Content and curation policy for the Zenodo-communities (D2.2)

Project	HORIZON-ZEN - EU research programme beneficiary depositing solution in Zenodo.
Work package	WP2
Deliverable no.	D2.2 (Public)
Authors	Lars Holm Nielsen (CERN)
Reviewers	Jose Benito Gonzalez Lopez (CERN) Alexandros Ioannidis (CERN)
Version	1.0
Date	2023-10-31
DOI	https://doi.org/10.5281/zenodo.8419426

Table of contents

Introduction

Document defining the scope of the content and curation level as well as compliance with Horizon Europe contractual requirements.

FAIR metadata
Deliverable *D2.3 FAIR-enabling deposit workflow design* (due M8) will define the discipline-specific metadata requirements and thus this is not considered in this document.

Content and curation policy

Introduction

The EOR-community serves as a repository for research outputs (publications, data, software, posters, presentations etc) which have been funded under an EU research funding programme such as Horizon Europe, Euratom or earlier Framework Programmes.

The community is managed by the European Research Executive Agency (European Commission).

[Zenodo's general policies](#) and [Terms of Use](#) apply to all content.

Scope

The community accepts all digital research objects which is an output result from one of EU's research and innovation funding programmes. The funding programmes currently include:

- Horizon Europe (including ERC, MSCA) and earlier Framework Programmes
- Euratom

In line with the principle *as open as possible, as closed as necessary* both public and restricted content is accepted. See note on how [Zenodo handles restricted content](#).

Content submission

EU programme beneficiaries are eligible to submit content to the community. The community supports two types content submissions:

- Submission via a project community (through User Interface or programmatic APIs).
- Automated harvesting from existing Zenodo content.

Project community

A representative of an EU programme beneficiary may request a project community and invite other project participants as members of the community. The project community is linked to one or more European Commission grants. All records in the project community are automatically integrated into the EOR-community immediately.

Automated harvesting

Records found among Zenodo's existing content will be automatically integrated if they are found to comply with the requirements. The submissions through this method are integrated into the community with delay in a fully automated way.

Descriptive information

Minimal metadata requirements

Records in the community are required to comply with the following minimal metadata requirements:

- **Visibility:** Both public and restricted (with or without embargo and/or access request)
- **Resource types:** All resource types.
- **Creators:** Creators SHOULD be identified with a persistent identifier (e.g. ORCID, GND), and affiliations SHOULD be identified with a persistent identifier (e.g. ROR, ISNI, …)
- **Licenses:** Public and embargoed records MUST specify a license.
- **Subjects:** Records SHOULD specify one or more fields of science from the [European Science Vocabulary](#).
- **Funding information:** Records MUST specify at least one grant from the European Commission.

Review & moderation

All submissions undergo automated curation checks. Submissions through project communities are reviewed by the project community.

Community curators may at any point edit metadata of the records in the community without notice through human or automated processing. The curators may at their sole discretion remove records from the community that are deemed not to comply with the content and curation policy or which are deemed of insufficient quality.

Updates

The content and curation policy is subject to change by the community owner at any time and without notice, other than through updating this page.

Data curation framework

The data curation framework foresees two methods by which content is submitted to the EOR-community:

- Project communities - curated by beneficiaries, content immediately available in EOR-community - use case for projects that want a full community on Zenodo.
- Automated harvesting - existing content deposited in Zenodo is harvested into the EOR-community with a delay - use case for a single record for a project being uploaded on Zenodo.

The two methods support complementary use cases.

Project communities

Scope

A project community serves as a repository for research outputs from one or more grants (allowed grants are defined by the EOR-community). All records in the community MUST be associated with at least one of the grants associated with the community.

The community is managed by the beneficiaries of the grant(s).

Eligibility

Any beneficiary may request a project community for their grant. A grant can only be associated with one project community (i.e. one grant has a designated project community).

A user is considered as a representative of a beneficiary if they have registered with an email address from the beneficiary organization. CORDIS is used as the authoritative source to determine if an organization is a beneficiary.

Lifecycle

A project community has a lifetime linked to the grant duration plus one year. The lifetime may be extended for 12-months on request of the project community's owner(s). The lifetime may also be extended by linking the community to a new grant. The maximum lifetime is 5 years after the project end date.

Once the project community has reached end of life the project community is automatically archived. Archiving means the project community no longer accepts further submission, members can no longer be managed, and settings cannot be changed. An archived community is still publicly visible. The community owner may also manually archive the community.

