



KADI DATA MANAGEMENT PLAN 1

Deliverable 6.3. HORIZON-INFRA-2021-DEV-01-02



Funded by
the European Union

kadi-project.eu

PROJECT NUMBER	OFFICIAL NAME	EXTERNAL COMMUNICATIONS NAME
HORIZON-INFRA-2021-DEV-01-02	KADI Knowledge and climate services from an African observation and Data research Infrastructure	KADI
DUE DATE OF DELIVERABLE	REFERENCE WP AND DELIVERABLE	BENEFICIARY IN CHARGE
28.02.2023	WP6 & Deliverable 6.3	ICOS ERIC
PROJECT URL		EU PROJECT OFFICER
https://www.kadi-project.eu		Pierre QUERTENMONT

LIST OF CONTRIBUTORS

NAME	INSTITUTION
Alex Vermeulen	ICOS Carbon Portal
Elena Saltikoff	ICOS ERIC
Werner Leo Kutsch	ICOS ERIC
Theresia Bilola	ICOS ERIC
Joerg Klausen	MeteoSwiss
Jonathan Padavatan	GCI WITS
Jessica Steinkopf	GCI WITS
Ingunn Skjelvan	NORCE
Yolandi Ernst	GCI WITS

Executive Summary

This first data management plan 1.0 (DMP) is a formal document that outlines how data will be handled both during and after the KADI project is completed based on the information available at the time of the publication. The goal of this data management plan is to consider the many aspects of data management, metadata generation, data preservation, and analyses. It ensures that data is well-managed in the present, and prepared for preservation in the future. Dissemination opportunities, targeted audiences and key messages will be identified and updated regularly. For the preparation of the present version of this deliverable (at the beginning of the KADI project) a number of reference data management documentation have been consulted by the authors and reviewed for the purpose to a self-consistent high-quality plan, based on best practices and experiences that authors consider related. This deliverable is related to the Task 6.1 dedicated to coordination, especially data management, including the necessary legal aspects such as GDPR when dealing with personal data. This Data Management Plan (DMP) outlines how the research data collected or generated during the KADI project will be handled by the consortium during and after the project. It describes which standards and methodology for data collection and generation will be followed and how data will be shared and preserved. The DMP is a public and evolving document which will gain more precision and clarity during the course of the project. It will be revised, if needed, with up-to-date information.

