|  |
| --- |
| **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 January 27, 2016.

**0. General questions**

1. What is the basic purpose of your RI, technically speaking?

ICOS (Integrated Carbon Observation System) is a pan-European research infrastructure for observing and understanding the greenhouse gas balance of Europe and its adjacent regions. The major task of ICOS is to collect and make available high-quality observational data from its state-of-the-art measurement stations operated with a long-term (20+ years) perspective. The measurement station networks span three themes - atmosphere, ecosystem and ocean - together, these provide information on greenhouse gas concentrations and exchange, meteorological and other environmental variables. Measurements are carried out on ecosystem sites, in tall atmospheric towers and on oceanic platforms and vessels.

* 1. Could you describe one or several basic use-cases involving interaction with the RI that cover topics 1-7?

ICOS expects to serve users from several different communities and categories, including “experts” (with background in atmospheric, ecosystem, climate and environmental sciences), “scientists” (with background in other fields, like medicine, geosciences, geography etc.), “educational” (teachers wanting to use data in courses, students needing data for reports & theses), “policymakers” (including governmental agencies), “companies” (wishing to use data for services, or interested in developing new measurement techniques), and “general public”. Each of these would have quite different needs and interests.

“Experts”: intensive use of ICOS observational data, potential provider of “elaborated products” (see below), may be interested in making measurements at ICOS sites, may participate in development of measurement practices. Interfaces between these users and ICOS include the measurement stations, experts at the ICOS Thematic Centers, and the ICOS Carbon Portal data center. For these users, easy-to-use data identification & citation information will be very important. Access to complete and relevant metadata, including provenance tracking, will be central to most, thus requiring comprehensive curation. Community support, including access to experts on all aspects of the data collection & processing, as well as services for data visualization & discovery, can also be of great interest.

* 1. Could you briefly describe what data (typology, coverage…) your RI is responsible for?

ICOS produces quality-controlled observational data, as well as results from modelling related to greenhouse gases, and syntheses reports. The data comes from national measurement networks, from our ICOS Thematic Centers and from modelers in the greenhouse gas research community. ICOS data products are used by atmospheric, climatological and environmental researchers for their research and also by policy makers, government agencies and organizations that need to know what is happening in terms of greenhouse gases in order to make decisions for the future. Our data is also available to anyone who wants to use it, for example, for educational or commercial purposes.

When fully deployed, ICOS RI will receive data from 80-100 stations, all operating following highly standardized protocols. These “raw” sensor data from the stations are processed at one of three Thematic Centers (Atmosphere, Ecosystem and Ocean), with support from the Central Analytical Laboratories. Quality-controlled data, aggregated at 30-minute or 1-hour level, are then distributed via the ICOS Carbon Portal data center.

The ICOS processed data products are mostly time series of observed parameter values, for example carbon dioxide concentrations, given for every 30-minute interval. Every observation station delivers such time series for 50-100 parameters, representing the measurements made every day of every year.

The elaborated products are outputs from different models that describe the combined influence of human activities and natural ecosystem processes on the emission and uptake of greenhouse gases from the Earth’s surface. The results are typically in the form of time series of maps, showing the variation in time and space of for example carbon dioxide emissions.

Finally, synthesis reports are summaries of the greenhouse gas balance of the European continent. These are made by combining ICOS results with other information, and are meant to both illustrate overall trends as well as highlight responses to specific climatic events, such as heat waves or droughts.

* 1. Could you describe which data are collected and how these are acquired, curated and made available to users?

The data handled and stored by the ICOS repository will mainly be of three kinds: 1) raw sensor data collected at the measurement stations associated with ICOS RI; 2) aggregated and quality-controlled observational data that are produced by ICOS expert centers based on the sensor data; and 3) so-called “elaborated” data produced by researchers external to ICOS, but based (in part) on ICOS observational data. The latter are typically results from calculations modeling global or regional greenhouse gas budgets.

All relevant ICOS data and ancillary data sets from external sources will be accessible through the facilities of the CP. The CP shall provide a "one-stop shop" for all ICOS data products. As such, the CP is envisioned as a virtual data center, i.e. a place where ICOS data can be discovered and accessed along with ancillary data and where users can post elaborated (value-added) data products that have been derived based on ICOS data. The CP will also have the ability to address all the requirements stemming from these aspects, including i) data security, ii) enforcement of the ICOS data policy and iii) user-friendly (and machine-friendly) internet-based and other computer-network-based interfaces.

