Suggested citation:
Thomas Staubitz; Sebastian Serth; Max Thomas; Martin Ebner; Markus Koschatnig-Ebner; Florian Rampelt; Alexander von Stetten; Andreas Wittke: A metastandard for the international exchange of MOOCs, 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. 147–161. DOI https://doi.org/10.25932/publishup-62415

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/
A Metastandard for the International Exchange of MOOCs
The MOOChub as First Prototype

Thomas Staubitz¹, Sebastian Serth¹, Max Thomas¹, Martin Ebner², Markus Koschutnig-Ebner², Florian Rampelt³, Alexander von Stetten⁴, and Andreas Wittke⁵

¹ Hasso Plattner Institute, University of Potsdam
Prof.-Dr.-Helmert-Str. 2–3, 14482 Potsdam, Germany
² Graz University of Technology
Münzgrabenstraße 36/1, 8010 Graz, Austria
³ Stifterverband
Tempelhofer Ufer 11, 10963 Berlin, Germany
⁴ Virtuelle Hochschule Bayern
Luitpoldstr. 5, 96052 Bamberg, Germany
⁵ Technische Hochschule Lübeck
Mönkhofer Weg 239, 23562 Lübeck, Germany

The MOOChub is a joined web-based catalog of all relevant German and Austrian MOOC platforms that lists well over 750 Massive Open Online Courses (MOOCs). Automatically building such a catalog requires that all partners describe and publicly offer the metadata of their courses in the same way. The paper at hand presents the genesis of the idea to establish a common metadata standard and the story of its subsequent development. The result of this effort is, first, an open-licensed de-facto-standard, which is based on existing commonly used standards and second, a first prototypical platform that is using this standard: the MOOChub, which lists all courses of the involved partners. This catalog is searchable and provides a more comprehensive overview of basically all MOOCs that are offered by German and Austrian MOOC platforms. Finally, the upcoming developments to further optimize the catalog and the metadata standard are reported.

1 The MOOChub

The MOOChub provides a centralized, searchable catalog of all courses on German and Austrian MOOC platforms. The dedicated goal of the MOOChub is to provide learners with a cross-platform overview of all available course offerings and enable
them to find suitable courses for their learning needs, regardless of the MOOC provider.

The genesis of the MOOChub goes back to two almost parallel activities. In 2015, in discussions between the TH Lübeck (MOOC platform oncampus, formerly mooin) and the TU Graz (MOOC platform iMooX) the idea emerged to mutually list each other’s MOOCs. Both platforms were dedicated to Open Educational Resources (OER) and free accessibility of online courses, so they saw a good opportunity to increase mutual awareness by listing each other’s offerings on their websites. They also experimented to offer an online course simultaneously on both platforms [10,11]. In parallel, the Hasso Plattner Institute’s MOOC platform openHPI and SAP’s MOOC platform openSAP set up a project named mammoc to combine the growing range of courses on an increasing number of world-wide MOOC platforms. Their vision was to offer learners a one-stop shop to search for new learning opportunities and centrally manage their course enrollments across various MOOC providers. In addition to courses from openHPI and openSAP as the directly involved platforms, MOOCs from other providers (such as Coursera, edX, or FutureLearn) were also included in the mammoc course catalog via proprietary interfaces. Finally, in cooperation with the TH Lübeck, the provider-independent exchange format for course metadata, which was initially developed as part of the mammoc project, premiered in early 2017.

Based on these preliminary projects and considerations, all major German-language MOOC platform operators first met in 2020 to discuss options for a centralized platform catalog. In a regular exchange, they discussed how learners could be provided with a common, cross-platform offering. This endeavor was not entirely trivial for several reasons: the platforms have different operators and operating models, different target groups, different course topics and, above all, different technical infrastructures. Therefore, the most important objective was to create a common understanding and define possible goals of joint activities. As a result, the founding members – namely the TH Lübeck (platform mooin/oncampus), the Hasso Plattner Institute (platform openHPI), the Stifterverband (platform KI-Campus), the Virtual University of Bavaria (platform OPEN vhb) and the TU Graz (platform iMooX) – agreed to operate an aggregator platform on which the free, open online course offerings of all partners are listed. Gradually, this aggregator was put into practice and is now operated under the name MOOChub. The TU Graz agreed to cover the technical part of the MOOChub platform development and operation. The Hasso Plattner Institute agreed to take the lead in the further development of the metadata format.