CONTENTS

EXECUTIVE SUMMARY	2
A DATA MANAGEMENT PLAN CREATED USING DMPONLINE	ERROR! BOOKMARK NOT DEFINED.
ABSTRACT	6
DATA SUMMARY	6
2. WHAT TYPES AND FORMATS OF DATA WILL THE PROJECT GENERATE OR RE-USE?	6
3. PURPOSE OF THE DATA GENERATION OR RE-USE AND ITS RELATION TO THE OBJECTIVES OF THE PROJECT?	6
4. WHAT IS THE EXPECTED SIZE OF THE DATA THAT YOU INTEND TO GENERATE OR RE-USE?.....	7
5. WHAT IS THE EXPECTED SIZE OF THE DATA THAT YOU INTEND TO GENERATE OR RE-USE?	ERROR! BOOKMARK NOT DEFINED.
6. TO WHOM MIGHT YOUR DATA BE USEFUL ('DATA UTILITY'), OUTSIDE YOUR PROJECT?.....	7
FAIR DATA	7
7. MAKING DATA FINDABLE, INCLUDING PROVISIONS FOR METADATA: WILL DATA BE IDENTIFIED BY A PERSISTENT IDENTIFIER?	7
8. MAKING DATA FINDABLE, INCLUDING PROVISIONS FOR METADATA: WILL RICH METADATA BE PROVIDED TO ALLOW DISCOVERY? WHAT METADATA WILL BE CREATED? WHAT DISCIPLINARY OR GENERAL STANDARDS WILL BE FOLLOWED? IN CASE METADATA STANDARDS DO NOT EXIST IN YOUR DISCIPLINE, PLEASE OUTLINE WHAT TYPE OF METADATA WILL BE CREATED AND HOW.	7
9. MAKING DATA FINDABLE, INCLUDING PROVISIONS FOR METADATA: WILL SEARCH KEYWORDS BE PROVIDED IN THE METADATA TO OPTIMIZE THE POSSIBILITY FOR DISCOVERY AND THEN POTENTIAL RE-USE?	8
10. MAKING DATA FINDABLE, INCLUDING PROVISIONS FOR METADATA: WILL METADATA BE OFFERED IN SUCH A WAY THAT IT CAN BE HARVESTED AND INDEXED?.....	8
12. MAKING DATA ACCESSIBLE - REPOSITORY: HAVE YOU EXPLORED APPROPRIATE ARRANGEMENTS WITH THE IDENTIFIED REPOSITORY WHERE YOUR DATA WILL BE DEPOSITED?.....	8
13. MAKING DATA ACCESSIBLE - REPOSITORY: DOES THE REPOSITORY ENSURE THAT THE DATA IS ASSIGNED AN IDENTIFIER? WILL THE REPOSITORY RESOLVE THE IDENTIFIER TO A DIGITAL OBJECT?	8
14. WILL ALL DATA BE MADE OPENLY AVAILABLE? IF CERTAIN DATASETS CANNOT BE SHARED (OR NEED TO BE SHARED UNDER RESTRICTED ACCESS CONDITIONS), EXPLAIN WHY, CLEARLY SEPARATING LEGAL AND CONTRACTUAL REASONS FROM INTENTIONAL RESTRICTIONS. NOTE THAT IN MULTI-BENEFICIARY PROJECTS IT IS ALSO POSSIBLE FOR SPECIFIC BENEFICIARIES TO KEEP THEIR DATA CLOSED IF OPENING THEIR DATA GOES AGAINST THEIR LEGITIMATE INTERESTS OR OTHER CONSTRAINTS AS PER THE GRANT AGREEMENT.	9
15. IF AN EMBARGO IS APPLIED TO GIVE TIME TO PUBLISH OR SEEK PROTECTION OF THE INTELLECTUAL PROPERTY (E.G., PATENTS), SPECIFY WHY AND HOW LONG THIS WILL APPLY, BEARING IN MIND THAT RESEARCH DATA SHOULD BE MADE AVAILABLE AS SOON AS POSSIBLE.....	9
MAKING DATA ACCESSIBLE - DATA:	9
16. WILL THE DATA BE ACCESSIBLE THROUGH A FREE AND STANDARDIZED ACCESS PROTOCOL?	9
17. IF THERE ARE RESTRICTIONS ON USE, HOW WILL ACCESS BE PROVIDED TO THE DATA, BOTH DURING AND AFTER THE END OF THE PROJECT?.....	9
18. HOW WILL THE IDENTITY OF THE PERSON ACCESSING THE DATA BE ASCERTAINED?	9
19. IS THERE A NEED FOR A DATA ACCESS COMMITTEE (E.G., TO EVALUATE/APPROVE ACCESS REQUESTS TO PERSONAL/SENSITIVE DATA)?.....	9