The CP is developed based on standard data interfaces to be an integrative access point for all ICOS users, ranging from experts to stakeholders and the general public. The CP supports standardized data exchange protocols and techniques. The CP provides the capability of advanced service composition techniques for web-based distributed processing of ICOS data to generate useful information (e.g. risk maps and integration and analysis with other types of datasets) for research, public users and decision-makers.

In this context the CP will promote and support the production of elaborated products by scientific communities from the ICOS data. These will also be distributed via the CP interface. Finally, organizing the long-term archiving of ICOS data products with the aim to both guarantee safe storage and future access, also after a possible cessation of the Research Infrastructure itself, is an important task of the CP.

* 1. What are the responsibilities of the users who are involved in this use case?

No answer provided.

* 1. What are the responsibilities of the users who are involved in the use case described as answers of 1.a question?

The main responsibility of all users is to follow the ICOS data license (akin to the Creative Commons 4 BY (SA)) and ensure that they acknowledge the data creators accordingly by properly citing the data products (using DOIs as applicable). In addition, ICOS welcomes users to contact the data producers (the principal investigators at the measurement stations and/or the experts at the Thematic Center involved) to provide feedback by discussing any issues with the data sets, such as any questionable or surprising data values, k (or even to extend an offer of collaboration or co-authorship).

1. What datasets are available for sharing with other RIs? Under what conditions are they available?

All ICOS data products are freely available for use by all interested parties, including other RIs. Aggregated “finalized” data sets (30-minute or 1-hourly time series of observations, or the spatial-temporal result data arrays of elaborated products) can be accessed via the ICOS Carbon Portal. Other types of data, including original “raw” measurement data, can be obtained via the Thematic Centers or the principal investigators of the observation stations.

1. **Apart from datasets, is your RI also willing to bring to ENVRIplus and/or other RIs:**
   1. *Software? In this case, is it open source?*

Yes. ICOS has as a goal to use open source software based on free software as far as possible. Copies of ICOS-developed open-source-based tools will be available via the involved party (the Carbon Portal or relevant Thematic Center).

* 1. Computing resources (for running datasets through your software or software on your datasets)?

Yes, but limited. ICOS is indeed planning to set up and maintain the computational facilities required to produce a limited (but consistent) base set of elaborated data products (partially) based on ICOS observations. (This work will be coordinated by the Carbon Portal; the necessary HPC resources may, however, be provided by data centers from EGI and/or EUDAT.) In principle, these facilities could be used at the behest of ICOS-external groups or RIs, which have themselves developed suitable model software, but would like ICOS to run the calculations for them.

* 1. Access to instrumentation/detectors or lab equipment? If so, what are the open-access conditions? Are there any bilateral agreements?

Yes, in principle. Non-ICOS researchers may certainly contact the principal investigator of any ICOS measurement site in order to discuss possibilities to install and operate their own equipment - provided that the site isn’t negatively affected and the ICOS measurement program is not disturbed. Researchers may also contact the Thematic Centers and the Central Analytical Laboratory to enquire about the use of their facilities. However, there is at the moment very little (if any) capacity to handle non-ICOS customers.

* 1. Users/expertise to provide advice on various topics?

Yes, in principle. ICOS representatives participate regularly in summer schools, graduate studies courses and similar events. There is however at the moment very little (if any) capacity at the Central Facilities to handle requests for “technical expert consultancy” from non-ICOS users. As ICOS matures, we may investigate the possibility to set up a more organized user support, for example in the form of FAQs on the web sites (of the Thematic Centers and the Carbon Portal) and possibly open up currently ICOS-internal training opportunities (courses, webinars etc.) also for non-ICOS users.

* 1. Access to related scholarly publications?

The Carbon Portal will host a database of all scholarly publications that have used ICOS data products or otherwise are relevant to ICOS. If these are open access, links to the full texts will be available. Of course all articles that are produced by ICOS itself, including reviews (synthesis reports) on the European greenhouse gas balance, will be made openly available.

* 1. Access to related grey literature (e.g. technical reports)?

These, too, will be curated, catalogued and made openly available in a similar manner as the scholarly publications - either by the thematic Centers or by the Carbon Portal. (Updated on 2016-01-27.)

1. All technical papers on measurement techniques and practices and other relevant reports will be made openly available. What objectives would you like to achieve through participation to ENVRIplus?

