

ePIC Document Database: Elements of Policy

Summary

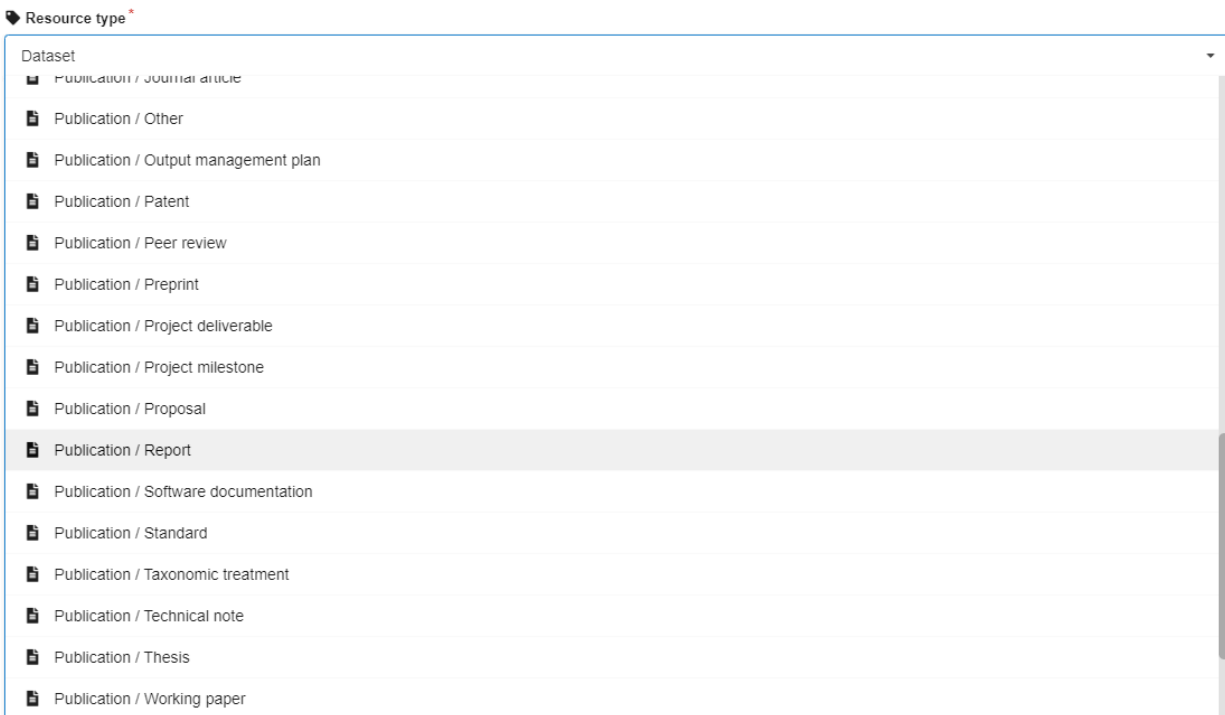
The ePIC Collaboration has established a community on zenodo.org. This document contains notes on current issues and elements of policy to be put in place. There have only been 7 non-void responses to the feedback collection survey so far, 6 of those positive.

Select policy items related to the ePIC community

1. **Joining:** we tested the procedure of joining the community based on local account creation (e-mail), ORCID and GitHub SSO integration. To be investigated – access to the membership data via the REST API that would allow to eventually automate cross checks and other procedures.
2. **Managing:** we tested the roles of the community members – owner, manager, curator and reader. Sending and accepting invitations is a simple process and there is e-mail notification for these actions.
3. **Review and Approval:** zenodo.org is not currently a valid choice for document development. Should we include a preference for the relevant platform into the policy? This could be GitHub (needs discussion).
4. **Limitations of the reader role:** while “readers” have access to restricted documents included in the community, they do not have access to the materials which are not approved yet by the curators. This is in line with zenodo.org being a repository rather than the document development tool and the general logic of the community curation. Possible ways to mitigate:
 - a. Add all ePIC members as “curators”
 - b. Conduct the review process elsewhere, e.g. along the lines of the proposal to the collaboration made two weeks ago

Keywords

1. **Motivation:** augmenting full text/elastic search with keywords has been demonstrated to result in less false positives and can make search more useful
2. **Global ePIC keyword:** to be assigned to any upload, example: “eic-epic”
3. **Semantic keywords:** human-readable tags may be helpful to navigate the ePIC collection on zenodo.org. NB. There is redundancy here since Zenodo already allows one to assign one of many predefined types of uploads to items added to the collection (see below).



We should converge on this element of policy. If we decided to establish semantic keywords, such definitions must be versioned and controlled (preferably on GitHub), and may look like the following examples:

- a. ANALYSIS-NOTE-2024-XXX
- b. PHD-THESIS-2024-XXX

- c. NB. It's much easier on the hands and eyes to use lowercase, e.g. phd-thesis, analysis-note, technical-note etc.

One scenario where such keywords would be useful is this: it would serve as a hint to the group of curators, to assign curation of an item to the correct person(s). The issue of serial numbers – see the section below. Should we start the registry of the keywords now? The semantics keywords need to be designed. Meaningful keywords describing the physics in the paper are pretty much a must.

Conferences must be assigned unique keywords e.g. qm2024. This is easy and makes managing conference materials a breeze.

4. **Serial numbers:** zenodo.org provides unique IDs to every published material. If we need strictly serial numbers in addition to that, this would require additional development/scaffolding e.g. a simple DB. These would logically be associated with the document type (see above). Or, can we have a running counter? That would then be similar to what Zenodo provides by default. NB. The serial numbers have their origins in the requirement for uniqueness, so technically this is redundant now since Zenodo does that already. Example: 10929193.

Misc

1. **Direct publishing by the curators:** the soon to be released feature of zenodo.org which includes an optional setting, which when activated will allow the curators to the direct “push” to the repository w/o review. We need to discuss if we need this – that would be useful in the bulk upload scenarios.