MAKING DATA ACCESSIBLE - METADATA:	10
20. WILL METADATA BE MADE OPENLY AVAILABLE AND LICENSED UNDER A PUBLIC DOMAIN DEDICATION CC0, AS PER THE GRANT AGREEMENT? IF NOT, PLEASE CLARIFY WHY. WILL METADATA CONTAIN INFORMATION TO ENABLE THE USER TO ACCESS THE DATA?	10
21. HOW LONG WILL THE DATA REMAIN AVAILABLE AND FINDABLE? WILL METADATA BE GUARANTEED TO REMAIN AVAILABLE AFTER DATA IS NO LONGER AVAILABLE?	10
22. WILL DOCUMENTATION OR REFERENCE ABOUT ANY SOFTWARE BE NEEDED TO ACCESS OR READ THE DATA BE INCLUDED? WILL IT BE POSSIBLE TO INCLUDE THE RELEVANT SOFTWARE (E.G., IN OPEN-SOURCE CODE)?	10
MAKING DATA INTEROPERABLE:	10
23. WHAT DATA AND METADATA VOCABULARIES, STANDARDS, FORMATS OR METHODOLOGIES WILL YOU FOLLOW TO MAKE YOUR DATA INTEROPERABLE TO ALLOW DATA EXCHANGE AND RE-USE WITHIN AND ACROSS DISCIPLINES? WILL YOU FOLLOW COMMUNITY-ENDORSED INTEROPERABILITY BEST PRACTICES? WHICH ONES?	10
24. IN CASE IT IS UNAVOIDABLE THAT YOU USE UNCOMMON OR GENERATE PROJECT SPECIFIC ONTOLOGIES OR VOCABULARIES, WILL YOU PROVIDE MAPPINGS TO MORE COMMONLY USED ONTOLOGIES? WILL YOU OPENLY PUBLISH THE GENERATED ONTOLOGIES OR VOCABULARIES TO ALLOW REUSING, REFINING OR EXTENDING THEM?	11
25. WILL YOUR DATA INCLUDE QUALIFIED REFERENCES [1] TO OTHER DATA (E.G., OTHER DATA FROM YOUR PROJECT, OR DATASETS FROM PREVIOUS RESEARCH)?	11
INCREASE DATA RE-USE:	11
26. HOW WILL YOU PROVIDE DOCUMENTATION NEEDED TO VALIDATE DATA ANALYSIS AND FACILITATE DATA RE-USE (E.G., README FILES WITH INFORMATION ON METHODOLOGY, CODEBOOKS, DATA CLEANING, ANALYSES, VARIABLE DEFINITIONS, UNITS OF MEASUREMENT, ETC.)?	11
27. WILL YOUR DATA BE MADE FREELY AVAILABLE IN THE PUBLIC DOMAIN TO PERMIT THE WIDEST RE-USE POSSIBLE? WILL YOUR DATA BE LICENSED USING STANDARD REUSE LICENSES, IN LINE WITH THE OBLIGATIONS SET OUT IN THE GRANT AGREEMENT?	11
28. WILL THE DATA PRODUCED IN THE PROJECT BE USEABLE BY THIRD PARTIES, IN PARTICULAR AFTER THE END OF THE PROJECT?	12
29. WILL THE PROVENANCE OF THE DATA BE THOROUGHLY DOCUMENTED USING THE APPROPRIATE STANDARDS?	12
30. DESCRIBE ALL RELEVANT DATA QUALITY ASSURANCE PROCESSES.	12
31. FURTHER TO THE FAIR PRINCIPLES, DMPs SHOULD ALSO ADDRESS RESEARCH OUTPUTS OTHER THAN DATA, AND SHOULD CAREFULLY CONSIDER ASPECTS RELATED TO THE ALLOCATION OF RESOURCES, DATA SECURITY AND ETHICAL ASPECTS	12
OTHER RESEARCH OUTPUTS	12
32. BENEFICIARIES SHOULD CONSIDER WHICH OF THE QUESTIONS PERTAINING TO FAIR DATA ABOVE, CAN APPLY TO THE MANAGEMENT OF OTHER RESEARCH OUTPUTS, AND SHOULD STRIVE TO PROVIDE SUFFICIENT DETAIL ON HOW THEIR RESEARCH OUTPUTS WILL BE MANAGED AND SHARED, OR MADE AVAILABLE FOR RE-USE, IN LINE WITH THE FAIR PRINCIPLES.	12
33. WHAT WILL THE COSTS BE FOR MAKING DATA OR OTHER RESEARCH OUTPUTS FAIR IN YOUR PROJECT (E.G., DIRECT AND INDIRECT COSTS RELATED TO STORAGE, ARCHIVING, RE-USE, SECURITY, ETC.)? HOW WILL THESE BE COVERED? NOTE THAT COSTS RELATED TO RESEARCH DATA/OUTPUT MANAGEMENT ARE ELIGIBLE AS PART OF THE HORIZON EUROPE GRANT (IF COMPLIANT WITH THE GRANT AGREEMENT CONDITIONS).	12
34. WHO WILL BE RESPONSIBLE FOR DATA MANAGEMENT IN YOUR PROJECT?	13

