RINGO | Readiness of ICOS

Readiness of ICOS for Necessities of integrated Global Observations

D6.7

Updated Data Management Plan

RINGO (GA no 730944)

Public Document

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 730944

Deliverable:

Author(s): Alex Vermeulen; Benjamin Pfeil
Date: 21.8.2018
Activity: Report
Lead Partner: ICOS ERIC
Document Issue: 2
Dissemination Level: PU
Contact: Alex Vermeulen (alex.vermeulen@icos-ri.eu) Benjamin Pfeil (benjamin.pfeil@gfi.uib.no)

	Name	Partner	Date
From	Alex Vermeulen Benjamin Pfeil	ICOS ERIC	21.8.2018
Reviewed by	Jouni Heiskanen	ICOS ERIC	21.8.2018
Approved by	Janne-Markus Rintala	ICOS ERIC	21.8.2018

Version	Date	Comments/Changes	Author/Partner
2.0	21.8.2018		

Deliverable Review Checklist

A list of checkpoints has been created to be ticked off by the Task Leader before finalizing the deliverable. These checkpoints are incorporated into the deliverable template where the Task Leader must tick off the list.

• Appearance is generally appealing and according to the RINGO template. Cover page has been updated according to the Deliverable details.	x
• The executive summary is provided giving a short and to the point description of the deliverable.	x
• All abbreviations are explained in a separate list.	x (n/a)
• All references are listed in a concise list.	x (n/a)
• The deliverable clearly identifies all contributions from partners and justifies the resources used.	x
• A full spell check has been executed and is completed.	x

DISCLAIMER

This document has been produced in the context of the *project* Readiness of ICOS for Necessities of integrated Global Observations (RINGO)

The Research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 730944.

Data Management Plan for READINESS OF ICOS FOR NECESSITIES OF INTEGRATED GLOBAL OBSERVATIONS

A Data Management Plan created using DMPonline

Creators: Alex Vermeulen (alex.vermeulen@icos-ri.eu), Benjamin Pfeil (benjamin.pfeil@gfi.uib.no)

Affiliation: Other

Template: European Commission (Horizon 2020)

Grant number: 730944

Project abstract:

RINGO, 'Readiness of ICOS for Necessities of Integrated Global Observations' is a 4-year H2020 project with a total budget of 4,719,680.00 euros. RINGO has 43 partners in 19 countries and consists of 5 work packages with specific emphasis on the further development of the readiness of ICOS Research Infrastructure (ICOS RI) to foster its sustainability. Five principal objectives RINGO's five principal objectives are scientific, geographcical, technological, data as well as political and administrative readiness: 1. Scientific readiness. To support the further consolidation of the observational networks and enhance their quality. This objective is mainly science-guided and will increase the readiness of ICOS RI to be the European pillar in a global observation system on greenhouse gases. 2. Geographical readiness. To enhance ICOS membership and sustainability by supporting interested countries to build a national consortium, to promote ICOS towards the national stakeholders, to receive consultancy e.g. on possibilities to use EU structural fund to build the infrastructure for ICOS observations and also to receive training to improve the readiness of the scientists to work inside ICOS. 3. Technological readiness. To further develop and standardize technologies for greenhouse gas observations necessary to foster new knowledge demands and to account for and contribute to technological advances. 4. Data readiness. To improve data streams towards different user groups, adapting to the developing and dynamic (web) standards. 5. Political and administrative readiness. To deepen the global cooperation of observational infrastructures and with that the common societal impact. Impact is expected on the further development and sustainability of ICOS via scientific, technical and managerial progress and by deepening the integration into global observation and data integration systems.

Last modified: 15-08-2018

Data Management Plan for READINESS OF ICOS FOR NECESSITIES OF INTEGRATED GLOBAL OBSERVATIONS - Detailed DMP

1. Data summary

State the purpose of the data collection/generation

ICOS collects observational data on greenhouse gas concentrations and fluxes in atmosphere, ecosystem and marine enviornments. This data is targeted to scientific use to increase our knowledge of the greenhouse gas cycles and budget of Europe and surrounding regions.

Explain the relation to the objectives of the project

The RINGO project aims at improving the methods and data used and generated by ICOS. The improved methods will increase the quality, amount and FAIRness of ICOS data.