A project community can be restored on request of the project community owner or the project coordinator.

Note, that the ability to create new versions of a record is not affected by the community being archived - i.e. record owners may create new versions of an existing record.

Members

The community owner is responsible for managing members of the community and granting management, curation and reading rights. The community owner may only grant curation rights to project participants (i.e. representatives from one of the beneficiaries).

In case a specific user e.g. terminates their employment with a beneficiary it's the responsibility of the community owner to remove their access to the project community.

Conflict resolution

In case of conflict between multiple beneficiaries, the project coordinator has the final decision power. A project coordinator can in addition request the project community to be transferred to another community owner.

Misuse

Any misuse of the project community by e.g. submitting content unrelated to the grant of the project, or by granting curation rights to non-beneficiaries will result in the project community being changed to a standard Zenodo-community and records indexed in the EOR-community may be removed. Reinstantion of the project community requires a new request.

Moderation

The creation of a project community is a moderated process. A request for the creation of a project community is reviewed by curators of the EOR-community and subject to their acceptance.

Descriptive information

A project community is associated with one or more grants. Once associated with a grant the following descriptive information is recorded for the project community (from CORDIS):
- Grant
- Subject areas (from European Science Vocabulary).
- Organizations

Automated harvesting

The automated harvesting of content serves the use case where a project only uses Zenodo for some research outputs but not all. The primary challenge for the automated harvesting is that

content is not curated compared to the project communities, and thus is susceptible to misuse by users wanting to e.g. promote their own records.

The automated harvesting will apply the automated curation checks to ensure that records comply with the minimal requirements. Using the existing Zenodo corpus of ~113.000 records in Zenodo we will investigate potential ways to exclude or flag records for manual curation.

Currently the following checks can be applied:

- Publication date is significantly outside the project period.

Compliance with Horizon Europe contractual requirements

EU programme beneficiaries are subject to the Horizon Europe contractual requirements as defined in their grant agreement (see [Annotated Grant Agreement v1.0 Draft - April 2023](#)). *Article 17 Communication, Dissemination and Visibility* defines specific rules in Annex 5 (p. 278 in the Annotated Grant Agreement). For easy reference the text is included in the appendix.

Automated curation checks

The following automated curation checks will be implemented in order to assist users in complying with the Horizon Europe contractual requirements:

- Required the fields: DOI, Creators, Title, Publication date, License, Funding information.
- Creators/contributors SHOULD have a persistent identifier (preferably an ORCID).
- Affiliations SHOULD have a persistent identifier (preferably a ROR).
- Funding information MUST provide at least one of the following options:
    - A link to a European Commission grant from the OpenAIRE database
    - A custom grant linked to the European Commission as funder with the following required information: project title, number, and acronym.
- Related works SHOULD specify related research output (in particular peer-reviewed publications).
- In addition for peer-reviewed scientific publications:
    - Publishing information MUST be provided (i.e. require journal title or book title).
    - Visibility SHOULD be set to public visibility to provide immediate open access.
    - License SHOULD be a CC-BY license (or equivalent). Monographs MAY exclude commercial use and derivative works (CC-BY-NC or CC-BY-ND).
- Research data:
    - Required field: Description
    - If visibility is public, the license SHOULD be CC-BY or CC0 (or equivalent) following the principle as open as possible, as closed as necessary.

FAIR Signposting

Machine-actionability of deposited records is a contractual requirement despite being somewhat vaguely defined in the grant agreement.

Zenodo is already making deposited records machine-actionable through a number of different methods (e.g. embedding of Schema.org JSON-LD metadata in record landing pages - see further details on https://about.zenodo.org/principles/.

As part of HORIZON-ZEN we'll also implement FAIR Signposting Level 2 in line with COAR Next Generation Repositories initiative to increase machine-actionability of deposited records.

Appendix

Article 17, Annex 5 - Specific rules on open science

Extract from the [Annotated Grant Agreement v1.0-DRAFT April 2023](#):

Open Science
<u>Open science: open access to scientific publications</u>
The beneficiaries must ensure open access to peer-reviewed scientific publications relating to their results. In particular, they must ensure that:
- at the latest at the time of publication, a machine-readable electronic copy of the published version or the final peer-reviewed manuscript accepted for publication, is deposited in a trusted repository for scientific publications
- immediate open access is provided to the deposited publication via the repository, under the latest available version of the Creative Commons Attribution International Public Licence (CC BY) or a licence with equivalent rights; for monographs and other long-text formats, the licence may exclude commercial uses and derivative works (e.g. CC BY-NC, CC BY-ND) and
- information is given via the repository about any research output or any other tools and instruments needed to validate the conclusions of the scientific publication.