35. HOW WILL LONG TERM PRESERVATION BE ENSURED? DISCUSS THE NECESSARY RESOURCES TO ACCOMPLISH THIS (COSTS AND POTENTIAL VALUE, WHO DECIDES AND HOW, WHAT DATA WILL BE KEPT AND FOR HOW LONG)?	13
DATA SECURITY	13
36. WHAT PROVISIONS ARE OR WILL BE IN PLACE FOR DATA SECURITY (INCLUDING DATA RECOVERY AS WELL AS SECURE STORAGE/ARCHIVING AND TRANSFER OF SENSITIVE DATA)?.....	13
ETHICS.....	13
37. ARE THERE, OR COULD THERE BE, ANY ETHICS OR LEGAL ISSUES THAT CAN HAVE AN IMPACT ON DATA SHARING? THESE CAN ALSO BE DISCUSSED IN THE CONTEXT OF THE ETHICS REVIEW. IF RELEVANT, INCLUDE REFERENCES TO ETHICS DELIVERABLES AND ETHICS CHAPTER IN THE DESCRIPTION OF THE ACTION (DOA).	13
38. WILL INFORMED CONSENT FOR DATA SHARING AND LONG-TERM PRESERVATION BE INCLUDED IN QUESTIONNAIRES DEALING WITH PERSONAL DATA?.....	14
OTHER ISSUES.....	14
39. DO YOU, OR WILL YOU, MAKE USE OF OTHER NATIONAL/FUNDER/SECTORIAL/DEPARTMENTAL PROCEDURES FOR DATA MANAGEMENT? IF YES, WHICH ONES (PLEASE LIST AND BRIEFLY DESCRIBE THEM)?	14

Abstract

Africa bears the brunt of the effects of climate change. It is also the continent that requires the most work to achieve the ambitious goals set in the Paris Agreement and the UN Sustainable Development Goals. The EU-funded KADI project will provide concepts for developing the best science and science-based services in Africa to accelerate common actions for climate change mitigation. The consortium consists of partners from Africa and Europe and combines different experiences. It will provide an inclusive design for a pan-African climate observation system based on a comprehensive documentation of past and existing observing capability and a broad information network as the basis for successful and sustainable cooperation that connects all stakeholders. This Data Management Plan was created using DMPOnline.

Data Summary

1. Will you re-use any existing data and what will you re-use it for?

Yes. For the deliverable 2.5 “Integrated data set from Kenyan observations” (Task 2.4) we will use:

- Metadata sets available in GAWSIS OSCAR/Surface, metadata set of existing ecosystems GHG flux observations.
- QC'd data sets available at World Data Centres for the purpose of further quality control, trend analysis.
- Meteorological data as made available by the Kenyan Meteorological department and other stakeholders in Kenya for the purpose of validating model output and other scientific applications.

2. What types and formats of data will the project generate or re-use?

- The riverine and ocean data collected in Task 2.2: Coastal Southern African Biogeochemistry will be provided as text and numerical information.
- The climate risk data in Task 2.3: will be text and numerical information.
- The data to be re-used from World Data centre data files are in UTF-8 text and NetCDF formats.
- The climate model will re-use ERA5 reanalysis data as input data. ERA5 is the latest reanalysis product from the European Centre for Medium-Range Weather Forecasts (ECMWF). This data is available in NetCDF format.

3. Purpose of the data generation or re-use and its relation to the objectives of the project?

Identification of criteria and requirements critical for a successful and sustainable operation of future atmospheric and ecosystem observations in East Africa, characterization of footprints, trend and gap analysis and recommendations for future expansion of observing capability.

4. What is the expected size of the data that you intend to generate or re-use?

The data from the pilots generated in WP2 will be from different origins: from model output, from coastal areas, from cities, and from atmospheric/ecosystem measuring sites. The coastal sites will over the length of the project likely produce several Gigabytes of data, however, the final decision regarding this will be done when the equipment has been installed at the different measuring locations.