Of most relevance here is the active involvement of ICOS in Themes 1 (Technical Innovation) and 2 (Data for Science). Through Theme 1, ICOS will benefit from collaboration on new sensor techniques, on metrology-related issues like quality assessment and harmonization, on finding common technological solutions to enable measurements in extreme environments, and finally to enable close co-operation and joint operation of observational sites. In Theme 2, the desired outcomes include an intensive information exchange with other RIs on relevant issues (e.g. all 7 Theme 2 topics), and the development of tools and best practices that can be directly implemented and used in ICOS data management activities.

1. **What services do you expect ENVRIplus technology to provide your RI with?**

Focusing on Theme 2, at the end of the ENVRIplus project ICOS would like to access and use tools & services in the fields of metadata curation (including “recipes” for cataloging and storage), data object identification & citation, and collection & handling of provenance information.

1. **What plans does your RI already have for data, its management and exploitation?**

In addition to offering ICOS users tools and functionality to discover, search, visualize and download all ICOS data products via the Carbon Portal, ICOS is also planning to set up and operate its own long-term sustainable community data repository (data archive). The ICOS Repository will be based on the Open Archival Information System (OAIS) reference model. In OAIS terms, the repository functionality will include most of the main functions of a data archive: Ingestion, Management and Access, as well as relevant parts of the Administration, Preservation Planning and Management layers. The only function that is envisaged to be (partly) outsourced is the long-term archival storage, which is foreseen to take place at an external trusted data center operating the EUDAT B2SAFE service. However, replicates of raw data will be held at the Thematic centers and most higher level data objects will also be held locally at the CP, in order to support services allowing users fast and efficient search & discovery, visualization and access to ICOS data products.

As part of ICOS participation in EUDAT2020, we are investigating the possibility to attain a so-called Data Seal of Approval (DSA) status for the repository at CP. A first-step self-assessment has been completed (October 2015) and is being evaluated by EUDAT2020 WP2 partners

* 1. Are you using any particular standard(s)?

Currently, the preferred formats for data dissemination in the communities that are seen as the main users of ICOS data are comma-separated (CSV) ASCII text files for time series observational data and NetCDF files for elaborated product (geocoded) multi-dimensional array data. The ICOS Carbon Portal will offer data in these formats, but may also provide other file types on demand. Metadata has traditionally been disseminated in a wider range of formats, ranging from simple text files via spread sheets (like Excel) to data bases. ICOS aims to build up its data object metadata database around semantic technologies, such as OWL and RDF.

* + 1. Strengths and weaknesses

“Simple” formats (like ASCII text files) do not require proprietary software to read and modify, but may not be very storage-efficient. Structured formats, like NetCDF, have wide usage in many communities, and there are therefore several different tools & libraries available for users. More advanced array database formats can be very efficient for machine-operated services, but can be tricky to use for human users and will be therefore only be accessible to users through user-friendly web based interface.

* 1. Are you using any particular software(s)?

The different components of ICOS (Thematic Centers, Central Analytical Laboratory and Carbon Portal) all use a number of software packages, based on a wide range of programming languages and platforms. While Windows and other Microsoft products are widely used for administrative purposes, Linux and associated (free) software is generally the norm for much of the data handling components. However, there is a clearly stated aim that open source software should be used as much as possible throughout the RI. Examples include the Alfresco Community content management platform, Drupal for web site content management, and EddyPro software for flux evaluation.

* + 1. Strengths and weaknesses

Proprietary software may be convenient and widely available (e.g. through various campus licenses at the participating institutes), but there are clearly risks (sudden end of support, obsolete or incompatible versions etc.). Open source solutions can suffer from dwindling user communities and lack of any technical support.

* 1. Are you, as part of a future plan, considering changing any of the following current (please provide documentation/links for all the below which apply):
     1. standard(s)

No.

* + 1. software

Perhaps. For some parts of the system choices are still open, in other parts alternatives will be considered when need occurs.

* + 1. working practices

These are subject of continuous change as the natural evolution of implementing improvements.

1. What part of your RI needs to be improved in order:
   1. For the RI to achieve its operational goals?

More integration of data provenance over the whole infrastructure and streamlining of the metadata exchange

* 1. For you to be able to do your work?

Too early to tell.

1. Do topics [1-7] cross-link with your data management plan?

ICOS doesn’t yet have a comprehensive Data Management Plan as such, but all of the 7 topics are in one way of the other covered in the internal discussions and documentation of the RI. The Concept Papers of the Thematic Centers, Central Analytical Laboratory and Carbon Portal address many relevant DM issues, such as data security, data processing, data storage and data discoverability & dissemination - although perhaps not in a clear-cut way. In addition, the Technical and Scientific Description of ICOS, as well as the ICOS Data Life Cycle documents both outline the tasks, roles and responsibilities of the ICOS components with respect to data management.