The guiding principles for the MOOChub include:

- Joint regular exchange and cooperation to strengthen education
- Openness of educational offers
• Interoperability, based on the Groningen Declaration
• High quality offerings
• Supporting lifelong learning
• Achieving Bologna Process goals through digital technologies

The following objectives were derived from this:

• Making educational offerings visible by providing an overview of all available online courses (the various MOOC offerings) of the partners
• Common authentication of learners across platforms
• Exchange on standards and frameworks to support the recognition and crediting of digital educational offerings in higher education and vocational fields
• Creation of technical standards, both at the level of the platforms and their interfaces with each other
• Creation of standardized processes and workflows
• Common quality standards for digital education offerings
• Pooling of resources and thus creation of a sustainable educational offers

At the current state, the first of these objectives has already been successfully implemented in the form of the common course catalog with the MOOChub platform. For this purpose, it was necessary to develop a common course catalog metadata format for the description of the online courses. In this publication, we want to present this course catalog format as well as its development and application. Finally, we discuss the experiences so far and point out possible further steps.

In addition to the MOOChub and the associated MOOC platforms using the course catalog format for the data exchange, other providers as well as aggregators were invited and encouraged to use and contribute to this format. In addition to the founding members previously mentioned, the course offerings of the openSAP, LERNEN.cloud and eGov-Campus platforms, which are also free and openly accessible, were added to the MOOChub catalog. On the aggregator side, the Open Educational Resources Search Index (OERSI) and the Kompetenznavigator Schleswig-Holstein use the course catalog metadata format to list courses

[8] https://www.findig.sh/
from the various MOOC platforms in their catalogs. OERSI is operated by the library of Leibniz University Hannover and the university library center of North Rhine-Westphalia, Germany. The “Kompetenznavigator” is a project of the Administration Laboratory at the “Zentrale Einrichtung für Angewandte Forschung” (ZEAF) of the University of Applied Sciences for Administration and Service in Schleswig-Holstein, Germany. Further discussions with other providers and aggregators are held on a regular basis and we are always open for an exchange with interested parties. Most recently, the format has also been approved as one of the metadata exchange formats that will be used by the new “Nationale Bildungsplattform” (NBP), a project funded by the German Ministry of Education, MERLOT, a project funded by the German Ministry of Economics, and Digital.Campus Bayern, a project funded by the Bavarian State Government.

2 Existing metadata formats for learning resources

Before we started to develop a common course catalog format for easy delivery of course metadata for MOOC platforms, we explored which existing metadata formats for learning resources were available. The two most widely used standards so far are the IEEE Standard for Learning Object Metadata [16], which is also part of the Sharable Content Object Reference Model [25, 22], and the more modern community-based approach of schema.org[9]. Both standards, however, prove not to be completely suitable for describing online courses and the specific characteristics. In addition to analyzing the standards, we compared how the partners and other established MOOC platforms provide metadata for their courses and which standards they used. It became apparent early-on that although there was no suitable standardized approach so far, there was a great deal of interest among the platforms to collaborate and establish a standardized exchange format.

Early research on a MOOC aggregator identified the difficulties in this regard as early as 2015 and showed the range – from Linked Data to processing the rendered web page – in collecting necessary data [17]. According to the researchers, a proprietary interface for retrieving the course catalogs was only available at a later point in time across all major MOOC platforms [6]. The multitude of different standards and formats in the learning context (and their specific advantages and disadvantages) has been the subject of research even before the increased popularity of MOOCs [19]. To some extent, there are also attempts to establish standards in related areas (e.g. for micro-credentials at the European level; [2]). Sometimes, in-

ternational standards such as LOM are adapted for specific use cases, for example by the “Kompetenzzentrum für interoperable Metadaten” (KIM, [3][21]).

