/ZFHE-18-01/02.

[8] R. Eynon and L.-E. Malmberg. "Lifelong learning and the Internet: Who benefits most from learning online?" In: *British Journal of Educational Technology* 52.2 (2021), pages 569–583. ISSN: 1467-8535. DOI: 10.1111/bjet.13041.

[9] *IEEE Standard for Learning Object Metadata*. Standard. Washington, US: IEEE Computer Society, Nov. 2020.

[10] *Interational Standard Classification of Education - Fields of education and training 2013 (ISCED-F 2013) Detailed field descriptions*. Standard. P.O. Box 6128, Succursale Centre-Ville, Montreal, Quebec H3C 3J7, Canada: UNESCO Institute for Statistics, 2015.

[11] *Information technology — Learning, education and training — Metadata for learning resources — Part 2: Dublin Core elements*. Standard. Geneva, CH: International Organization for Standardization, Jan. 2011.

[12] S. S. Khanal, P. Prasad, A. Alsadoon, and A. Maag. "A systematic review: machine learning based recommendation systems for e-learning". In: *Education and Information Technologies* 25.4 (July 2020), pages 2635–2664. ISSN: 1573-7608. DOI: 10.1007/s10639-019-10063-9.

[13] *MarkEtplace foR LifelOng educaTional dataspaces and smart services provisioning (MERLOT)*. URL: https://merlot-education.eu/en/homepage/ (last accessed 2016-10-01).

[14] *mooc-list*. 2012. URL: https://www.mooc-list.com/ (last accessed 2016-10-01).

[15] *MOOChub*. URL: https://moochub.org/ (last accessed 2016-10-01).

[16] M. M. Nauta. "The development, evolution, and status of Holland's theory of vocational personalities: Reflections and future directions for counseling psychology." In: *Journal of Counseling Psychology* 57.1 (2010), pages 11–22. ISSN: 1939-2168, 0022-0167. DOI: 10.1037/a0018213.

[17] OECD. *Getting Skills Right: Future-Ready Adult Learning Systems*. Paris: Organisation for Economic Co-operation and Development, 2019.

[18] *Revised Fields of Science and Technology (FOS) Classification in the Frascati Manual*. Standard. OECD - Working Party of National Experts on Science and Technology Indicators, Feb. 2007. DOI: 10.1787/9789264239012-en.

[19] *Open Educational Resources Search Index (OERSI)*. URL: https://oersi.org/ (last accessed 2016-10-01).

[20] A. Pohl, A. Klinger, B. Hartmann, C. Schuurbiers, F. Steeg, M. Oellers, M. Stengel, M. Hoffmann, S. Rörtgen, S. Kulla, and T. Bülte. *Allgemeines Metadatenprofil für Bildungsressourcen (AMB)*. May 2023. URL: https://dini-ag-kim.github.io/amb/draft/ (last accessed 2016-10-01).

[21] F. Reimers, editor. *Primary and secondary education during Covid-19: disruptions to educational opportunity during a pandemic*. OCLC: 1269494853. Cham: Springer, 2022. ISBN: 978-3-030-81500-4.

[22] *schema.org*. URL: https://schema.org/ (last accessed 2016-10-01).

[23] D. Shah, L. Pickard, and R. Ma. *Massive List of MOOC Platforms Around The World in 2022*. Jan. 2022. URL: https://www.classcentral.com/report/mooc-platforms/ (last accessed 2016-10-01).

[24] *xAPI/Project TinCan*. URL: https://xapi.com/ (last accessed 2016-10-01).