

# RINGO

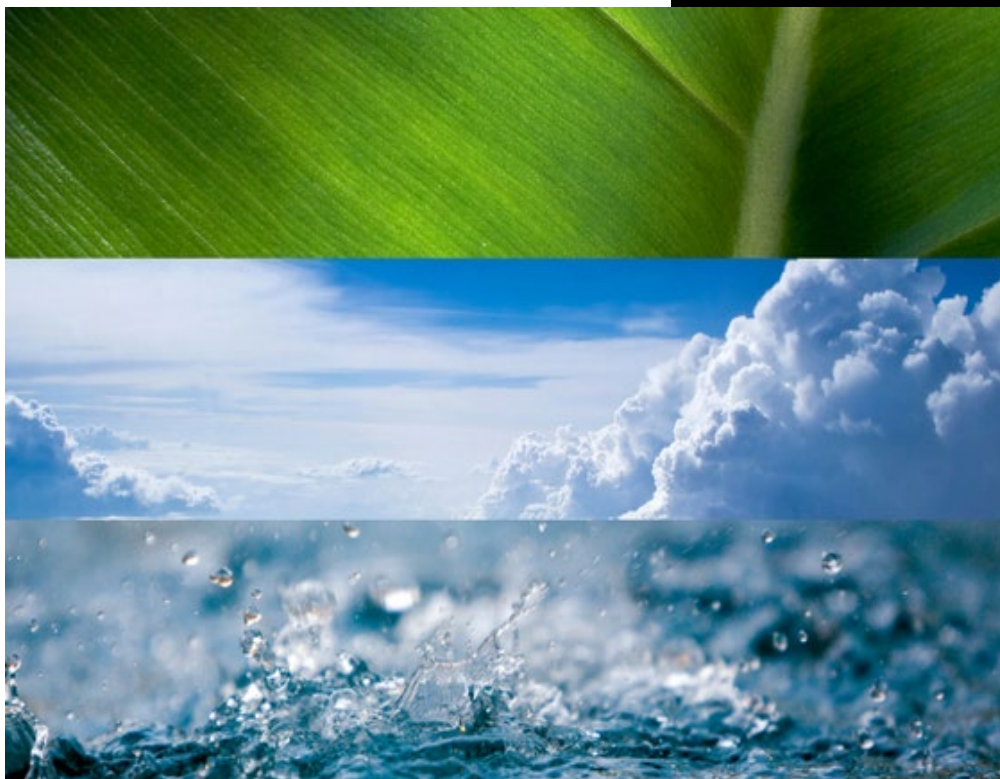
Readiness  
of ICOS

Readiness of ICOS for Necessities of integrated Global Observations



## D6.12

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



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

RINGO (GA no 730944)

Public Document/define level

A Data Management Plan created using DMPonline

**Deliverable:**

**Author(s):** Alex Vermeulen ([alex.vermeulen@icos-ri.eu](mailto:alex.vermeulen@icos-ri.eu)); Benjamin Pfeil ([benjamin.pfeil@gfi.uib.no](mailto:benjamin.pfeil@gfi.uib.no))

**Date:** 24.02.2020

**Activity:** WP6

**Lead Partner:** ICOS ERIC

**Document Issue:** 1

**Dissemination Level:** Public

**Contact:** [alex.vermeulen@icos-ri.eu](mailto:alex.vermeulen@icos-ri.eu)

	Name	Partner	Date
From	Alex Vermeulen	ICOS ERIC	24.02.2020
Approved by	Werner Kutsch	ICOS ERIC	25.02.2020

### 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.
- The executive summary is provided giving a short and to the point description of the deliverable.
- All abbreviations are explained in a separate list.
- All references are listed in a concise list.
- The deliverable clearly identifies all contributions from partners and justifies the resources used.
- A full spell check has been executed and is completed.

### 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. All Information in this document is provided "as is" and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability. For the avoidance of all doubts, the European Commission has no liability in respect of this document, which is merely representing the authors view.

Amendments, comments and suggestions should be sent to the authors.

## 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 five work packages with specific emphasis on the further development of the readiness of ICOS Research Infrastructure (ICOS RI) to foster its sustainability. The principal objectives are scientific, geographical, 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.