Particularly relevant for the further use of learning materials described by the corresponding metadata, is often their free availability as Open Educational Resources (OER). Projects such as MERLOT[10] therefore have been collecting free educational resources for over 20 years and have been describing the relevance of equally free interfaces for retrieving the metadata [24]. In this context, the support of standards by content providers is considered a prerequisite for building a comprehensive collection of learning materials [5]. As illustrated by the OpenupEd platform, this can also help to counterbalance the leading, mostly U.S.-based content providers at the European level [13]. In addition to the OpenupEd platform (operated by the European Association of Distance Teaching Universities [7]), there were further efforts between 2016 and 2019 with EMMA as European Multiple MOOC Aggregator [1] to collect courses from the European context and offer them simultaneously in multiple languages.

Besides those initiatives focusing on European offerings, other platforms without such a focus were created, listing as many courses from various providers as possible. These usually include advertisements or commercial offers to the listed platforms. When learners enroll in a paid course by clicking on one of the affiliated links shown by the MOOC aggregators, MOOC platforms pay a commission to the respective aggregator. Examples of MOOC aggregators financed by affiliate links and advertisements include Class Central[11] Coursesity[12] and MOOC-list[13]. Therefore, these providers can allocate more resources in developing and maintaining their platform, including an adaption to more formats. With the MOOChub, we envision an ad-free catalogue of OER resources, inviting partners to join our efforts by providing data in an open schema.

3 Collaborative development of a common course catalog format

The development of the MOOChub schema required intensive discussions with all partners and is based on the experiences from the mammooc project [23]. The metadata format[14] developed in this project formed the conversational foundation for
defining the current MOOChub schema for data exchange. Further development involved identifying the existing similarities between the course offerings, working out trade-offs for differences, and considering new developments in the MOOC context. Given the differences between offerings of the involved MOOC providers, the format was defined to include a basic set of mandatory fields on one hand and a set of optional fields on the other. This allows adapting the format to the specific use cases of each MOOC provider. At the same time, though, the format allows for a largely homogeneous presentation of the courses from all providers on the MOOChub website as well as by other aggregators.

The selection and naming of the fields in the MOOChub schema are based on the preliminary work performed by the mammooc project and the existing “Course” standard of schema.org. However, minor changes were made to adapt to the special circumstances of MOOCs, for example by allowing to specify several dates for the course start, deadlines, or availability. Schema.org was introduced in 2011 by Google, Microsoft, Yahoo, and Yandex to create a uniform vocabulary to be used on the web. Further development of the standard is done in an open process with the community. Based on the preliminary work, it was agreed to use the schema.org format as the basis for the MOOChub schema. Given that the MOOChub schema only extends schema.org and as long as these extensions are optional, compatibility with the original standard is guaranteed. The use of such a widely accepted basis facilitated the agreement process among stakeholders immensely. Another advantage of using schema.org as a basis is that this also has a positive impact on search engine rankings. Providers who are only compatible to the original schema.org standard can thus use at least the basic functionality of the MOOChub schema. However, additional features that require special data from the optional fields may be reserved for providers that support the full format.

In theory, as it is often the case, the practical implementation diverged at some point from the theoretical foundation. During the ongoing development and further distribution of the MOOChub schema, elements were included in the implementation that are no longer compatible with schema.org. For this reason, a revision of the format is currently being prepared to restore compatibility with schema.org.

