|  |
| --- |
| **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*

* *Dario Papale,*

*ICOS Ecosystem Thematic Centre & University of Tuscia*

* *Benjamin Pfeil,*

*ICOS Ocean Thematic Centre & University of Bergen*

This version from January 27, 2016.

**3. Cataloguing**

**CASE 1: Atmospheric Thematic Center**

The ATC has chosen not to provide specific answers to this part.

**CASE 2: Ecosystem Thematic Center (Dario Papale, University of Tuscia)**

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

Yes, description of all the sensors and their position, characteristics, calibration. We use a system named BADM that is an international standard in the community where each info is a variable.

* 1. *Data processing system*

Yes, we need to catalogue the processing system. It is not yet defined the system, at the moment code versioning and parameters collection are organization are the starting point

* 1. *Observation event and collected sample*

Yes, similar to the observation system and using the BADM

* 1. *Data processing event*

Probably yes, in coordination with the ICOS CP

* 1. *Data product*

As point d

* 1. *Paper or report product*

As point d

* 1. *Research objects or feature interest (e.g. site, taxa, ...)*

Not so relevant

* 1. *Services (processing, discovery, access, retrieval, publishing, visualization, etc.)*

No, it is a task for the ICOS Carbon Portal.

*For EACH used or required catalogue, consider the following questions:*

1. **Item descriptions:**
   1. *Which fields do you use for describing items?*

They are variable dependent, for instrument there are model, serial number, position and then a number of additional info sensor specific

* 1. *Which standards do you apply for these fields (format or standard model)?*

We use the BADM system that has different input formats (excel, web, APP) and are imported in a database structure

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

Yes, impossible to cite, it is variable and parameter dependent, hundreds of options all together.

* 1. *Do you maintain a cross-link or inter-links between:*
     1. *Catalogue items (e.g. between observation system - observation event - result dataset)?*

Yes, the observation system and events are used in the results calculation

* + 1. *Fields for item description and actual items (e.g. between dataset fields - dataset access services or between sample fields - label on the sample)?*

Yes, the BADM system is designed in this way, it is hierarchical so all the info relevant for the same group (sensor, sample etc) are linked.

* 1. *Which repositories/software do you use to manage your metadata?*

Script in C-Sharp, database SQL

1. **Inputs:**
   1. *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.*

Currently yes because we are in a test phase. Not sure if we can build the system in a completely independent and automatic way. We check the error report from the import system and check the reasons in order to better explain to the data providers.

* 1. *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)?*

I think it is not relevant for ETC.

* 1. *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?*

When exactly the same data are submitted they are imported and classified as duplicate. If even just a parameter is different then they are considered a new version and the previous retired.

1. **Human outputs:**
   1. *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)?*

Not relevant for ETC

* 1. *Do you evaluate the accessibility, quality, and usage of your catalogue by using a dashboard or value-added products?*

Not relevant for ETC

**If so, do you provide/need any of the following (please describe shortly as applicable):**

* + 1. *Popularity or Usage feedback?*
    2. *Any other synthesis, indicators or statistics?*
  1. *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.*

Not relevant for ETC

1. **Machine outputs:**
   1. *Do you provide/need machine interfaces for accessing your catalogues? If so, which protocols are implemented?*

They are simple ASCII format

* 1. *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.*

Not relevant for ETC.

**CASE 3: Ocean Thematic Center (Benjamin Pfeil, University of Bergen)**

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

Yes

* 1. *Data processing system*

For parts of the data by now, for all by 2016 when funding for OTC comes in

* 1. *Observation event and collected sample*

For parts of the data by now, for all by 2016 when funding for OTC comes in

* 1. *Data processing event*

For parts of the data by now, for all by 2016 when funding for OTC comes in

* 1. *Data product*

Yes SOCAT and GLODAP where part of the ICOS OTC data is included

* 1. *Paper or report product*

Just for parts of the data that are included in SOCAT or GLODAP

* 1. *Research objects or feature interest (e.g. site, taxa, ...)*

Don’t really know what is meant – but standardized vocabulary is currently being implemented.

* 1. *Services (processing, discovery, access, retrieval, publishing, visualization, etc.)*

For parts of the data by now, for all by 2016 when funding for OTC comes in

*For EACH used or required catalogue, consider the following questions:*

1. **Item descriptions:**
   1. *Which fields do you use for describing items?*

No answer provided.

* 1. *Which standards do you apply for these fields (format or standard model)?*

No answer provided.

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

Yes, NERC Vocabulary Server

* 1. *Do you maintain a cross-link or inter-links between:*
     1. *Catalogue items (e.g. between observation system - observation event - result dataset)?*

By now just for parts of data but will be implemented for the all OTC data

* + 1. *Fields for item description and actual items (e.g. between dataset fields - dataset access services or between sample fields - label on the sample)?*

Not sure what is meant

* 1. *Which repositories/software do you use to manage your metadata?*

By now we use PANGAEA which uses 4d (http://www.4d.com)

1. **Inputs:**
   1. *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.*

We use the SOCAT automation by now which will be adjusted to ICOS OTC needs. All changes (QC comments, flags, metadata edits) are being tracked

* 1. *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)?*

Not yet but will most likely be implemented in the future (oai-pmh)

* 1. *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?*

We use unique cruise identifiers (expocode) for each data set but we plan on using more advanced for each sensor (e.g. SensorML). SensorML will enable us to track changes for other variables as well e.g. temperature and salinity where other communities perform QC. Duplicates are now being sorted by contacting the PI to get the latest version but the feedback from other communities that handle non CO2 parameters is often not reflected. In the future we will have automated processes that will track changes directly

1. **Human outputs:**
   1. *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)? In the future: multi criteria search, spatial and temporal subsetting, facets, standardized vocabs and presets*

No answer provided.