* 1. If so please provide the documentation/links

The documents are available on request from the ICOS Head Office (contact Marjut Kaukolehto) and/or ICOS Carbon Portal (contact Maggie Hellström).

1. Does your RI have non-functional constraints for data handling and exploitation? For example any of the following (please provide the documentation/links):

This is a difficult and rather vague question. As with any RI, there is a fixed budget frame which does put limits on what equipment can be purchased, how many people can be employed, how much manpower can be spent on specific projects, etc.

* 1. Capital costs

Capital costs for the Central Facilities of ICOS, including also the Carbon Portal, are to a large extent contributed by the host countries. Most of the investments in hardware required for the data handling in ICOS during the coming 2-3 years have already been made.

* 1. Maintenance costs

Not sure exactly what should be counted here? In any case, issues related to moderate levels of maintenance costs are not expected to have a negative impact on ICOS data handling.

* 1. Operational costs

Operational costs for the Central Facilities of ICOS, including also the Carbon Portal, are to a large extent contributed by the host countries, with a smaller contribution coming from the member countries (based on the number of observation stations). This should guarantee that adequate funds for data handling within ICOS are available at least in a 3-5 year perspective.

* 1. Security

Adequate provisions for securing the collection, computing and storage facilities used within ICOS are being made. We do not foresee security issues providing any specific constraints on the data handling and management of ICOS. However, the issue of Trust is of great importance - users of ICOS data products must be able to trust that these have been produced in compliance with all applicable ethical rules of science, throughout all steps of their life cycle.

* 1. Privacy

The data that ICOS produces are not directly concerned with individuals or with sensitive information of other kinds. However, some of the metadata either contains personal information (names, e-mail addresses and other contact information etc.) related to individuals involved in the data life cycle. There is a possibility that collected data that can be tied to a specific location may be traced back to e.g. local land owners. The potential “conflict” between keeping some of this data internal to ICOS, and on the other hand making as much descriptive metadata available to data users, is however currently not considered as a constraint on ICOS data handling.

In addition, some potentially personal information - visitors’ IP addresses and, for registered users, also contact information, search and download histories etc. - will be collected at the Carbon Portal. The handling of this information must be done securely in compliance with applicable laws and regulations. To minimize the risks of misunderstandings, clear and transparent information about the ICOS privacy policy will be available on the web site.

* 1. Computational environment in which your software runs

No. Adequate resources are available - both “in-house” at the involved partners and through collaborations with external e-Infrastructure providers.

* 1. Access for scrutiny and public review

No, not as far as we can see.

1. Do you have an overall approach to security and access?

Yes and no. The ICOS Carbon Portal is developing a single-sign-on system that will be used (as needed) to control and monitor user identification, authorization and authentication for data and computational resources that require this. Other ICOS components (Thematic Centers) are using systems that are local to their host institutes for these purposes.

1. Are your data, software and computational environment subject to an open-access policy?

The finishing touches are now being made to the ICOS Data Policy and Data License documents, but it is clear that all ICOS data products will be under “open access” (CC4.0 BY), and available to users of all categories. The intention is also to make all software codes that are developed as part of ICOS development and operation openly accessible, under relevant licensing. The software should also, as far as it is possible, be based on free Open Source alternatives and solutions. The computational environments, however, will in general not be open for non-ICOS users, except by individual agreement.

1. What are the big open problems for your RI pertinent to handling and exploiting your data?

There are no “big” problems as such. However, ICOS is still in a developing stage with regards to selecting the appropriate storage systems for its data and metadata holdings. While the long-term archiving is planned to be done at external, EUDAT-operated centers, there are currently a number of different storage platforms for “day to day data handling” in use across the infrastructure. This lack of platform and format homogeneity may have negative impacts on the effectiveness of data and information transfer between partners, not the least with regards to the planned centralized metadata repository.

1. Are you interested in any particular topic [1-7] to discuss in more detail?

ICOS is just getting started, and so does not have ready-to-go solutions for all its data handling & management needs. Currently, we are very interested in learning more about “best practices” and concrete use cases, especially concerning metadata collection and storage, ontologies and vocabularies, and technical solutions for storing and serving (geocoded) data in multi-dimensional array formats.

* 1. If so, would you like us to arrange a follow up interview with more detail questions about any particular topic to be discussed?

No, more interviews are not likely to be useful at this stage. But we would appreciate help to identify experts who are willing to work with us on a consultancy basis!