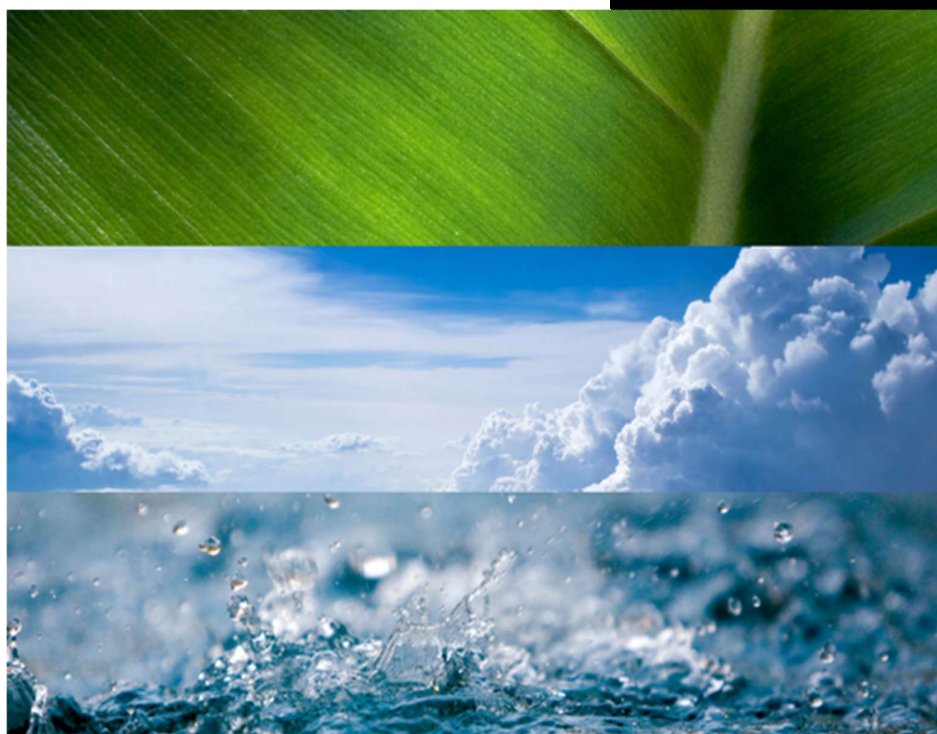




# Initial Data Management Plan



RINGO (GA no 730944)

PUBLIC; R

---

**Deliverable: Initial Data Management Plan (DMP)**

**Author(s): Alex Vermeulen, Benjamin Pfeil**

**Date: 30 June 2017**

**Activity: WP 6, Task 6.4**

**Lead Partner: ICOS ERIC**

**Document Issue: 1.0, will be updated at Month 18**

**Dissemination Level: PU; R**

**Contact: ICOS ERIC CP, alex.vermeulen@icos-ri.eu**

---

	<b>Name</b>	<b>Partner</b>	<b>Date</b>
From	Alex Vermeulen		29 June 2017
Reviewed by	Evi-Carita Riikonen		30 June 2017
Approved by	Jouni Heiskanen		30 June 2017

<b>Version</b>	<b>Date</b>	<b>Comments/Changes</b>	<b>Author/Partner</b>
1.0	30 June 2017		ICOS ERIC

**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

---

## Contents

1. INTRODUCTION .....	2
2. Data summary .....	3
3. FAIR data .....	3
3.1 Data findability, including provisions for metadata.....	3
3.2 Data accessibility.....	3
3.3 Data interoperability.....	4
3.4 Data re-use.....	4
4. Allocation of resources .....	4
5. Data security .....	4
6. Ethical aspects.....	4
7. Other .....	4

## 1. INTRODUCTION

This initial Data Management Plan describes the existing and planned data management, data access and data security policies of ICOS RI.

The structure of this report consists of a general overview on the data management of the RINGO project as a whole, as well as a more detailed description of data management.

The mission of the European Research Infrastructure 'Integrated Carbon Observatory System' (ICOS RI) is to enable research to understand the greenhouse gas (GHG) budgets and perturbations. The ICOS RI is a distributed research infrastructure that provides the long-term observations required to understand the present state and predict future behaviour of the global carbon cycle and GHG emissions. ICOS RI ensures the continuous, high-precision and long-term greenhouse gas measurements in Europe and adjacent key regions of Africa and Eurasia. The backbone of ICOS RI are the three measurement stations networks: the ICOS atmospheric, ecosystem and ocean networks. Together they are organized within national measurement networks. Technological developments and implementations, related to GHGs, will be promoted by the linking of research, education and innovation.

ICOS Central Facilities (ICOS CF), which are Atmospheric Thematic Centre (ATC), Ecosystem Thematic Centre (ETC), Ocean Thematic Centre (OTC) and the Central Analytical Laboratories (CAL) have the specific tasks of collecting and processing the data and samples (e.g. flask or radiocarbon samples in Atmosphere or soil and plant tissue samples in Ecosystem observational network) received from the national measurement networks.

ICOS ERIC is the legal entity of ICOS RI established to coordinate the operations and the data of ICOS distributed research infrastructure, and to develop, monitor and integrate the activities and the data of ICOS RI. The ICOS Carbon Portal is the ICOS data centre from where ICOS data and ancillary data sets will be published and be accessible for the users. Carbon Portal is responsible for handling and providing ICOS data products. Carbon Portal is being designed and envisioned as the single access point for environmental scientists to discover, obtain, visualise and track observation measurements produced from the observation stations as quickly as possible.