Specify the types and formats of data generated/collected

The RINGO project collects several different type of data.

Some is CO2 ambient mole fraction and ecosystem flux data from the pre-ICOS period that will be reprocessed into higher quality data, very similar to the current ICOS Level 2 data and INGOS datasets for CH4 and N2O. This data is stored in the WMO and Fluxnet community defined Comma and Tab seperated data format as clear ASCII text.

Observation data of vertical profiles of greenhouse gas mole fractions from observations using air cores is stored into a community defined ACSII format, the development, definition and documentation of this processing is part of RINGO.

Raw data is usually also some comma or tab separated ASCII file.

Specify if existing data is being re-used (if any)

Raw instrument pre-ICOS data is used to generate the higher quality Level 2 data as described in the previous section.

Specify the origin of the data

All raw data is generated by approved, pre-ICOS or candidate ICOS instrumentation.

State the expected size of the data (if known)

Raw data volume per site is about 40 MB (ecosystem), 20 MB (atmosphere) or several Kb (marine) per day. The Level 2 data products are about 1-2 MB per year.

Outline the data utility: to whom will it be useful

Scientist in many fields from climatology, biogeochemistry, biology, agriculture, forestry etc.

General public.

Scholars, students.

Policy makers.

2.1 Making data findable, including provisions for metadata [FAIR data]

Outline the discoverability of data (metadata provision)

All raw data and Level 2 data products from RINGO will be published together with the relevant metadata through the ICOS Carbon Portal and follow its FAIR principles and mechanisms. The metadata is exposed through the CP search app, the B2FIND service and in the future other portals of portals (e.g. EOSC; GEOSS)

Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?

All data objects get assigned at ingestion a Persistent Identifier (PID) based on the handle system. This PID contains te sha-256 checksum of the data object and resolves into a landing page that contains the relevant metadata and a link to the data object.

Outline naming conventions used

Each data object has an object specification that links to the metadata that describes the data format and describes the data content ontology (data columns, variables, units). File names are preserved but are considered redundant for machine to machine interpretation and not interpreted. Each community follows its own conventions for file naming to support the internal processing.

Outline the approach towards search keyword

Search keywords will be linked to the data object specification. All metadata elements can be searched for as keyword using the open SPARQL query and the Carbon Portal search app.

Outline the approach for clear versioning

Part of the metadata for each data object is the data version. Each data object of higher version links to its previous (and eventual next) version.

Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how

All metadata is open linked data based on RDF, accesible through an open SPARQL endpoint (W3C standards). Naming convention of the metadata entries follows ISO19115 where possible.

2.2 Making data openly accessible [FAIR data]

Specify which data will be made openly available? If some data is kept closed provide rationale for doing so

All raw and higher level data products are provided following the ICOS data licence which is CC4BY. Raw data (Level 0) objects are not directly downloadable through the web pages, there we ask to contact the thematic center first in order to provide the user with the best service to interpret the data and to inform us of the actual data requirements, but access is allowed and will be given without further conditions.

Experimental data from for example the RINGO project is in principle kept closed until the end of the program, and will be provided open access according the ICOS Data Policy after the project end. The RINGO experimental data is open to project participants.

Specify how the data will be made available

All data is made available through the ICOS Carbon Portal through standard https transfer. The PIDs of the data can be found through the interactive search interface at Carbon Portal or by queries through the open SPARQL endpoint or though other portals (of portals) like the B2FIND or GEOPortal. All PIDs resolve through the Handle (or DOI) system into a landing page that contains all relevant metadata and a link to the actual data object for direct download.

Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?

All metadata and data can be accessed through restful APIs with standard web browsers and internet tools such as wget or curl and javascript or python code. All ICOS Carbon Portal code is open and licensed under GPL v3 through github.

Specify where the data and associated metadata, documentation and code are deposited

Metadata: At Carbon Portal server https://meta.icos-cp.eu as RDF and B2FIND (CKAN, OAI-PMH)

Data: At Thematic Centers, Carbon Portal data service https://data.icos-cp.eu and B2SAFE at CSC (Finland) and KFz Jülich (Germany)

Code: https://github.com/ICOS-Carbon-Portal

Specify how access will be provided in case there are any restrictions

Raw data (Level 0) objects are not directly downloadable through the link on the landing pages, there we ask to contact the thematic center first in order to provide the user with the best service to interpret the data and to inform us of the actual data requirements, but access is allowed and will be

given without further conditions. Access to the raw data is easily obtained through a small manipulation of the link resolved from the PID.

