



I-CISK
HUMAN CENTRED CLIMATE SERVICES

Deliverable D7.2
Data Management Plan

January 2022





Innovating Climate services through Integrating Scientific and local Knowledge

Deliverable Title: D7.2 – Data Management Plan
Author(s): Louise Petersson (SMHI), Micha Werner (IHE)
Contributing Authors(s): Ilias Pechlivanidis (SMHI), Ilyas Masih (IHE), Marc vd Homberg (RC510), Marije Schaafsma (VUA)
Date: 23 January 2022
Suggested citation:
Availability: PU: This report is public [Please select]
 CO: Confidential, only for members of the consortium (including the Commission Services)

Document Revisions:

Author	Revision	Date
Louise Petersson & Ilias Pechlivanidis	First Draft	21 December 2021
Micha Werner	Second Draft	23 January 2022
Louise Petersson & Ilias Pechlivanidis	Final	28 January 2021



Executive Summary

This document presents the data management plan (DMP) of the I-CISK project. The DMP is a living document during the whole lifetime of the project. The report describes the strategies concerning the acquisition, storage, management, and distribution of the project data. In particular, it details how I-CISK will respect the FAIR (Findable, Accessible, Interoperable and Re-usable) data principles according to the principles of the “Open research data pilot” EU initiative.

I-CISK will use the Zenodo open research data repository, to provide open access to all publications from the project, as well as the underlying datasets (with all datasets based on personal data irreversibly anonymised). Data will be uploaded to Zenodo within the European Commission Funded Research (OpenAIRE) Community, as well as a community specifically created for the I-CISK project. Uploaded datasets will be assigned a Digital Object Identifier (DOI) to ensure it can be properly referenced, and provided with the relevant metadata that includes the grant agreement number.

The report also lays down the principles for sharing data internally in the project between the collaborating partners. To facilitate collaboration and sharing of data within a secure environment, I-CISK has established a dedicated and access-controlled instance to be used by the project within the Research Drive facility; a secure, cloud-based collaborative environment made available through SURF. SURF is the collaborative IT environment of the universities in the Netherlands and public research institutes.

Furthermore, the report describes the legal framework that governs the management of data within the project, as laid down by the EU General Data Protection Directive (GDPR) and the applicable laws of third countries where the I-CISK project works. The report also details the ethical aspects related to the management of data within the consortium.

The purpose of the DMP is twofold. First, it provides guidance to the partners that are collaborating within the project on how to collect, handle, curate and preserve data, and to ensure that data is properly handled, respecting the principles laid down by GDPR and the ethics strategy of the project. Second, the report provides external parties that are interested in the project with information on the data collected and generated in the project, and how to access that data and if relevant re-use it for their own activities. Importantly, the DMP provides details how their data is processed, curated and stored to those collaborating with the project through the multi-actor platforms in the seven Living Labs that are the foundation of the research and innovation undertaken by the project.

Table of Contents

1	Introduction.....	1
2	Data summary, management and repositories.....	2
2.1	Data collection and generation in I-CISK	2
2.2	Data types and sources.....	2
2.3	Data formats and size	3
2.4	Data repositories	3
2.5	Guiding principles for data collection, storage and quality assurance.....	4
2.5.1	Legal Framework	4
2.5.2	Principles for collecting and processing of open access data	5
2.5.3	Principles for collecting and processing of restricted access data	5
2.5.4	Principles for collecting and processing of personal data	6
2.5.5	Data Quality Assurance	7
3	FAIR Data Management.....	8
3.1	Making data findable, including provisions for metadata.....	8
3.2	Making data openly accessible.....	8
3.3	Making data interoperable.....	9
3.4	Increase data re-use (through clarifying licences).....	9
3.4.1	Licensing of I-CISK Datasets	9
3.4.2	Longevity of I-CISK Datasets	9
4	Allocation of resources.....	11
5	Data security.....	12
6	Ethical Aspects.....	13
	Appendix 1 I-CISK Data	15
	Appendix 2 I-CISK Data sharing agreement.....	17
	Appendix 3 Access to I-CISK Research Drive	20

List of Figures

Figure 1. Invitation email to Research Drive	20
Figure 2. Select the “set password” option.	20
Figure 3. Log in with your guest account.....	21
Figure 4. I-CISK Research Drive main folder structure.	22

List of Tables

Table 1 Recommended data formats..... 3

Glossary

Acronym	Definition
API	Application Programming Interface
CMIP	World Climate Research Programme's Coupled Model Intercomparison Project
CORDEX	Coordinated Regional Climate Downscaling Experiment
CS	Climate Services
DOI	Digital Object Identifier
DMP	Data Management Plan
EU	European Union
FAIR	Findable, Accessible, Interoperable and Re-usable
GDPR	General Data Protection Directive
GEOSS	Global Earth Observation System of Systems
I-CISK	Innovating Climate services through Integrating Scientific and local Knowledge
IHE	IHE Delft Institute for Water Education
LL	Climate Services Living Labs
NetCDF-CF	Network Common Data Form – Climate and Forecast
NGO	Non-Governmental Organisation
OGC	Open Geospatial Consortium
SMHI	Swedish Meteorological and Hydrological Institute
VUA	Vrije Universiteit Amsterdam

1 Introduction

Research data can be defined as any information that has been collected, observed, or created for scientific purposes. This type of data normally includes statistics, results of experiments and simulations, measurements, observations resulting from fieldwork or remote sensing techniques, survey results, interview recordings, and images. In the context of this document, we focus on data in digital form, even if non-digital formats such as laboratory notebooks and diaries are usually part of this category. The data management plan (DMP) allows to identify risks for the future sharing and reusing of the data collected and produced during the I-CISK activities, and thus to define ways to solve corresponding issues or to mitigate these risks as early as possible. Considering this, in I-CISK, the data management plan is a document that describes the data collected and generated by the project, the procedures used by the project to curate data, and how the project is fostering the adoption of the FAIR principles by ensuring that rules are defined to improve data findability, accessibility, interoperability, and reusability.