* 1. *Do you evaluate the accessibility, quality, and usage of your catalogue by using a dashboard or value-added products?*

Yes already for SOCAT for ICOS OTC soon

**If so, do you provide/need any of the following (please describe shortly as applicable):**

* + 1. *Popularity or Usage feedback?*

Not yet

* + 1. *Any other synthesis, indicators or statistics?*

Yes for SOCAT, webfeedback and webstatistics

* 1. *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.*

No not yet

1. **Machine outputs:**
   1. *Do you provide/need machine interfaces for accessing your catalogues? If so, which protocols are implemented?*

Just for parts of the data for the OTC that are available through SOCAT we provide OIA-PMH, DIF and ISO

* 1. *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.*

Not relevant

**CASE 4: Carbon Portal view (Margareta Hellström, Lund University)**

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

The components of the ICOS observation systems will be recorded & documented in the metadata catalogue, yes. The aim is that all relevant information (down to instrument types and serial numbers) will be accessible.

* 1. Data processing system

No, at this time there are no plans to use catalogues for storing this type of information.

* 1. Observation event and collected sample

Most ICOS measurements are continuous, but there are also periodic or singular measurements made. Lists of physical samples will be created, with all associated metadata. The results of sample analyses will be treated as data objects.

* 1. Data processing event

No, at this time there are no plans to use catalogues for storing this type of information.

* 1. Data product

Yes. All ICOS data products will be included in a catalogue.

* 1. Paper or report product

Yes. All ICOS-produced reports and articles will be catalogued. ICOS also aims to build up a registry (catalogue) listing all publications that are at least partly using ICOS data.

* 1. Research objects or feature interest (e.g. site, taxa, ...)

A registry of all ICOS observation stations is being set up.

* 1. Services (processing, discovery, access, retrieval, publishing, visualization, etc.)

There are plans to create a catalogue listing all computer-accessible services, including REST-based APIs and similar.

*For EACH used or required catalogue, consider the following questions:*

**NOTE:** Here we are currently only providing one example:

**DATA PRODUCTS**

1. Item descriptions:
   1. Which fields do you use for describing items?

Data objects will primarily be described using Dublin Core v1.1 fields: Title, Creator, Subject, Description, Publisher, Contributor, Date, Type, Format, Identifier, Source, Language, Relation, Coverage, and Rights.

* 1. Which standards do you apply for these fields (format or standard model)?

Generally, ICOS strives to be compliant with ISO19115 for its metadata holdings. The catalogues of data objects will at least expose fields listed in the Dublin Core Metadata Element Set, Version 1.1.

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

This is under discussion.

* 1. Do you maintain a cross-link or inter-links between:
     1. Catalogue items (e.g. between observation system - observation event - result dataset)?

Yes, at least with respect to observational data products. Here there will be clear links from a (published) dataset to the observation station, allowing further lookup all the way to the instrumentation used.

* + 1. Fields for item description and actual items (e.g. between dataset fields - dataset access services or between sample fields - label on the sample)?

No answer provided (unclear question).

* 1. Which repositories/software do you use to manage your metadata?

ICOS is currently building up its own metadata ontology and associated database, based in part on OWL and RDF.

1. Inputs:
   1. 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.

Yes, all catalogues will have interfaces for human interaction with the metadata - including forms for data input and editing.

* 1. 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)?

A large portion of metadata about data objects will be automatically harvested throughout the data life cycle, and uploaded to the catalogue (either event-triggered or via periodic updates). The exact protocols to be used are not yet decided.

* 1. 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?

In the case of “true” duplicates (identical copies, stored at different locations), these will be given PIDs, with a pointer to the “primary object” kept in the replicate object’s handle registry metadata entry as well as in the central ICOS data object metadata database. If a “content copy” (same informational content, but a different file or numeric format) is created, it will be treated as a separate object, given a PID and registered - again noting the PID of the “primary object”. (Updated on 2016-01-27.)

1. Human outputs:
   1. 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)?

The ICOS Carbon Portal discovery function will support advanced searches based on measurement theme (atmospheric, ecosystem and ocean); measurement station (name, country, geographic location); observed or derived parameter (name, unit, ...); parameter type (concentration, flux, meteorological, ...) and date & time interval.

* 1. Do you evaluate the accessibility, quality, and usage of your catalogue by using a dashboard or value-added products?

We will be tracking the usage of the catalogue by e.g. storing all searches in a database. In addition, a user feedback form is being implemented.

**If so, do you provide/need any of the following (please describe shortly as applicable):**

* + 1. Popularity or Usage feedback?

Summaries of selected usage metrics will be available on the ICOS web site.

* + 1. Any other synthesis, indicators or statistics?

Annual reports illustrating more comprehensive summaries of data product and service usage will be provided.

* 1. 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.

It is the intention that all catalogues will be freely readable to all interested parties.

1. Machine outputs:
   1. Do you provide/need machine interfaces for accessing your catalogues? If so, which protocols are implemented?

Yes, the databases will support SPARQL queries returning results in e.g. JSON or RDFS triples.

* 1. 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.

Yes. ICOS is working on becoming fully INSPIRE and GEOSS compliant.