In addition to the technical aspect, it should be mentioned that the chosen “grassroots approach” has proven to be very successful in the implementation so far. The steadily growing number of partners using the course catalog format, both on the MOOC platform side and on the aggregator side, immensely strengthens the argumentation standpoint towards new potential partners. The existing and demonstrably well-functioning format is often gratefully accepted.
4 Course catalog format description

The exchange of metadata for the courses is based on the JavaScript Object Notation (JSON). The MOOC providers make this data available via an Application Programming Interface (API). In this regard, both the use of an API and the JSON format are in line with the current industry standard for metadata exchange. In the JSON format, fields are defined as key-value pairs and nested objects. Here, each key corresponds to a unique, predefined identifier. Depending on the definition, specifying values in accordance with the respective data type is mandatory or can be voluntary. Figure 1 shows an excerpt from the JSON format.

```
36 },
37 "instructors": [
38 {
39 "name": "Prof. Dr. Christoph Meinel"
40 }]
41 "learningObjectives": []
42 "duration": "P2M4D"
43 "partnerInstitutes": []
44 "moocProvider": {
45 "name": "openHPI",
46 "url": "https://open.hpi.de/"
47 }
```

Figure 1: Excerpt from a JSON file with course data according to the MOOChub schema

4.1 Required fields

The use of the MOOChub schema requires participating MOOC providers to provide a mandatory set of metadata for each course. This set includes:

- “name”: the course title. It is stored in the form of a string and is mostly used as a heading in the catalog display.


• “courseMode”: the course type. The course type is specified in the form of a string. Currently, “MOOC” is the only valid value. With an upcoming revision of the MOOChub schema, the list of allowed values will be extended in the future.

• “instructors”: the instructor(s). This is a list of the complex data type “instructor”. This data type was defined specifically for the MOOChub schema and is therefore not compatible with the original standard. We will deal with this and other necessary changes in the chapter “Discussion and Outlook”.

• “moocProvider”: the provider of the course. This is also a separate complex data type.

• “url”: the URL to the course content. A string in the form of a valid Internationalized Resource Identifier (IRI according to RFC 3987, [8]) is stored here.

• “courseLicenses”: the course licenses. The specification is made as an array of the complex, own data type “courseLicense”. By using multiple licenses, different usage scenarios of MOOC providers including multi-licensing are covered.

• “access”: the access modifier. In an array of strings, the modifiers “free”, “paid”, “member-only” and “other” can be specified. This field indicates whether the course is available with or without fee, or if it is only offered to a certain group of participants (such as students enrolled at a university).

This minimum data set is intended to ensure that courses can be represented consistently in the MOOChub catalog. Other aggregators and search services can also rely on the delivery of this data.

4.2 Optional fields

In addition to the mandatory fields that have been listed in the previous chapter, an additional set of metadata can be provided optionally. These fields include, among others:

• “description”: a description of the course content. A string can be specified here to provide a more detailed description of the course. The string can include HTML tags.

• “language”: the language of the course. The language versions of the course can be stored in this array of strings. The language must be specified as an abbreviation according to BCP 47 (RFC 6497, [4]).
Course catalog format description

• “workload”: the estimated weekly course completion time for learners in hours. This field allows an integer value (if known) or “null” as value (if explicitly not known).

The described extensions to the course catalog format help aggregators and learners to search for courses attributes and, therefore, are also a useful addition for learners to find most suitable courses. Hence, it is advisable for MOOC providers to include as much metadata as possible – including the optional fields.

Nevertheless, providers who do not want to or cannot provide all data are not prevented from using the MOOChub schema.

4.3 Data types

For efficient processing of the data, not only the identifiers of the fields but also the data types of the passed values must be standardized. Most fields can simply contain strings. In some cases, however, special specifications must be applied. Dates, for example the start date of a course, can only be passed in a date format according to ISO 8601 (2019) or RFC 3339 [20]. The URL reference to the course homepage has to be delivered as a valid IRI. In addition, simple HTML markup [15] can be used as a markup language, for example in the human-readable description of the course content.

Some fields require complex data types that have been embedded as additional objects in the JSON format. An example of such a complex data type is the specification of the instructor(s) (“instructor”). This field is defined by the instructor’s name (as a string), the instructor’s type (identified as a person or organization), the instructor’s role, an image if applicable (stored as a URL), and a brief description of the person or organization. The “instructors” field itself is created as an array, so that multiple instructors can be added.