The document is structured as follows:

- Section 2 - Data Summary: description of the purpose and use of data within I-CISK, and data handling and storage procedures.
- Section 3 - FAIR Data Management: description of how I-CISK will implement the FAIR (Findable, Accessible Interoperable and Re-usable) principles
- Section 4 - Allocation of resources: description of associated costs and personnel
- Section 5 - Data security: how the security of data will be ensured
- Section 6 - Ethical aspects: how personal and sensitive information will be duly protected
- Appendix 1 - I-CISK data: description of the types of data that will be collected and produced during the project, including a description of the metadata of the attached data catalogue¹
- Appendix 2 - Template of a data sharing agreement
- Appendix 3 - Access to I-CISK Research Drive

¹ <https://vu.data.surfsara.nl/index.php/s/QWvniuQKMPg315B>

2 Data summary, management and repositories

2.1 Data collection and generation in I-CISK

The central concept of I-CISK is the innovation of climate services through integration of scientific and local knowledge. In the project, the primary motivation for the collection of data is to represent these different knowledges. Within the context of the seven Living Labs that are central to the project's approach, the data collected provide an essential input into the co-production of (pre-operational) climate services. Additionally, data will be collected and analysed to support the research for the integration of scientific and local knowledge and human-climate feedbacks, and testing and validating scientific hypotheses. Data will also be collected to support evaluation of the impacts of the research and innovation actions of the project within the Living Labs, as well as in the climate services sector in Europe and beyond.

2.2 Data types and sources

Within the project, three main types of data and sources that will be collected are identified:

- Global/Regional scientific hydro-climatic data: I-CISK will make extensive use of large-scale hydro-climatic datasets representing the current state of knowledge of the earth system (in particular hydrological, meteorological and environmental data). These data cover large geographic regions (e.g. regional/global) and will be primarily sourced from open access repositories, such as Copernicus Climate Data Store, GEOSS, as well as other global repositories. For selected repositories, the source data may be proprietary, while derived and processed datasets are open access. Where data is not open access, and where required, a data sharing and processing agreement will be made with the data providers (see Appendix 2 for a template).
- Local scientific hydro-climatic data: In the context of the seven Living Labs, extensive use will be made of existing hydro-climatic data, which will be collected through the actors and stakeholders that form the multi-actor platforms established in each of the Living Labs. A concise overview of the specific data sources will be established as part of the operationalising of the Living Labs. This will vary across the seven Living Labs depending on the specific objectives as well as the actors and stakeholders involved. These data may be sourced from public agencies (e.g. National Hydrological and Meteorological Agencies) as well as (semi) private sector companies (e.g. irrigation agencies, hydropower companies), or NGO's. Data may also be collected by citizen observatories. Most of these data will be existing datasets, although limited and focused data campaigns, including citizen observatories, will be developed in the project. Where data is not open access, and where required, a data processing agreement will be made with the data providers (see Appendix 2 for a template).
- Social data representing local knowledge and perceptions: In the research developed by the project extensive use will be made of social data to establish a common understanding of local knowledges and perceptions. These data will be sourced through a range of methods, including primary data collection campaigns undertaken by the project (group discussions, interviews, workshops, role play games, demonstration of climate services), as well as secondary data obtained through local and national agencies (e.g. statistics offices, census data, socio-economic databases). The collection of these data will include human involvement, limited to adults, both women and men, who are able to give fully informed and prior consent for their voluntary participation in the project activities. The project will follow the "data minimization principle", where a very limited amount of personal data will be collected from those adults who participate in the I-CISK project. These data will remain confidential and only managed by designated partners in the project, and must be irreversibly anonymised before being made public. Further details of the procedures established in the managing of social data are provided in the subsequent sections.

2.3 Data formats and size

Data collected and generated in the project is expected to cover a range of data types, in particular for local data collected in the context of the Living Labs, which may be made available in a wide range of propriety formats.

For storing and managing the data within the consortium, as well as datasets published in public repositories (see next section), data will be formatted wherever possible using standard data formats. These recommendations follow where applicable published recommendations for data formats in research projects (<https://www.openaire.eu/data-formats-preservation-guide>; <https://www.ukdataservice.ac.uk/manage-data/format/recommended-formats>).

Table 1 Recommended data formats

Type of data	Recommended Format (s)
Geospatial data (static)	Geo-referenced TIFF (.tif, .tiff) ESRI Shapefile (.shp, .shx, .dbf, .prj, .sbx, .sbn optional).
Geospatial data (dynamic)	NetCDF-CF and CMIP6 compliant
Tabular data	Comma Separated Values (CSV)
Textual data	Markdown format (https://commonmark.org/) eXtensible Markup Language (XML)
Images	TIFF 6.0 uncompressed (.tif). JPEG or PNG if original created in this format
Audio	Free Lossless Audio Codec (FLAC) (.flac) MPEG-1 Audio Layer 3 (.mp3) if original created in this format
Video	MPEG-4 (.mp4)
Documents & Publications	PDF/UA, PDF/A or PDF (.pdf) MS Word & LaTeX

At the time of writing the initial version of the DMP, the size of datasets is not fully clear. Additional information as this becomes available will be added in updates of this DMP.

