Releasing commonmeta-py v0.8

Martin Fenner [ORCID]

Published September 12, 2023

Citation

Fenner, M. (2023). Releasing commonmeta-py v0.8. *Front Matter.* https://doi.org/10.53731/xszpd-6z265

Keywords

News

Copyright

Front Matter

Today I am happy to announce the release of commonmeta-py v0.8, the next major release of the Python scholarly metadata conversion library. There are numerous changes in this release compared to v0.7.1 released in March, in particular:

- Added support for metadata conversions from the JSON Feed and InvenioRDM formats.
- Updated commonmeta JSON schema to v.10.1. The biggest changes are added support for file metadata and contributor roles.
- Many bug fixes and small improvements.

JSON Feed

JSON Feed is a syndication format for blogs and other periodical content and uses JSON instead of XML serialization used by the RSS and Atom formats. The Rogue Scholar blog archive that I started earlier this year makes heavy use of JSON Feed and uses it to convert blog post metadata to Crossref XML and then register DOIs for them. For the about 5,000 DOIs for blog posts that I have registered so far, I used GitHub Actions and the commonmeta-ruby library. As the number of blog posts registered every day is constantly increasing, I need to refactor the Rogue Scholar backend to properly handle that, and I decided to build a dedicated Python API to replace the GitHub Actions workflow. This work will start in October, and adding JSON Feed support to commonmeta-py is an important step.

Files metadata

One big addition in commonmeta v0.10, and supported in the new release of commonmeta-py, is metadata for content associated with a scholarly resource. In the simplest case, this is a direct download link to a publication or software, but it can also mean download links to multiple files each with file size, file type, and checksum. The best implementation is currently the new InvenioRDM commonmeta-py format, but files metadata are also supported in the schema.org format, and partially in DataCite and Crossref formats. Files metadata are particularly important for automated machine access to content, whereas human users are typically first directed to a landing page with links to download content. To properly use this functionality, the content should be available with an open license such as CC-BY, MIT, or CC Zero – licenses have been supported in commonmeta since the first release.

Contributor roles

Authorship of scholarly content has become more complex over the years, with many publications typically requiring multiple authors, often with dedicated roles. The Contributor Roles Taxonomy (CRediT), now hosted by NISO, was started 10 years ago to address this complexity and has been adopted by an increasing number of publishers. One remaining problem is that CRediT was developed for text publications and has limited support for other publication types, e.g. datasets or software. Another problem is the different terminologies used. ORCID use **contributor** and has added support for CRediT in 2021. Crossref uses

Front Matter

contributor and has defined different contributor roles that are different from CRediT. DataCite (based on work in Dublin Core) uses the concepts of **creator** and **contributor**, where creators are *the main researchers involved in producing the data, or the authors of the publication* whereas a contributor is *the institution or person responsible for collecting, managing, distributing, or otherwise contributing to the development of the resource.*

A complex problem, but one important step forward would be to align these different taxonomies in commonmeta. Commonmeta v0.10 has therefore dropped the **creator** property in favor of **contributor**, added support for contributor roles from CRediT, Crossref, and DataCite, and added support for multiple contributor roles. The various metadata formats supported in commonmeta implementations such as *commonmeta-py* can then use a subset of these contributor roles. An example would be the **editor** role which is used in Crossref, DataCite, BibTeX, schema.org, and citation style language (CSL) metadata. Going forward commonmeta can consolidate these roles and add new roles needed for particular use cases and metadata formats, e.g. the **maintainer** role for software used by codemeta.

References

Fenner, Martin. (2023). *Commonmeta-ruby* (v3.0.1) [Computer software]. Zenodo. https://doi.org/10.5281/ZENODO.7752775

Fenner, M. (2021). *First InvenioRDM Long-Term Support (LTS) version released today and Front Matter is joining as a participating partner.* https://doi.org/10.53731/r8c26t1-97aq74v-ag66m

Allen, L., Scott, J., Brand, A., Hlava, M., & Altman, M. (2014). Publishing: Credit where credit is due. *Nature*, *508*(7496), 312–313. https://doi.org/10.1038/508312a

Hosseini, M., Kerridge, S., Allen, L., Kiermer, V., & Holmes, K. L. (2023). *Enhancing Understanding and Adoption of the Contributor Roles Taxonomy (CRediT).* https://doi.org/10.31222/osf.io/n6249

DataCite Metadata Working Group. (2021). *DataCite Metadata Schema Documentation for the Publication and Citation of Research Data and Other Research Outputs v4.4.* 82 pages. https://doi.org/10.14454/3W3Z-SA82