## TABLE OF CONTENTS

1. DATA SUMMARY .....	4
2.1 MAKING DATA FINDABLE, INCLUDING PROVISIONS FOR METADATA [FAIR DATA] .....	5
2.2 MAKING DATA OPENLY ACCESSIBLE [FAIR DATA].....	6
2.3 MAKING DATA INTEROPERABLE [FAIR DATA].....	7
2.4 INCREASE DATA RE-USE (THROUGH CLARIFYING LICENSES) [FAIR DATA].....	7
3. ALLOCATION OF RESOURCES .....	8
4. DATA SECURITY .....	9
5. ETHICAL ASPECTS .....	9
6. OTHER .....	9

## 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 environments. This data is targeted to scientific use to increase our knowledge of the greenhouse gas cycles and the 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 types of data:

- CO<sub>2</sub> 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 ICOS datasets for CH<sub>4</sub> and N<sub>2</sub>O. This data is stored in the WMO and Fluxnet community defined Comma and Tab separated 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 ASCII format, the development, definition, and documentation of this processing is part of RINGO.
- Additional 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 sites), 20 MB (atmosphere, 10 sites) or several Kb (marine, five sites) per day. The Level 2 data products are about 1-20 MB per year.

*Outline the data utility: to whom will it be useful:*

- Scientists 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, through the Datacite through DOIs of the collections of the datasets 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?*

At ingestion, all data objects get assigned a Persistent Identifier (PID) based on the handle system. This PID contains the 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. Collections of data, per station and per release of one or more years, will be generated and minted a Datacite DOI and the metadata of these collections, including attribution information, will be stored at Datacite.

### *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 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, accessible through an open SPARQL endpoint (W3C standards). Naming convention of the metadata entries follows ISO19115 where possible. Collections and their metadata will also be findable from the Datacite search engine.

## 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 originating from the RINGO project 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 the RINGO project is in principle, kept closed for users from outside the relevant RINGO task until the end of the project, but will be provided with open access according to the ICOS Data Policy after the project end. Descriptive metadata will always be openly available for all data.

*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 through 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)
- Datacite for collections of data
- 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*

While access to RINGO raw data (Level 0) objects is open, and not subject to any restrictions or conditions, these data are not directly downloadable via the corresponding landing pages. Instead, interested users are requested to contact the relevant ICOS thematic center, which will provide the data as well as assistance on how to best use and interpret it.

## 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.*

RINGO and ICOS data and metadata follow standards that are well-anchored in the Earth Science community, including relevant ISO specifications and the INSPIRE directive. This ensures smooth uptake and usage of the data products by all designated scientific user communities, including support for automated harvesting of data and metadata by machine-driven processing e.g. in the cloud.

The metadata associated with all data objects stored in the ICOS Carbon Portal is stored in a RDF database, based on an open ontology based on OWL, which 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 data objects in B2SAFE through CKAN. The B2FIND repository is again linked to the GEOPortal for global access to the metadata from 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 interdisciplinary 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. Some data, e.g. related to development work being performed as part of the RINGO DoA, will be made available after the end of the project, see section 2.2 above.

*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 license is available to all third parties. RINGO experimental data results are restricted to the consortium until the end of the project. They will become available under CC4BY 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 the 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 CO<sub>2</sub> and CH<sub>4</sub> mole fractions at the ICOS Atmosphere Thematic Centre, Atmos. Meas. Tech., 9, 4719-4736, <https://doi.org/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 infrastructure that is foreseen to exist for at least 20-25 years. This would guarantee operation and data availability until 2040. Following OAIS recommendations, periodic consultations with the designated scientific end-user communities are planned to ensure that the data remain fully FAIR throughout this time.

### **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/CDI 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. In general Carbon Portal is responsible for the data identification and minting of PIDs and DOIs, publishing of data and metadata and archiving to the trusted repository. The Thematic Centers of ICOS are responsible for the data curation and provenance. The project participants and task leaders are responsible for the provenance of the raw data and in some cases of curation of the experimental results.

#### *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 covered 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 backed up at all individual instances of ICOS, these locations are the 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 minted by ICOS CP that contains the AES256 checksum of the data for unique identification and consistency check. NOTE: the (primary) repository in this context is the Carbon Portal (which operates the “ICOS repository”), not the EUDAT data center!

## 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. All other information is at

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