Within the project, large hydro-meteorological and climate datasets will likely be accessed. However, only sub-sets (in time and space) will be used within this research. Wherever possible, the principle in the project will be to leave original datasets in their source repositories and avoid creating redundant copies of these original data. Sub-sets or processed datasets within the context of the research will be stored. Where possible, relevant scripts and procedures that allow processed data to be unambiguously re-created from the source datasets will be stored, thus avoiding the need to store data redundantly within the project data repositories.

2.4 Data repositories

I-CISK will adopt the following repositories for data storage, access and management:

- **The I-CISK Research Drive** is a collaborative platform that will be used by the consortium partners and invited collaborators to share internal information and upload intermediate versions of data sets and deliverables. This repository is established within the Research Drive² environment that is made available through the SURF network. SURF is the collaborative IT environment of the universities in the Netherlands, and an instance of Research Drive has been created to be used by the project free of charge within the SURF instance of the Vrije Universiteit (project partner VUA). Access is controlled by login and password

² <https://www.surf.nl/en/research-drive-securely-and-easily-store-and-share-research-data>

(dual factor authentication), and the instance created is for exclusive use of the officially designated project collaborators. Data uploaded to this repository will include data that is anonymised such that it cannot be traced back to the individual participants of the Living Labs. Data that has not yet been anonymised will also be uploaded to the Research Drive, with appropriate access control to that data so that it can only be accessed subject to authentication. Guidance on data processing procedures is provided in the following sections.

- **The Zenodo³ repository** will be used to publish project outputs, including datasets that are public and non-confidential. Using the Zenodo repository, all public data generated by the project will be published with a Digital Object Identifier (DOI) and a common dataset of metadata (based on the Dublin Core⁴). Keywords will be sourced by the standard dictionaries, like the USGS water dictionary. Additionally, Zenodo will assign persistent identifiers for public datasets and provide a long-term archival for these. Versions of each dataset will report main and minor changes with dot notation. Main version changes will occur after significant changes in the data (e.g. change in the data structure, massive correction or update, changes in the procedure for data collection or generation, etc.), while minor version changes will occur after data updates or limited correction. Any changes will also be mentioned in the description metadata field. Naming convention is reported, whenever applicable, in the data catalogue (see Appendix 1).

To allow project outputs, including project reports, deliverables and datasets to be easily found within Zenodo, datasets will be associated with the H2020 I-CISK Innovating Climate Services Community⁵ and the European Commission Funded Research (OpenAIRE) Community⁶.

- **The GitHub⁷ repository.** Software developed in the I-CISK project will be Open Source, and will be made available through GitHub. GitHub is the largest and most used development platform used to host open source software, and provides advanced tools for software version control. Software will be uploaded to a project space created within Github (https://github.com/icisk_eu).
- A key output of the project is the development of (pre-operational) climate services in the **I-CISK cloud web platform**. These services include information products that incorporate processed data. These information products may be developed using a combination of open access, proprietary and possibly anonymised personal data, and are publicly available. The I-CISK cloud web platform is formally not a data repository. However, as data from all three sources may be processed and are accessible to the public, the project will apply the same guiding principles to these data as outlined below for data that is published to Zenodo, as that is also a publicly available repository.

2.5 Guiding principles for data collection, storage and quality assurance

2.5.1 Legal Framework

Data collected, processed and published by I-CISK follows procedures that comply with the General Data Protection Regulation (GDPR)⁸. GDPR applies in all member states of the EU including the Netherlands, the EU

³ A data and publication repository, developed by CERN in the OpenAIRE project, freely available to all research programs. (<https://zenodo.org/>)

⁴ A set of “core metadata” for simple and generic resource descriptions (<http://dublincore.org>)

⁵ https://zenodo.org/communities/icisk_eu/

⁶ <https://zenodo.org/communities/ecfunded>

⁷ <https://github.com/>

⁸ Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32016R0679>

Member state where IHE as lead beneficiary and employer of the appointed Data Protection Officer is legally founded. The procedures I-CISK implements comply with GDPR as a minimum, or with national/regional regulations (including those in the LL outside the EU), and local customs and values. Where legal stipulations of the countries outside the EU are more stringent, the stricter definition will be applied. Data that has a human dimension will be subjected to appropriate anonymisation, compliance with GDPR. The GDPR grants specific rights to the individuals of whom personal data is processed within the project. This includes the right for any individual to:

- Obtain information on the processing of personal data;
- Gain access to the personal data that is held, and receive personal data in a machine-readable format ("data portability");
- Request incorrect, inaccurate or incomplete personal data to be completed, as well as that personal data is deleted when it is no longer required or if it is used in unlawfully;
- Request the restriction of processing of personal data in specific cases, and request that decisions based on automated processing concerning or affecting them as an individual and based personal data are made by natural persons, and not only by computers, thus reserving the right to express an opinion and to contest the decision.

The rights described above can be exercised by any individual involved in the research and at any time, through the direct counterpart within the I-CISK project with whom consent for participation in the research has been established as detailed below, typically the lead beneficiary for the Living Lab, as well as through the appointed Data Protection Officer, the name and contact details of whom is made available in the consent agreement.

2.5.2 Principles for collecting and processing of open access data

Open access data collected and processed to support the research and innovation actions in I-CISK can be collected (including sub-setting) and uploaded to the I-CISK Research Drive repository, together with appropriate metadata and attribution.

These data can be stored under the *Project Data (Open Access)* folder, which is accessible to the full consortium, and can be shared through a link on request by external parties.

Where processed data is explicitly identified to be used in products that could be considered for intellectual property, copyright, patents, etc, and where the licence of that data allows, then the data that can be stored in the restricted access folder *Project Data (Restricted Access)* and not shared or published before the end of the project.

2.5.3 Principles for collecting and processing of restricted access data

For data that is collected with restricted access, including proprietary data, a data sharing agreement should be established with the data provided, detailing how the data will be accessed, shared, processed and of any applicable restrictions or licences. A template for a data sharing agreement is provided in Appendix 2.