4.4 Versioning and pagination

In addition to the specification of the MOOChub schema in JSON format, the API also includes a versioning agreement for the data format of the respective MOOC platform. This ensures that changes in the course catalog format can be implemented independently by the MOOC platforms and the MOOChub, and that the exchange of the metadata automatically adjusts using the compatible versions. To keep the configuration effort as low as possible and to avoid URL customization, version negotiation takes place via HTTP Content Negotiation (RFC 9110, [12]). This allows the MOOChub as a client to specify a preferred version of the course catalog format, which the MOOC platform as a server uses to respond to the request whenever possible. Using this mechanism, different MOOC platforms
can (temporarily) deliver different versions, so that the synchronization efforts required to implement breaking changes to the MOOChub schema can be reduced to a minimum for all involved partners. Optionally, the end of support for a specific API version can be defined using the HTTP Sunset Header (RFC 8594, [26]).

In addition to the versioning described above, the MOOChub schema allows the course catalog to be split across multiple pages when retrieving it through the API. This allows the MOOChub to gather the courses from the individual MOOC platforms in a resource-efficient manner and process them in smaller page sizes. Therefore, we specified that an API should use the JSON API standard [18], which allows linking to subsequent pages directly in the response.

**Figure 2:** Filter options on the MOOChub website

---

Figure 3: Further aggregators using the MOOChub metadata format

5 Discussion and outlook

Once all platforms involved had agreed to cooperate with each other and jointly promote the digital education through MOOCs, the actual technical implementation presented a comparatively small hurdle. As soon as the common course catalog format was established, the self-confidence and the range of arguments vis-à-vis other platforms and aggregators increased with each new partner. The steadily growing interest is underlined by the increasing access figures.

The “grassroots approach” taken to establish this course catalog format more firmly, bit by bit, through a steadily growing user base, has been very successful. With some justification, we can now claim that the format has become a de-facto standard for MOOC platforms in the German-speaking area. In principle, there is little standing in the way of expanding the user base to include other online

---

17Screenshot of the websites [https://oersi.org/](https://oersi.org/) and [https://www.findig.sh/](https://www.findig.sh/) as of March 7th, 2023
course formats. For some time now, discussions have also been taking place at the European level, although there is still a comparatively large need for coordination and harmonization of the various needs. To this end, the major European MOOC providers have recently submitted a joint project application that deals with the creation of a European portal solution comparable to MOOChub. If this application is approved, the development of a European solution based on the current MOOChub schema is very likely. Additional project grants potentially leading to further standardization of the course catalog format are currently evaluated or have already been approved. Among others, the MOOChub schema was recently approved as one of the supported metadata exchange formats as part of the NBP project of the German Federal Ministry of Education and Research. There, particular attention is also being paid to how such a format can serve as a basis for recommendation services and AI-supported learning path creation. Another goal here is also to move the format from the status of a de-facto standard to the status of an official standard in collaboration with the German Institute for Standardization (DIN). Next to the NBP, the Digital Campus of the German federal state of Bavaria has also agreed to use the MOOChub schema to exchange metadata of online courses.

On the technical side, we still have to implement several necessary changes to the MOOChub schema. First, the compatibility with the original schema.org standard needs to be restored. This requires renaming some of the existing fields. While a preview of the updated schema is already available, we expect that the transition of the MOOChub and all partners will be finished by the end of 2023.

In addition to these relatively straightforward adjustments, the requirements for the MOOChub format have expanded. To integrate the courses into learning paths, further information is required, such as thematic keywords and classification of the courses into competence levels. AI-based recommendation services also need such unified course data. Smart algorithms, therefore, must be enabled to create entire learning paths for individuals based on different courses from various providers.

The MOOChub and its described standardization of the course catalog format is, therefore, a necessity to enable modern learning as well as to support learners and teachers in the best possible way in selecting or creating suitable learning opportunities.

Acknowledgment

The paper at hand is based on a paper in German language that has been published in the Austrian journal “Zeitschrift für Hochschulentwicklung” in March 2023 [9].
References


160