Beneficiaries (or authors) must retain sufficient intellectual property rights to comply with the open access requirements.

Metadata of deposited publications must be open under a Creative Common Public Domain Dedication (CC 0) or equivalent, in line with the FAIR principles (in particular machine-actionable) and provide information at least about the following: publication (author(s), title, date of publication, publication venue); Horizon Europe or Euratom funding; grant project name, acronym and number; licensing terms; persistent identifiers for the publication, the authors involved in the action and, if possible, for their organisations and the grant. Where applicable, the metadata must include persistent identifiers for any research output or any other tools and instruments needed to validate the conclusions of the publication.

Only publication fees in full open access venues for peer-reviewed scientific publications are eligible for reimbursement.
<u>Open science: research data management</u>
The beneficiaries must manage the digital research data generated in the action ('data') responsibly, in line with the FAIR principles and by taking all of the following actions:
- establish a data management plan ('DMP') (and regularly update it)
- as soon as possible and within the deadlines set out in the DMP, deposit the data in a trusted repository; if required in the call conditions,this repository must be federated in the EOSC in compliance with EOSC requirements

- as soon as possible and with in the deadlines set out in the DMP, ensure open access — via the repository — to the deposited data, under the latest available version of the Creative Commons Attribution International Public License (CC BY) or Creative Commons Public Domain Dedication (CC0) or a licence with equivalent rights, following the principle 'as open as possible as closed as necessary', unless providing open access would in particular:
    - be against the beneficiary's legitimate interests, including regarding commercial exploitation,or
    - be contrary to any other constraints, in particular the EU competitive interests or the beneficiary's obligations under this Agreement; if open access is not provided (to some or all data), this must be justified in the DMP
- provide information via the repository about any research output or any other tools and instruments needed to re-use or validate the data.

Metadata of deposited data must be open under a Creative Common Public Domain Dedication (CC 0) or equivalent (to the extent legitimate interests or constraints are safeguarded), in line with the FAIR principles (in particular machine-actionable) and provide information at least about the following: datasets (description, date of deposit, author(s), venue and embargo); Horizon Europe or Euratom funding; grant project name, acronym and number; licensing terms; persistent identifiers for the dataset, the authors involved in the action, and, if possible, for their organisations and the grant. Where applicable, the metadata must include persistent identifiers for related publications and other research outputs.

Open science: additional practices
Where the call conditions impose additional obligations regarding open science practices, the beneficiaries must also comply with those.

Where the call conditions impose additional obligations regarding the validation of scientific publications,the beneficiaries must provide (digital or physical) access to data or other results needed for validation of the conclusions of scientific publications, to the extent that their legitimate interests or constraints are safeguarded (and unless they already provided the (open) access at publication).

Where the call conditions impose additional open science obligations in case of a public emergency, the beneficiaries must (if requested by the granting authority) immediately deposit any research output in a repository and provide open access to it under a CC BY licence, a Public Domain Dedication (CC 0) or equivalent. As an exception, if the access would be against the beneficiaries' legitimate interests, the beneficiaries must grant non-exclusive licenses —under fair and reasonable conditions —to legal entities that need the research output to address the public emergency and commit to rapidly and broadly exploit the resulting products and services at fair and reasonable conditions. This provision applies up to four years after the end of the action (see Data Sheet, Point 1).

Plan for the exploitation and dissemination of results including communication activities

Unless excluded by the call conditions, the beneficiaries must provide and regularly update a plan for the exploitation and dissemination of results including communication activities.

Example records

Following are examples of existing records highlighting some of the diversity:

- Many grants and many communities: https://doi.org/10.5281/zenodo.7877359
- Dataset with grant link but no community: https://doi.org/10.5281/zenodo.5589597
- Long-lived record, one grant: https://doi.org/10.5281/zenodo.8284953
- Difficult to to check if correct: https://doi.org/10.5281/zenodo.5938877
- Misuse: https://doi.org/10.5281/zenodo.2783994
  - Was linked to OPENAIR - OPtimisation for low Environmental Noise impact AIRcraft (234313) - 2009-2014 but with publication date in 2019.