All restricted access data should be uploaded to the I-CISK Research Drive repository, together with appropriate metadata and attribution, and stored under specifically created folders under the *Project Data (Restricted Access)* folder within the I-CISK Research Drive repository. The access rights to this folder will be established by the manager of the repository (see section 5 Data security), and can be allocated or denied to specified user groups or individuals within the project.

Copies of restricted access data can only be held on the local drives using the Research Drive synchronisation and data access protocols (see section 5 Data security). All copies outside this secure environment should be deleted.

2.5.4 Principles for collecting and processing of personal data and data reflecting traditional and local knowledge

For the collection and storing of personal data, a detailed procedure is established to safeguard individual rights as well as ethical considerations. For all processes involving the collection of personal data, a clear audit trail is created, and stored on file under a specifically created folder under the *Project Data (Restricted Access)* folder within the I-CISK Research Drive repository. The access rights to this folder are set by the manager of the repository (see section 5 Data security), and can be allocated or restricted to specified user groups or individuals as appropriate.

The following steps are to be followed for each data collection activity that includes the collection of personal data and the participation of natural persons in the research.

- A clear description of the scope and purpose of the data collection activity must be documented, including the procedures and criteria be used to identify/recruit the participants to the research. The document detailing these procedures and criteria must be made available for review by the appropriate ethics committee of the project, and must be stored in the access restricted folder created to store the data.
- A procedure to gain explicit consent of natural persons that are subjects of the collection and processing of data must be documented, including the explicit provision that participation is restricted to adults and is fully voluntary. The document detailing this procedure must be made available for review by the appropriate ethics committees, and must be stored in the access restricted folder created to store the data. These procedures can be established following templates of informed consent/assent forms and information sheets available in the *Project Resources* section of the I-CISK Research Drive. Note that these are currently available in English only, but may be translated in applicable languages relevant to the respective Living Labs.
- Data that is collected from participants to the research that reflects their traditional and/or local knowledge is the intellectual property of those participants, or the holders of that knowledge. A documented agreement must be established with the knowledge holders about what knowledge will be used, for what purposes, who can use and share it (and how) and in general protected to avoid misuse by other people, or use without clear agreement with the knowledge holders. This is, for example recognized by the Convention on Biological Diversity, the Nagoya Protocol, and international conventions to protect cultural heritage. This is particularly relevant where use of traditional or local knowledge may be for commercial purposes. Knowledge holders should clearly understand how this knowledge, in original or in transformed form, will be used along the chain by the private sector or other users, and how the knowledge holders will have due credit and will benefit.
- In the preceding steps, documented procedures should be made available for review prior to implementation of the activity to the appropriate ethics committees (see also Deliverable D8.1). At a minimum this is the ethics committee that has been established by the project, chaired by the Deputy Project Coordinator. Where applicable these should be made available to the ethics committee or competent authorities of the beneficiary organisation of the researcher or researchers implementing the activity. Comments and approval from the review by the ethics Committee(s) must be documented and stored together with the data collected.
- The I-CISK Research Drive is hosted and accessible within the European Union. For data collected outside the EU, in particular personal data collected in the Living Labs in Georgia and Namibia, explicit confirmation must be sought with the mandated authorities that the data can be transferred in compliance with the national laws, and subjected to any processing as required (including anonymisation). This confirmation should be documented and stored with the data in the respective (restricted access) folder.
- Any personal data stored within the restricted access sections of the I-CISK Research Drive will remain confidential with access granted only to designated individuals or partners in the project. These data must be irreversibly anonymised before being made public and stored in any part of the repository that is

accessible by all beneficiaries and possibly shared with third parties, or uploaded to Zenodo as (part of) a published and publicly accessible dataset.

- Copies of personal data can only be held on the local drives using the Research Drive synchronisation and data access protocols (see section 5 Data security). All copies outside this secure environment should be deleted.

2.5.5 Data Quality Assurance

Quality assurance will be performed through the following steps:

- Collected data
 1. Storage of raw datasets (i.e. without any further processing) in a dedicated folder in the I-CISK Research Drive repository;
 2. Data check and editing for assuring positional, attribute and temporal quality, completeness and consistency, under the responsibility of the project partner listed as reference for each dataset in Appendix 1;
 3. Compilation of metadata (both generic and domain-specific, where applicable) reporting a brief summary with the editing done;
 4. Storage of the final version of the datasets in a dedicated folder in the I-CISK Research Drive repository;
 5. Uploading of the final version of the datasets on the Zenodo repository, if relevant for maintenance also after the project lifetime.

- Generated data
 1. Compiling metadata (both generic and domain-specific, where applicable) reporting a brief summary specifying the generation process (lineage);
 2. Storing the final version of all generated datasets in a dedicated folder in the I-CISK Research Drive repository;
 3. Uploading of the final version of selected datasets on the Zenodo repository, if relevant for maintenance also after the project lifetime.

3 FAIR Data Management

I-CISK will manage data in accordance with the FAIR Data Management Principles recommended by the European Commission. In this chapter we describe how the project implements FAIR principles, ensuring that the data generated by the project is accessible to the research community to be accessed and reused. Access and re-use of the data generated by the project is subject to whether data is public or confidential, and subject to the licence that is applicable to the particular dataset. Chapter 2 outlined procedures established within the project through which confidential data is managed to protect the privacy of the participants to data collection campaigns implemented by the project, in particular those with the stakeholders and actors in the Living Labs.