The design of the overall ICOS RI data management is challenged by the complicated requirements of dataflow from the distributed acquisition via centralized processing and quality assessment to the publication of the data products. The ICOS national observation stations are highly distributed; data are semantically diverse, organisational features of the National Networks are different from country to country, observational measurements and resulting data life cycles are varying between observational networks. Data definitions, transfers and responsibilities have been discussed within ICOS RI for several years. These discussions have been documented in numerous documents.

## 2. Data summary

**This section specifies the purpose of the data, the formats and origin, the size and to whom it will be useful.**

There are different kinds of data generated in the project. First of all, there are experiments where new observational technologies are developed and tested, either in the lab or in the field. The data is used to evaluate the performance of the instrumentation and/or methods. Data and metadata are essential for documentation and publishing of the results in the literature. Data formats will be very similar to the current standard data formats used in ICOS, except for new instrument specific raw data formats that are often proprietary. Order of magnitude of the data generated is several Gigabytes.

In several work packages methods are developed for new or existing measurement strategies. For this existing (ICOS or pre-ICOS) measurements and/or model simulations are used. The model data will be generated using existing models. The data will be used to evaluate the different alternatives for measurement strategies. Data and metadata are essential for documentation and publishing of the results in the literature. Data formats will be very similar to the current standard data formats used in ICOS and mainly consist of netcdf files. Order of magnitude of the data is tens of Gigabytes.

In WP5 the historical data sets (pre-ICOS) will be re-evaluated for a limited number of stations by going back to the original raw data and a re-analysis using methodologies as close as possible to the current data processing standards and strategies of ICOS RI, including evaluation of the uncertainties in the individual measurements. The data will be used and offered to the users as improved, quality controlled and recalibrated data sets extending the ICOS dataset to the pre-ICOS period. This dataset will be essential for inverse modelling experiments to complement the ICOS dataset with at least 10 years of historical data. Order of magnitude of the datasize is one gigabyte. This data will be published through the ICOS Carbon Portal following the ICOS data license (CC4BY) and data policy.

## 3. FAIR data

### 3.1 Data findability, including provisions for metadata

**This section outlines the discoverability and identifiability of the data and the use of persistent and unique identifiers.**

All data will be curated using standard EUDAT B2 services, making sure that all data is discoverable through B2FIND. All final (Level 2) datasets will be shared through the ICOS Carbon Portal, which will be fully implementing the FAIR principles. All EUDAT and ICOS CP services make use of ePIC handle for identification of all data objects. (Collections of) Level 2 products will also be minted doi identifiers based on Datacite. All ICOS data object metadata is shared with the GEOSS portal.

Naming conventions are not relevant due to the use of persistent identifiers and the linked machine-readable description through metadata complying with INSPIRE and ISO19115 as a subset. New versions of data objects receive of course their own unique persistent identifier and in the metadata the appropriate link to the older version is added. Same for the reference of an updated version in the newer version.

Keywords are part of the metadata, following the appropriate (community) standards.

### 3.2 Data accessibility

**This section specifies the extent of open access, how the data is made available, what methods and tools are used.**

Experiment and model data might but will be openly accessible only after the end of the projects as soon as the results have been published. All publications will be open access.

Whenever possible data will be openly accessible following the ICOS CC4BY license. Through the EUDAT B2 services of B2FIND, B2DROP and B2SHARE and the ICOS Carbon portal all metadata and data can be found and accessed.

Where relevant and possible with regards to property rights developed software will be made available through the open source repository Github or similar using a GPL license.

### **3.3 Data interoperability**

**This section covers what data and metadata vocabularies, standards and methodologies are used to facilitate interoperability.**

All RINGO and ICOS data and metadata are designed for interoperability and in all cases follow and in some cases even form de-factor (community) standards. All metadata will available in INSPIRE compliant form. All ICOS Carbon Portal data is available as linked open data and through an open SPARQL endpoint. The RINGO project specific data will be available through the EUDAT B2 services following the same standards.

### **3.4 Data re-use**

**This section specifies data licencing, availability and length of time for re-use.**

Wherever possible the data will be shared right after production following the Creative Commons 4.0 International License with Attribution (CC4BY). Experimental data test data will in some cases only become available after the end of the project or publication of the results, whatever comes first, and will be shared used the same CC4BY license.

The CC4BY licenses guarantees maximum re-use (and redistribution) while maintaining the traceability of the use and credit to the data providers and their sponsors.

Data quality assurance and control is central and the raison d'être of ICOS and the RINGO project. About 80% of the efforts spent in the ICOS Thematic Centres is directed at data quality assurance.

ICOS RI has a time horizon of at least 20 years, the data will remain useful and usable beyond that period. For example, now the time-series generated since 1957 of CO<sub>2</sub> concentrations at Mauna Loa are still being used.

## **4. Allocation of resources**

The costs of making data fair can be estimated to be 100% of the effort by ICOS Carbon Portal and 25% of the operational costs of the ICOS Thematic Centres. About 10% of the RINGO budget is used for improvement of the interoperability of the ICOS metadata.

The cost of long term preservation of the data is at this moment impossible to estimate. On the long-term the costs of storage are foreseen to go down tremendously. At this moment, the storage costs for ICOS are foreseen to be in the order of 50 k€ per year.

## **5. Data security**

ICOS and RINGO produce non-sensitive data. Personal information is processes and stored according to the ICOS privacy policy. For secure storage ICOS relies on the European e-Infrastructure of EUDAT and EGI.

## **6. Ethical aspects**

Not relevant for the RINGO data.

## **7. Other**

There are several documents that are related to the data management plan developed in RINGO. These are ICOS Data Policy document, ICOS Data lifecycle document, ICOS Data License, and ICOS measurement protocol documentation for Atmosphere, Ocean and Ecosystem community.