

Cataloguing in IS-ENES2

Context of cataloguing in IS-ENES2

Summary of IS-ENES2 requirements for cataloguing

Detailed requirements

1. Do you use catalogues or require using catalogues for the following items?

- Observation system
- Data processing system
- Observation event and collected sample
- Data processing event
- Data product

Metadata catalogue for specification of data products.

- Paper or report product
- Research objects or feature interest (e.g. site, taxa, ...)
- Services (processing, discovery, access, retrieval, publishing, visualization, etc.)

2. For each used or required catalogue, consider the following questions:

- Item descriptions:

-> Which fields do you use for describing items?

ESGF: Use MD, in preparation: Citation MD

LTA: Use MD, Citation MD, Contacts, Rights, access&storage, in preparation: provenance.

->Which standards do you apply for these fields (format or standard model)?

ISO, DIF, DC, etc.

->Do you use controlled vocabularies for these fields? If so, please cite the vocabulary service providers.

ESGF: NetCDF-CF (www.cfconventions.org) and lists in central repository (remake is in progress, github.com/ES-DOC/esdoc-cim-cv and ES-DOC/esdoc-cv and other)

LTA: Just internal lists.

->Do you maintain a cross-link or inter-links between:

- Catalogue items (e.g. between observation system - observation event - result dataset)?

Working on cross-link between data and simulation-Metadata, Model-Metadata.

LTA: cross-links to various publications (with DOI).

- Fields for item description and actual items (e.g. between dataset fields - dataset access services or between sample fields - label on the sample)?
- Which repositories/software do you use to manage your metadata?

CMS = plone, cKAN.

LTA: Oracle DB, JavaSP.

ESGF: Lucene indexing, postgres DB.

- Inputs:

->Human inputs: Do you provide/need facilities for editors/reviewers to maintain the metadata in the catalogues (e.g. forms, validation workflow, etc.)? If so, please describe them briefly.

LTA: Oracle SQLdeveloper.

->Machine inputs: Do you use/ need automated harvesting to populate your catalogues? If so, which protocol do you use (e.g. csw, oai-pmh, other, specific)?

ESGF: 1) auto Metadata harvesting from netCDF file headers to DB; 2) Lucene solr cloud for Metadata aggregation and presentation to user; 3) external harvesting from DB possible.

->How do you manage duplicates? i.e. Do you apply governance rules in a network of catalogues, do you use unique identifiers, or take other actions?

Checksums and unique IDs.

- Human outputs:

-> What specific feature is provided/required in your web discovery function (multi-criteria search, graphical selection components (e.g. map, calendar), facets, keyword or natural language)?

ESGF: faceted search.

->Do you evaluate the accessibility, quality, and usage of your catalogue by using a dashboard or value-added products? If so, do you provide/need:

- Popularity or Usage feedback?
- Any other synthesis, indicators or statistics?

No.

If so, please describe them shortly.

- Is the catalogue freely readable or do you apply a specific authorization scheme? If you are applying a specific authorization scheme, please cite the authentication system (SSO) and the authorization rules.

All Metadata are free and open.

4. Machine outputs:

- Do you provide/need machine interfaces for accessing your catalogues? If so, which protocols are implemented?

Metadata: OAI-PMH: ISO, DC, DIF, etc.

- Do you need to fit in Applicable Regulations requirements (e.g. INSPIRE) or embedding frameworks (GEOSS, OBIS)? If so, please cite the regulation, applicable interface requirements, impacts (format, performance, access policy, etc.) on your catalogues.

Partially e.g. to provide Metadata as part of the world data center federation

Formalities (who & when)

Go-between	Yin Chen
RI representative	Sylvie Joussaume < sylvie.joussaume@lsce.ipsl.fr > Francesca Guglielmo < francesca.guglielmo@lsce.ipsl.fr >
Period of requirements collection	Oct -Nov 2015
Status	Completed