2.3 Making data interoperable [FAIR data]

Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.

Sorry to correct, but interoperability can only be defined in relation to an actual implementation and solution, and is thus by definition dependent on technology and standards. All ICOS data objects are stored in the B2SAFE repository. the associated metadata linked to the data is stored in a RDF database, based on an open ontology based on OWL, that is part of the RDF database. Read-only access to the metadata repository is given through an open SPARQL endpoint. All metadata is also exported to the B2FIND repository where it is also linked with the PIDs of the dataobjects in B2SAFE through CKAN. The B2FIND repository is again linked to the GEOPortal for global access to the metadata form other portals and portals of portals.

The landing pages of the data object will allow for content negotiation to deliver the metadata in the format and vocabulary of the respective community standards. This translation using equivalences will be dynamic and online, will be anchored in the ontology and thus open and easy to maintain and update.

Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?

As described in the previous item, we plan to support all relevant standard vocabularies by mapping the ontologies to the ICOS standard dynamically.

2.4 Increase data re-use (through clarifying licenses) [FAIR data]

Specify how the data will be licenced to permit the widest reuse possible

Data will be provided in general according to the ICOS data Policy and using the Creative Commons Attribution 4.0 International (CC4BY) licence.

Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed

In general the data is available directly after ingestion/generation.

Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why

All data that falls under the CC4BY licence is available to all third parties. RING experimental data results are restricted to the consortium until the end of the project. They will become available within 2 years after the end of the project.

Describe data quality assurance processes

Quality assurance is at the heart of ICOS and the reason of existence of the research infrastructure. The quality assurance procedures are described in the relevant papers and reports that are published by ICOS, the Thematic Centers and their contributors.

ETC Guidelines and instructions: http://www.icos-etc.eu/icos/documents/instructions

ATC ICOS station specifications: https://icos-atc.lsce.ipsl.fr/node/99/27248

ATC Data processing: Hazan, L., Tarniewicz, J., Ramonet, M., Laurent, O., and Abbaris, A.: Automatic processing of atmospheric CO2 and CH4 mole fractions at the ICOS Atmosphere Thematic Centre, Atmos. Meas. Tech., 9, 4719-4736, doi:10.5194/amt-9-4719-2016, 2016

OTC: https://otc.icos-cp.eu/data-levels-quality-access

Specify the length of time for which the data will remain re-usable

ICOS is a long term infrastucture that is foreseen to exist for at least 20-25 years. This would guarantuee operation and data availability until 2040.

3. Allocation of resources

Estimate the costs for making your data FAIR. Describe how you intend to cover these costs

Most of the work to make ICOS and RINGO data FAIR is performed at the Carbon Portal and the ICOS Thematic Centers and concerns at least 50% of the cost of these. Total cost is thus roughly 3 M€ per year.

Hardware costs are about 10 k€ per year. EUDAT services like B2FIND and B2SAFE are about 50 k€ per year.

Clearly identify responsibilities for data management in your project

The data management responsibilities are clearly described in the ICOS Data Policy document and the ICOS Data Lifecycle document. The latter is under continuous development.

Describe costs and potential value of long term preservation

The trusted repositories at B2SAFE from EUDAT at CSC and KFA Jülich will preserve all ICOS data objects for the foreseeable future. B2SAFE will be part of the EOSC service portfolio. Cost is coverered from the ICOS budget that is secured for the long term (>20 years).

4. Data security

Address data recovery as well as secure storage and transfer of sensitive data

Transfer of sensitive data is not applicable. Data is backupped at all individual instances of ICOS, be it stations, experiments, thematic centers and Carbon Portal. All raw data and higher level data is streamed at ingestion to a trusted repository from EUDAT CDI (B2SAFE) that replicates the data over two centers in Europe (in Finland and Germany), that each also provide a full backup. All data objects are identified with a persistent identifier that contains the AE256 checksum of the data for unique identification and consistency check.

5. Ethical aspects

To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former

None

6. Other

Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

Documentation for Atmosphere, Ocean and Ecosystem community, see section 2.4

https://github.com/ICOS-Carbon-Portal/meta