

Cataloguing requirements

Introduction

<Insert here a brief introduction to this topic>

<Introduction to the questions asked pertaining to general / pervasive requirements and setting the context of topic-specific requirements. Collation and integration of any pertinent properties and requirements that are consistent across all of the research infrastructures addressed by ENVRIplus requirements gathering.>

<insert here who is responsible for steering and editing this page. But they need to get their go-betweens to agree they have covered the points, e.g. for General requirements>

with help from go betweens and others he co--opts.

Overview and summary of cataloguing requirements

<The overview and summary should be written (integrated and distilled) by the topic leader(s), highlighting commonalities and reporting significant variations. It should be refined and agreed by the go-betweens who contributed to this topic. In particular, they should check that critical points have not been missed and that a balance has been attained.>

Regarding the possible items to be managed in catalogues, the RI have shown interest as follow:

- **Observation systems and lab equipment:** most RI manage equipment which requires management (scheduling, maintenance, monitoring, ...) and some of them are managing or would like to manage this with Information System. Some are already in a standardized approach (OGC/SWE, SSN).
- **Data processing procedures and systems, software:** a very few or none mentioned an interest to support this in a catalogue. However, should be done to manage provenance from product datasets to procedures.
- **Observation events:** not explicitly mentioned as a requirement most of time. However, should be done to manage provenance from observation results to observation systems.
- **Physical samples:** mentioned by a few especially in bio-diversity field.
- **Processing activities:** not explicitly mentioned.
- **Data products or results:** widely mentioned as being done by existing systems (EBAS, EARLINET, CLOUDNET, CKAN, MAdrigal, DEIMS). Widely standardized (ISO191XX). Copliamce required somethmes with INSPIRE. Once with WIS.
- **Publications:** widely mentioned. However very manage the publication on thier own. Links for provenance between publication and datasets is quite common.
- **Persons and organizations:** not explicitly mentioned. However this is reference information which is required for the other described items (datasets, observation systems, ...) and for provenance (contact points).
- **Research objects:** mentioned once as feature of interest (airports for IAGOS).

As a consequence, the following strategy to fulfill ENVRI+ requirements for cataloguing can be proposed. The strategy considers 3 levels of cataloguing.

References catalogues: which are not developed by ENVRI+ or in RI but are pre-existing infrastructures containing reference information to be used. They can be considered as gazetteer, thesaurus or directories... Among them we consider catalogues for:

- persons and organizations,
- publications
- research objects.

Federated catalogues: catalogue which are pre-existing and partly harmonized in RI but need to be federated by ENVRI+. Among them we consider:

- data products or results
- results and observation system and lab equipments. It would be interesting to promote the management of metadata on to improve the provenance.
- physical samples
- data processing procedures, systems and software components metadata management should be promoted to improve the provenance

At last, **activity records** can be considered, observation events, processing activities, usages logs. They should be provided by RI and harmonized at ENVRIplus level to link together the catalogues and fullfil the provenance requirements. However it is very challenging to achieve this fine grained description. However the tracking of usage of datasets in scientific papers is widely mentioned by RI. This activity record can be used harmonized in ENVRIplus and used as a model of what could used for provenance later on. Then first which dataset most contributed to scientific paper, but later on which dataset is most downloaded, which equipment is most used...

Research Infrastructures

The following RIs contributed to developing cataloguing requirements

<Delete from the following list any that were not able to contribute on this topic>

<Add an interest inducing sentence or two, to persuade readers to look at the contribution by a particular RI. e.g., What aspect of the summary of requirements, or the special cases, came from this RI. Check with RIs that they feel they are correctly presented.>

ACTRIS: *<e.g., This RI ... and therefore has XYZ <Topic> requirements, with a particular emphasis on ...>*

Relies on 3 major databases:

- EBAS: for gas
- EARLINET: for lidar observations
- CLOUDNET: for clouds profiles

Has to comply with WMO Information System to integrate GEOSS.

AnaEE:

implements standards for metadata harmonization:

- SSN for sensor systems
- ISO191XX for dataset descriptions

EISCAT-3D:

Relies on following components:

- madrigal system is used
- EPAS database

ELIXIR:

Interested in sample description.

EMBRC:

Interested in lab equipment cataloguing and activities scheduling.

For paper publication: Web Of Knowledge, Scopus

Standards: GBIF and Darwin-Core

EMSO:

Interested in platform, sensor registry management.

Uses Pangea, SeaDataNet, Copernicus Marine for data management.

SWE, ISO19XXX standards are used

EPOS:

CERIF as a internal system for every catalogue aspects.

Metadata standards through CKAN and eGMS services (e.g. dublin-core).

Euro-ARGO:

Central system (JCOMMOPS) for platforms, deployment descriptions.

complies with SWE and ISO191XX standards through seadatanet.

EUROFLEETS2:

complies with SWE and ISO191XX standards through seadatanet.

ESONET: see EMSO

EUROGOOS:

Coordinates RI above network level, does not coordinate inside networks.

FIXO3: see EMSO

IAGOS:

Complies with SWE and ISO19XXX (geonetwork) through AERIS portal.

INSPIRE compliance required

ICOS:

One single integrated Information System provides everything for catalogues

INTERACT:

Tilda component uses CKAN for dataset description

openAIRE is used to publish

INSPIRE compliance required

IS-ENES2:

CKAN is used

ISO191XX, OAI-PMH, DIF and Dublin-Core standard are used

JERICO: see SeaDataNet

LTER:

DEIMS platform

standard OGC/CSW

SEADATANET:

SWE and ISO191XX standards used

geonetwork

SIOS:

ISO191XX standard used