The downscaled climate model experiments will involve four experimental runs for each region, each generating 25TB of data, totalling 100TB of data for the southern Africa region and another 100TB for the East Africa region. However, we will only store the surface variables relevant to the project output which will be approximately 500GB for each region, totally 1TB of data to be stored in the long run.

The origin of some of the data will be confirmed after the data collection.

5. To whom might your data be useful ('data utility'), outside your project?

The coastal data collected in Task 2.2 will be of interest to the broader community of ocean researchers, e.g., in context of Southern Ocean Carbon – Climate Observatory (SOCCO).

The Integrated data set from Kenyan observations will be useful for benchmarking for any authority planning their national GHG and climate observation network, especially in Africa. It can also act as a consolidated input dataset for students and climate change researchers, as well as bodies planning climate services for the future applications.

Fair Data

6. Making data findable, including provisions for metadata: Will data be identified by a persistent identifier?

Yes, all data to be retained will be provisioned with a persistent identifier, in the case of using the ICOS repository this will be, at a minimum, a Handle PID (ePIC). In most cases also a Data cite DOI PID will be minted. In some cases, data will be published using other public archives with similar functionality that mints PIDs or DOIs, like Pangaea or Zenodo.

7. Making data findable, including provisions for metadata: Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

Yes, all PIDs will resolve to a landing page that contains human and machine-readable rich data that allows the users to find and access the data using detailed (faceted) search criteria, selecting on keywords, variable names and other elements of the rich metadata provided at ingestion. Station metadata archived in

OSCAR/Surface conform to the WMO WIGOS Metadata Standard. Stations are identified by a globally unique WIGOS Station Identifier. The data generated by the climate model runs will be in NetCDF format, which contains all relevant metadata.

8. Making data findable, including provisions for metadata: Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use?

Yes, keywords will be provided, users are encouraged to select keywords from the GCMD (Global Change Master Directory) standard. In the Carbon portal data portal data files can be selected based on a faceted search that include the keywords.

9. Making data findable, including provisions for metadata: Will metadata be offered in such a way that it can be harvested and indexed?

Yes, at ICOS Carbon Portal all metadata is accessible to both human and machine to machine communication, using the following standards: iso19115, schema.org, and by content negotiation as xml, json-ld, rdf/turtle, rdf/xml. All metadata is served from the landing pages and the open sparql endpoint.

10. Making data accessible - Repository: Will the data be deposited in a trusted repository?

The ICOS Carbon Portal is the trusted data repository for the large ICOS community and beyond. All data is securely copied using the EUDAT CDI B2SAFE service, that stores two copies of each digital artefact at two locations (one at CSC, Finland and one at FZJ, Germany). Next to that ICOS Carbon portal stores its own copy of the data, with a separate backup at two different buildings at Lund University, Sweden. Some metrological climate data are also openly available from the National Centre for Environmental Information (NCEI), Ashville (NC).

11. Making data accessible - Repository: Have you explored appropriate arrangements with the identified repository where your data will be deposited?

ICOS ERIC is in full control of the ICOS Carbon Portal as owner of the repository. ICOS is a long-term research infrastructure, legally presented by ICOS ERIC, that is foreseen to exist permanently, for at least 25 years.

12. Making data accessible - Repository: Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object?

Yes, each data object is assigned a persistent identifier that contains (part of) its digital SHA256 checksum. Through the landing page that the PID resolves into, a direct link to the digital object is provided together with licence and access metadata.

14. Will all data be made openly available? If certain datasets cannot be shared (or need to be shared under restricted access conditions), explain why, clearly separating legal and contractual reasons from intentional restrictions. Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement.

All raw and higher-level data products are provided and will follow the ICOS data licence, which is CC4BY.

15. If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g., patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

Experimental data from the project is in principle kept closed until the end of the project and will be provided open access according the ICOS Data Policy after the project end. Data can be uploaded by ICOS Carbon Portal and all metadata will be publicly available, but data access will be restricted under a moratorium. The restricted data will be open to project participants.