3.1 Making data findable, including provisions for metadata

In the I-CISK project the Zenodo repository is the main tool that is used to ensure the research data from the project is findable and can be re-used by other researchers, subject to licensing. This includes data that is generated by the research conducted in the project and related to specific project outputs, including publications and selected deliverables (where applicable). Datasets uploaded to Zenodo will be associated with the *H2020 I-CISK Innovating Climate Services Community* that has been established for the project, as well as the *European Commission Funded Research (OpenAIRE) Community*. Software that is created within I-CISK is open source will be made available through GitHub.

All datasets uploaded to Zenodo will be assigned a Digital Object Identifier (DOI), while Zenodo additionally provides version control. A naming convention will be used to identify datasets associated to the project. This format is yet to be finalised, but will follow a clear convention that contains: the status of the dataset (raw, processed, calculated); an identifier of the country or geographic region; the institution to which the author of the data is affiliated; a short description (two to three words, concatenated using camel case); the project acronym (H2020_icisk); and a version number.

The version number is structured as are version numbers for software: MAJOR.MINOR.PATCH (e.g. 1.02.03). Uploads to Zenodo of new versions should be done, where possible, under the same DOI as the original version.

3.2 Making data openly accessible

The term “Open Data”, technically defined as “data that can be freely used, re-used, and redistributed by anyone - subject only, at most, to the requirement to attribute and sharealike”, typically applies to textual and non-textual materials, including datasets, statistics, transcripts, survey results, and the metadata associated with these objects. More specifically, the Open Research Data Pilot of H2020, to which I-CISK adheres, applies to two types of data:

1. the data needed to validate the results presented in scientific publications, including associated metadata;
2. other data (e.g. curated data not directly attributable to a publication, or raw data), including associated metadata.

Data generated in I-CISK is by default and unless otherwise specified made available as open access data (see also Article 29.3 of the Grant Agreement). All data relevant to scientific publications as well as selected deliverables (where relevant), will be published in the Zenodo repository, together with associated metadata. Data with restricted access will be uploaded to the I-CISK Research Drive, which is accessible through a user identification to all project partners and if applicable accredited stakeholders. Any open-source software and

tool developed within I-CISK needed to access data will be published on Zenodo or other public software code repositories, like GitHub.

Aspects of accessibility are also respecting low bandwidth regions where data access is limited due to reduced download rates, as is expected for the Namibia Living Lab. If data access will be limited, other options for appropriate information delivery will be considered.

3.3 Making data interoperable

I-CISK strives to integrate data and information from different disciplines and domains. In order to provide a common understanding of data within the project itself, the use of the Dublin Core Metadata Element set vocabulary will be adopted. Whenever possible and useful, more discipline-specific metadata will also be adopted, as those defined by OGC for geospatial data. Where possible, interfaces shall use and be based on open standards (e.g. Web Coverage Service, OGC APIs) to maximise the interoperability of the data. I-CISK intends to streamline interoperability standards of the to the Copernicus Climate Change Service as well as the developments of the European Common data spaces, which are part of the European digital strategy.

3.4 Increase data re-use (through clarifying licences)

The I-CISK project follows the principle that datasets published by the project are open access, and can be accessed, mined, exploited and disseminated free of charge. This includes the use of data for commercial purposes. The datasets published by the project will be regulated using the Creative Commons Licensing⁹.

3.4.1 Licensing of I-CISK Datasets

By default, datasets will be provided under the CC-BY-SA (Attribution-Share-Alike), which allows anyone to adapt and redistribute the data, provided this is done so under the same conditions as the original licence.

For any dataset that is made available through a different licence than the default, for example CC BY-NC (Attribution Non-Commercial) which restricts the use of the data, this should be done so on agreement with the I-CISK innovation manager (Apostolis Tzimas, EMVIS).

In the case any commercial product (e.g. patent) is developed, relevant IPR and ethical issues will be addressed. For this, a detailed exploitation and business plan will be developed as indicated in the project proposal, in consultation with the appointed innovation manager.

3.4.2 Longevity of I-CISK Datasets

Datasets should be made available on the public repository (Zenodo) at the same time of publishing the publication that is supported by the dataset. For publications, this includes the time of publication of the pre-print.

Datasets published in Zenodo are currently stated to be re-usable on the platform for 20 years, with a guarantee that should the platform cease to exist, all data hosted by the platform will then be migrated to other repositories.

Public (including anonymised) datasets hosted on the I-CISK Research Drive will be hosted up to six months beyond the end date of the project (currently 31 October, 2025) to allow for additional time that may be needed for the completion of scientific publications stemming from the research in the project. An exception is made for images and videos, used under consent from participants where this includes images of natural persons, which will be kept for four years beyond the end of the project, in compliance with the Grant

⁹ <https://creativecommons.org/>

Agreement (article 26). Where consent is withdrawn, the data shall be deleted from the repositories without due delay.

Non-anonymised data sets (with restricted access) will be stored for as short as possible, and removed following anonymisation, unless a limited set of personal data should be kept for explicit reason. Any decisions to keep data longer should be documented and stored with the data.

For personal data that is stored (with restricted access), the data shall be deleted from the repositories without due delay in the case of any objection raised by the natural person or persons this data pertains to, and as agreed with the Data Protection Officer.

4 Allocation of resources

Data curation, storage, archiving, re-use and management within a secure environment incurs both direct costs and indirect costs, related to the IT infrastructure, and the staff time dedication both internal and external to the project to manage the repositories used. I-CISK has carefully selected repositories that do not burden the project with costs for the IT infrastructure.

- The Zenodo repository, which is employed for making datasets and research outputs FAIR is made available free of charge to H2020 research projects.
- The I-CISK Research Drive facility that is used to host (anonymised) datasets internal to the project partners, is a facility made available to collaborative research projects. This is made available free of cost to the project, provided that less than 2TB is used. Additional 2TB blocks can be made available at a cost of 500 EUR/year. We do not, however, foresee that the 2TB limit will be exceeded. It should be noted that the costs for Research Drive is supported by the Vrije Universiteit Amsterdam as institutional costs, with no separate charge to the I-CISK project.
- The Github repository for making software developed by the project available is also available free of charge.

