|  |
| --- |
| **Requirements survey topics:**   1. General questions 2. Identification and citation 3. ***Curation*** 4. Cataloguing 5. Processing 6. Provenance 7. Optimization 8. Community support |

**ENVRIplus Theme 2:**

Requirements information gathering exercise

***ICOS (Integrated Carbon Observation System)***

*RI representative(s):*

* *Margareta Hellström,*

*ICOS Carbon Portal & Lund University*

This version is from December 1, 2015.

**2. Curation**

1. Will the responsibility for your RI’s curation activities be shared with other organizations?

No, all data curation needs of ICOS will be handled within the organization - primarily by the Carbon Portal, with support for specific parts (like quality control) by the Thematic Centers. The plan is to build up an ICOS Community Repository based on the Open Archival Information System (OAIS) standard, and to apply for a Data Seal of Approval for this repository. (The latter is being explored as part of ICOS involvement in the EUDAT2020 project.)

1. Does the curation cover datasets only or also:
   1. Software?

Not quite decided yet. But some nodes of ICOS (including the Carbon Portal) are using external repositories like GitHub to manage, store and make accessible the software components that are produced for the RI.

* 1. Operating environment?

Not at this time.

* 1. Specifications/documentation?

Yes. All ICOS-produced documentation, including reports, measurement protocols and similar, will be curated by ICOS.

1. What is your curation policy on retaining/discarding
   1. Datasets?

Datasets that have been “registered” and given a persistent and unique identifier (from e.g. ePIC or DataCite) should never be deleted, only deprecated. The respective landing pages should be updated with information about the reasons why the dataset is deprecated, and give pointers to the superseding version. (Only objects that are found to be completely corrupted due to some technical fault should be considered for actual deletion.)

* 1. Software?

Software versions that have been used to process data (that were subsequently “published”) should never be deleted, but kept for posterity. It can also be noted that some ICOS nodes are applying Docker technology to manage components of the data processing chains, and these could be covered by ICOS curation activities.

* 1. Operating environments?

Not strictly necessary to keep copies of these, but detailed documentation on e.g. required libraries, applied critical updates etc. should be provided.

* 1. Documents?

As with data and software, all documents that are in some way associated with “published” ICOS data products must be retained, even if their contents are superseded by newer versions.

1. How will data accessibility be maintained for the long term? E.g. what is your curation policy regarding media migration?

Storage of ICOS data objects will take place in two different systems. The long-term archiving will be done at external HPC data centers operating the EUDAT B2SAFE service. Any media migration and similar routine operations aimed at ensuring the integrity and sustainability of the stored data objects are outside of ICOS control (but we trust them to do their job properly). “Local” storage, at the Carbon Portal and at the Thematic Centers, will be organized and maintained in as safe a manner as possible, including regular backups etc.

Exactly what happens after the ICOS RI reaches its end is not clear, but ICOS stakeholders have expressed their intention that all ICOS data should continue to be discoverable and accessible to researchers for a foreseeable future.

1. Do you track all curation activities with a logging system?

This is desirable, but not implemented (as we have yet to decide on the specific repository system to use).

1. What metadata standards do you use for providing the following (please supply documentation):
   1. Discovery

We plan to use Dublin Core as a minimum for all ICOS data objects, enhanced as applicable with ICOS-specific information related to data types, parameter types, data content etc.

* 1. Contextualization (including rights, privacy, security, quality, suitability...)

What can be fitted into Dublin Core will be provided there. Else ICOS-specific fields will be added.

* 1. Detailed access-level (i.e. connecting software to data within an operating environment)?

Not decided yet.

1. If you curate software how do you do it? Preserving the software or a software specification?

Current plans are to do both, as appropriate. As an example, all software (code) developed at the ICOS Carbon Portal are stored in GitHub, with full versioning control. This approach is, however, not yet implemented across all of ICOS.

1. What provisions will you make for curating workflows or other processing procedures/protocols?

At the moment, there are no ICOS-wide procedures or protocols in place for curating workflows, as the Thematic Centers have developed their own domain-specific workflow administration systems. But we (the Carbon Portal) are planning to look into possible standards here.

1. If you curate the operating environment how do you do it? Preserving the environment or an environment specification?

Not foreseen at this time.

1. What steps in tooling, automation and presentation do you consider necessary to improve take up of curation facilities and to reduce the effort required for curation?

The primary task is to select a suitable software package for setting up a centralized and common system for ICOS that is to be used for collecting, arranging and storing metadata about all the items that require curation. Currently, the different “data producing” components of ICOS all have their own systems and policies, and these must be brought together under the “hat” of the ICOS Carbon Portal data center. This work is now in progress.