Making data accessible - Data:

16. Will the data be accessible through a free and standardized access protocol?

All data (and metadata) is available using standard http protocol, using semantic web and linked open data technology.

17. If there are restrictions on use, how will access be provided to the data, both during and after the end of the project?

There could be embargoes for scientific projects (e.g., doctoral theses or publication). The data will be open after the embargo ends. Personal data is protected following the GDPR regulations.

18. How will the identity of the person accessing the data be ascertained?

If user identification is required for data with restricted access, users will have to identify through the Carbon Portal login, using either Edugain, ORCID or local account login. All other data is available without user identification. We log data access purely based on IP and date/time.

19. Is there a need for a data access committee (e.g., to evaluate/approve access requests to personal/sensitive data)?

Not required for the environmental data collected within the framework of this project.

Making data accessible - Metadata:

20. Will metadata be made openly available and licensed under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why. Will metadata contain information to enable the user to access the data?

Yes, all ICOS metadata is licensed through CC0.

21. How long will the data remain available and findable? Will metadata be guaranteed to remain available after data is no longer available?

ICOS is a long-term research infrastructure, legally presented by ICOS ERIC, that is foreseen to exist permanently, for at least 25 years. ICOS Carbon Portal will preserve data and metadata for the lifetime of ICOS and beyond. After the termination of ICOS all data and metadata will be transferred to a static website and/or other trusted repositories during the decommissioning period of at least one year.

OSCAR/Surface is the operational metadata portal for surface-based observation assets of the World Meteorological Organization (WMO). The platform is operated by MeteoSwiss with a long-term perspective. WMO is the data owner and will make provisions for long-term archival of OSCAR/Surface in the highly unlikely event of termination of the activity.

22. Will documentation or reference about any software be needed to access or read the data be included? Will it be possible to include the relevant software (e.g., in open-source code)?

Data providers can add additional documents or links to software (e.g., in a GitHub repository) to any data object. Software can be added to the data collection/set.

Making data interoperable:

23. What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?

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 an 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 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. Keywords follow

in general the GCMD standard and data formats follow the community standards such as cf-netcdf, FLUXNET, WMO GAW etc.

OSCAR/Surface station metadata can be exported as WIGOS Metadata Standard compliant XML files. These are also harvest by the EU-funded NextGEOSS portal.

24. In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies? Will you openly publish the generated ontologies or vocabularies to allow reusing, refining or extending them?

The ICOS ontology is modelled in OWL and can be queried through the open sparql endpoint. The ontology is published at <https://lov.linkeddata.es/dataset/lov/vocabs/cpmeta>. All metadata is mapped to schema.org and iso19115.

OSCAR/Surface metadata are compliant with OGC O&M / ISO19156.

25. Will your data include qualified references [1] to other data (e.g., other data from your project, or datasets from previous research)?

[1]A qualified reference is a cross-reference that explains its intent. For example, X is regulator of Y is a much more qualified reference than X is associated with Y, or X see also Y. The goal therefore is to create as many meaningful links as possible between (meta)data resources to enrich the contextual knowledge about the data. (Source : <https://www.go-fair.org/fair-principles/i3-metadata-include-qualified-references-metadata/>)

At the moment we only allow for association of data with other data without further qualification. The idea is that the cross-referencing can be performed better using the rich metadata attached to each data object.

Increase data re-use:

26. How will you provide documentation needed to validate data analysis and facilitate data re-use (e.g., readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)?

Data providers are encouraged to provide links to supporting documentation with the data together with relevant descriptions (abstracts) and title of the dataset that also include references to methodology. ICOS data itself is foreseen with detailed provenance data on instrumentation and methods used and includes detailed information on location, variables, units, data providers and their roles, all modelled over time.

27. Will your data be made freely available in the public domain to permit the widest re-use possible? Will your data be licensed using standard reuse licenses, in line with the obligations set out in the Grant Agreement?