Within I-CISK the responsibility for the management of data rests with the project coordinator, Micha Werner (IHE Delft), supported by the Data Management Officer (Ilias Pechlivanidis, SMHI).

5 Data security

The Zenodo repository, which will be used for publishing and maintaining final project outcomes, deliverables, and scientific publications, is hosted at CERN and it is subject to its rules for data security, as reported at <https://zenodo.org/policies>.

The I-CISK Research Drive which is used as the primary data repository to host both open access and restricted access datasets, has specifically been designed to offer a secure collaboration environment for project consortia such as I-CISK¹⁰. The I-CISK instance of Research Drive has been created as an access-controlled instance within the institutional Research Drive of the Vrije Universiteit Amsterdam (VUA) hosted by SURF¹¹, the collaborative IT organisation of universities and research organisations in the Netherlands.

Data on Research Drive is (backend) encrypted and access to the I-CISK instance of Research Drive is reserved to authorised persons, invited by the manager of the Research Drive instance. Access is user name and password authenticated, and as a standard dual-factor authorisation for access is activated. Access to specific folders created within Research Drive can be assigned to a specific group of users, or to a user individually. A sharable link can also be created to share the dataset with a third-party user, without access to the instance of Research Drive.

For each folder as well as files within a folder, the following access privileges can be granted. Note that access rights are inherited by sub-folders:

- Share: when enabled this allows a sharable link (URL) to a file or folder to be created
- Read: user has read access to the file or folder
- Edit: user can modify a file or folder
- Create: user can create a new folder or file (but not delete)
- Delete: user can delete a file or folder

To efficiently manage access to data, user groups have been created, with all members of these groups enjoying the same access rights.

- I-CISK_FullConsortium: All members of I-CISK registered on the I-CISK Research Drive (registration is done after invitation by the data manager(s)).
- I-CISK_WorkPackage#: All members of I-CISK identified as collaborating in a work package.
- I-CISK_LivingLab#: All members of I-CISK identified as collaborating in a Living Lab.
- I-CISK_ProjectCoordinationTeam: Participants in the Project Coordination Team
- I-CISK_ScientificTechnicalCoordination: Participants in the Scientific & Technical Coordination team

Additional groups can be established as the need arises. Any registered member may be identified as a part of multiple groups. Access to the I-CISK Research Drive is managed by the project coordinator (Micha Werner, IHE, or delegated assistant), supported by the Data Management Officer (Ilias Pechlivanidis, SMHI) and the Gender and Inclusivity Officer (Marije Schaafsma, VUA).

Research Drive allows data to be synchronised to a local folder of the computer of a registered user using the Own-Cloud sync client. For access restricted data within the I-CISK project this will only be done using the Cryptomotor. This will create a vault inside the sync folder within which the data is encrypted, and can only be used subject to authentication.

¹⁰ <https://www.surf.nl/en/research-drive-securely-and-easily-store-and-share-research-data>

¹¹ <https://www.surf.nl/en>

6 Ethical Aspects

I-CISK complies with ethical principles, internationally accepted scientific best practice and applicable international and EU law, to ensure high-quality research and enterprise culture, with the highest possible standards of integrity and practice. The research I-CISK undertakes will be conducted such that it meets the national legal and ethical requirements of the participating countries in which tasks raising potential ethical issues are to be conducted. I-CISK will comply with Decision No. 2013/743/EC, recognising the Charter of Fundamental Rights of the EU, adherence to ethical principles.

The research and innovation actions that I-CISK develops, requires the research consortium to collaborate with citizens, decision makers and other stakeholders in the seven Living Labs. In all these LL, human involvement will be limited to adults who are able to give fully informed and prior consent for their voluntary participation in the project activities (e.g. membership of Multi-Actor Platforms, participation in group discussions, interviews, workshops, role play games, demonstration of climate services). In all the project activities, the participants will be provided detailed information about the project objectives, methods and specific details of each activity, the role they are expected to play in specific activities and expected outputs and impacts. This will provide sufficient information to make an informed decision on participation. The I-CISK team will take responsibility for ensuring that all the participants are treated with equity and respect. Following the “data minimization principle,” a very limited amount of personal data will be collected from those adults who participate in the I-CISK project. This limited data pertains to name, gender, email contact, and representation (e.g. citizen, stakeholder group, decision maker, business sector). This data will be used with consent of the participants for the I-CISK activities and will not be shared beyond the project without their consent. These data will be stored in a restricted access & access-controlled partition of the collaborative platform.

Moreover, the “do no harm principle” and “anonymisation principle” will be followed when analysing data, and producing outputs such as publications, media reports and climate services products. This will ensure the protection of personal identities and avoidance of risks to any individual or group. I-CISK will not collect and/or process any sensitive personal data (e.g. health, sexual lifestyle, ethnicity, political opinion, religious or philosophical conviction). The research is not expected to lead to risk of misuse, stigmatisation of certain societal groups, political or financial retaliation. The research undertaken is exclusively focused on civil applications.

Two LL of I-CISK are outside of the EU (Georgia and Namibia). I-CISK will abide by the relevant national regulations for these countries and adhere to their research ethics and integrity guidelines. Similar to the LL in the European region, only minimum personal data will be collected from the participants from Namibia and Georgia. Additionally, no personal information will be shared beyond these LL without the consent of the participants. Moreover, the personal information will be anonymised prior to transferring research results into the European region (see also section 2.5 of this Data Management Plan).