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

28. Will the data produced in the project be useable by third parties, in particular after the end of the project?

Yes, all data and metadata will be stored permanently at the ICOS Carbon Portal. Station metadata will be permanently stored in OSCAR/Surface.

29. Will the provenance of the data be thoroughly documented using the appropriate standards?

Carbon portal provides provenance of all data objects following the PROV-O ontology.

30. Describe all relevant 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 Centres and their contributors.

The ICOS data life cycle is documented in the Improved ICOS data Life Cycle document (<https://doi.org/10.18160/D2JV-KB6B>).

31. Further to the FAIR principles, DMPs should also address research outputs other than data, and should carefully consider aspects related to the allocation of resources, data security and ethical aspects

OK.

Other Research Outputs

32. Beneficiaries should consider which of the questions pertaining to FAIR data above, can apply to the management of other research outputs, and should strive to provide sufficient detail on how their research outputs will be managed and shared, or made available for re-use, in line with the FAIR principles.

Relevant digital research outputs, such as videos, documents and presentations, will be published under the same conditions and methods as described for data in the previous sections. Other relevant material for internal project use can make use of the ICOS Next Cloud sharing facility that provides online collaboration tools and a safe and backed up data sharing space.

33. What will the costs be for making data or other research outputs FAIR in your project (e.g., direct and indirect costs related to storage, archiving, re-use, security, etc.)? How will these be covered? Note that costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions).

Costs are relatively minor and everything will be covered free of charge by ICOS ERIC in kind contributions.

The Wits data server will be used to store project data, as an in-kind contribution to KADI.

34. Who will be responsible for data management in your project?

Carbon Portal and the ICOS Head Office (Alex Vermeulen and Ville Kasurinen).

MeteoSwiss (Jörg Klausen) for OSCAR/Surface.

GCI-Wits (Jonathan Padavatan) for climate model data.

35. How will long term preservation be ensured? Discuss the necessary resources to accomplish this (costs and potential value, who decides and how, what data will be kept and for how long)?

Data will be archived and stay available through the ICOS Carbon Portal. The costs are estimated to be negligible because Carbon Portal has a high archiving potential through cooperation with the EUDAT Collaborative Data Infrastructure (or EUDAT CDI).

Data Security

36. What provisions are or will be in place for data security (including data recovery as well as secure storage/archiving and transfer of sensitive data)?

Default Carbon Portal data security: as explained duplicate copies of all data are kept in physically separate locations in different countries. Integrity of the data is ensured by the checksums connected to the data PIDs.

Ethics

37. Are there, or could there be, any ethics or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).

The ethical issues related to qualitative data collection, which includes the handling of personal data (GDPR, anonymity, confidentiality, informed consent), are discussed in the ethics deliverables. They are related to a collection of climate risk data in the Task 2.3.

- The grant proposal and agreement include an ethical questionnaire.
- ICOS-ERIC has ethical guidelines and a gender equality plan which can be referred to.
- ICOS-ERIC has an ethical advisory board that can provide guidance if needed.

In addition to these, each partner would ensure that the ethical guidelines are followed as pertaining to their country and/or institution context.

Due to the geographic dimensions of this project (including different continents and time zones), communication and qualitative data collection will partly rely on online platforms (as opposed to e.g., face to

face project meetings). This presents an ethical issue in the form of data storage and access (e.g., in situations where the contents are recorded). As in all activities in the project, GDPR and inclusivity are also applied in these situations.

38. Will informed consent for data sharing and long-term preservation be included in questionnaires dealing with personal data?

Yes. Any interview / survey / workshop participant will be provided, (in a relevant language), with information of the purpose of data collection and the handling of personal data, and the study's aims and content, and asked for an informed consent. Participants will be informed of the voluntary basis of participation. Obtaining informed consent will take into account the specificities related to a culturally perceived concept of personhood (individual consent versus collective consent).

Other Issues

39. Do you, or will you, make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones (please list and briefly describe them)?

Relevant material for internal project use can make use of the ICOS Next cloud sharing facility that provides online collaboration tools and a safe and backed up data sharing space.