I-CISK aims to share benefits emerging out of project activities among the project partners as well as citizens, stakeholders and decision makers in the Living Labs and beyond. For example, the pre-operational climate services developed by I-CISK in each Living Lab will be freely available through the I-CISK cloud web platform. The participants will also benefit from research results, which will be shared through various dissemination activities with all involved in the I-CISK project and beyond using means appropriate to end-user capacity and access.

Details on the defined processes in I-CISK to safeguard ethics and comply with the ethics requirements to the project are provided in Deliverable D8.1: Protection of Personal Data (POPD) & Humans (H). This includes defined processes with which data of natural persons is managed (see also section 2.5 of this Data Management Plan), as well as procedures and templates to identify and solicit consent from participants.

These templates are available on the I-CISK Research Drive and are annexed to Deliverable D8.1. We appreciate that the Living Labs include a variety of stakeholders as well as cultures, which means procedures may be adjusted to suit the local context. We will avoid issues of social profiling (including stigmatization), avoid pressure or undue influence, and provide an accurate and clear description of the study through Participant Information Sheets (see also consent above). If and when a sub-contracted party is involved in data collection, clear agreements will be put in place in advance about data transfer processes.

The project has established an ethics committee to address and monitor any ethics issues that arise. This committee is formed as a part of the Project Coordination Team, with the constitution described in Deliverable D8.1. Additionally, a Data Protection Officer has been appointed at IHE Delft (contact details are provided in Deliverable D8.1, which is confidential), or on request to the project coordinator (Micha Werner, IHE), the Data Management Officer (Ilias Pechlivanidis, SMHI) or the Gender and Inclusivity Officer (Marije Schaafsma, VUA).

Appendix 1 I-CISK Data

Sources of data in I-CISK

In I-CISK, scientific data will come from a variety of sources:

- Global/Regional Scientific hydro-climatic data, primarily sourced from open access repositories such as Copernicus Climate Data Store, GEOSS, as well as other global repositories.
- Local scientific hydro-climatic data, collected through the actors and stakeholders that form the multi-actor platforms established in each of the Living Labs.
- Social Data representing local knowledge and perceptions.

A concise overview of the specific data sources will be established as part of the operationalisation of the Living Labs. This overview is available in a shared document that can be found on the I-CISK Research Drive, and is accessible (Read only) through this link: <https://vu.data.surfsara.nl/index.php/s/QWvniuQKMPg315B>. Note that this overview is a living document and will evolve as the Living Labs are developed.

Metadata in the data catalogue

The data catalogue¹² details the data requirements that I-CISK partners have for each Living Lab. The columns (letters from A to Y) specify the characteristics of the Living Lab, as well as the data required (for validation and benchmarking), specifically:

- Columns A to E describe the Living Lab, reporting the following background information:
 - Living Lab:** LL name
 - Impact indicator:** Quantitative impact indicators used for quantifying the value of climate services
 - Impact model:** Existing impact models used to transform the climate-related decision triggers into the sector-specific impact indicators (using also other non-climate-related existing datasets for different Living Labs)
 - Decision trigger:** Climate-related decision variable or index used to define the impact indicator
 - Existing climate services or hydro-meteorological forecasts:** Yes/No/Not Available (Y/N/NA) - Existing local climate services
- Columns F to K describe the characteristics of the data that is used for validation of the benchmark climate service:
 - Variables for validation/supervision:** Climate or land surface variables used for the validation of the benchmark
 - Reference dataset for validation/supervision:** Dataset used for the validation of the benchmark
 - Data accessibility:** the official source from where to obtain the data, including format
 - Temporal resolution:** the temporal resolution of the validation data
 - Spatial resolution:** the spatial resolution of the validation data
 - Time coverage:** the time span covered by the validation data
- Columns L to P describe the data that partners ideally would use ('wish list') for the different Living Labs, including details of what characteristics the desired data should have:

¹²<https://docs.google.com/spreadsheets/d/1sHQYdMUBIXWmDHFc7C1qrT2kmmfhuCU0aPehfVdlmZE/edit#gid=1422100599>

- L. **Climate/land surface variables needed:** Climate or land surface variables used to estimate/simulate/predict the decision trigger
- M. **CS time horizon:** Time horizon of the related climate service, with options: i) sub-seasonal, ii) seasonal, iii) decadal, iv) projections, v) "observed" (or lag-0)
- N. **Time coverage (events):** Time span the data should cover (include in parenthesis specific events that may be needed, outside of this span)
- O. **Temporal resolution:** the desired temporal resolution of the climate variables required
- P. **Spatial resolution:** the desired spatial resolution of the climate variables required.
- Columns Q to Y describe the benchmark, i.e. large-scale climate services, used for the different Living Labs:
 - Q. **CS benchmark (models + data):** Name of the benchmark system used (i.e. combination of existing models + data)
 - R. **Post-processing:** Type of post-processing (e.g. bias adjustment)
 - S. **Reference dataset for post-processing:** Reference dataset used for post-processing
 - T. **Data accessibility (provider):** Official source from where to obtain the data (public data repository)
 - U. **Internal availability:** Data directly available from any I-CISK partner? Y/N (contact details)
 - V. **Temporal resolution:** the temporal resolution of the benchmark data
 - W. **Spatial resolution:** the spatial resolution of the benchmark data
 - X. **Time coverage:** the time span covered by the benchmark data
 - Y. **Notes:** Other key features of the benchmark data or comments (e.g. initialisation, lead times, ensemble members, RCP scenarios, etc.)

Appendix 2 I-CISK Data sharing agreement

The following template will be used to facilitate data exchange between the project partners and stakeholders providing data. This template is also provided as a stand-alone document on the I-CISK Research Drive in the Project Resources folder. This template may be amended to suit the local context, including translation into a more appropriate language if suitable. Note also that the template that is available on the I-CISK Research Drive will be refined in the course of the project and the template below reflects status as the time of writing this DMP. It is recommended that when establishing a data sharing agreement the template file from the I-CISK Research Drive is used, and amended when required to suit the local context. Note also that this data sharing agreement is applicable to proprietary data, as for open data, a data sharing agreement is not needed.

I-CISK Research Agreement

Parties

A Partner A (“**Abbrev**”), org. No. 123456-1234, address,
and
B Partner B (“**Abbrev**”), org. No. 123456-1234, address,

A-B are hereinafter collectively referred to as “Parties”, and individually as “Party”.

Background

Partner A is involved in the research project funded by the European Climate, Infrastructure and Environment Executive Agency called “Innovating Climate services through Integrating Scientific and local Knowledge” (I-CISK), Grant agreement 101 037 293 (as further described below, the “Project”).

One of the purposes of the Project is to ...

If applicable, state the organisations that will get access to the data via Partner A (the “Project Members”)

Add scientific rationale for the data exchange.

Parties therefore enter into this research agreement, which includes this background, in order to specify rights and responsibilities.

Article 1, Partner B’s obligations

State Partner B’s obligations: dataset, variables, period etc. to be extracted.

Article 2, Partner A’s obligations

State Partner A’s obligations.

Article 3, Payment

No monetary payment is involved in this agreement.

Article 4, Agreement Conditions, Data Licence and Data Management

Add conditions, data licence requirements etc. For example:

1. **Partner A** is granted a limited licence to use the Data in the Project. **Partner A** is granted a right to sub-license the Data to the Project members in accordance with the limitations stated in this Article 4. The Data shall only be used by **Partner A** and the Project members stipulated in this Agreement. The Data shall only be used within the Project and must, in all its existing forms and copies generated, be destroyed after Project ending.
2. **Partner B** shall be offered to participate as co-author in all publications based on the Data. **Partner A** or the aforementioned Project members should make direct contact on this matter with Contact Information.
3. **Partner B** makes no representation and gives no warranties about the Data and any reliance on them by **Partner A** or any Project member will be at their own risk.
4. The Data should be referred to in publications by citing the following papers:
 - Paper A...

Article 5, Amendments

Amendments to this agreement shall be made in writing and agreed between the Parties.

Article 6, Miscellaneous

Matters related to this Agreement, such as practical arrangements regarding the delivery of Data, shall be discussed and decided upon in a cooperative spirit. Neither Partner would hold the other liable to litigation due to this agreement. Differences or conflicts arising as a result of a breach in the agreement are to be resolved through dialogue.

Article 7, Contacts

The following people are the designated contact points for this Agreement:

Partner B

The contact for scientific work related to this agreement is *Contact information*.

The administrative contact for this agreement is *Contact information*.

Partner A

The contact for scientific work related to this agreement is *Contact information*.

The administrative contact for this agreement is *Contact information*.

Article 8, Entry into force

This Agreement shall enter into force when signed by both Parties and shall regulate the time period equivalent of the Project time period.

(Signatures on the following page)

This Agreement has been made in two (2) original copies, of which the Parties have taken one each.

Place, Date	Place, Date
Signature	Signature
_____	_____
Title, Name	Title, Name
Position	Position
Department	Department
Affiliation	Affiliation
Telephone:	Telephone:
Email	Email

Appendix 3 Access to I-CISK Research Drive

Access to the I-CISK Research Drive is by invitation only, from the data managers. An email request is sent from within the system to the invited user, with this email providing details on how to activate the account.

Step 1: Request access to the I-CISK Research Drive by contacting the project Coordinator (Micha Werner) or deputy coordinator (Ilyas Masih).

Step 2: Activate the account through the link in email received (see example below) “Create your Research Drive account”)

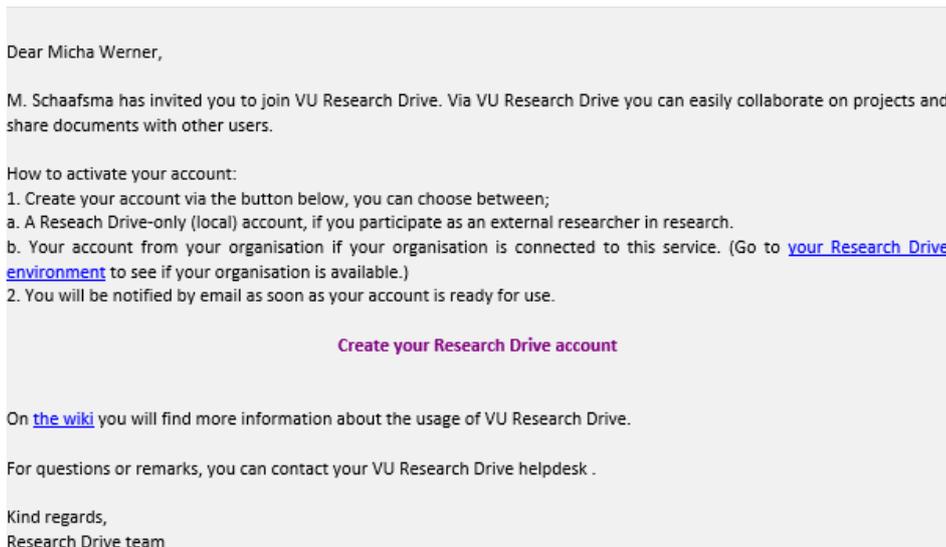


Figure 1. Invitation email to Research Drive

Step 3: You will be provided with options how to login. Select the “Set password” option and follow the instructions. Follow the instructions on the email provided to access the platform.

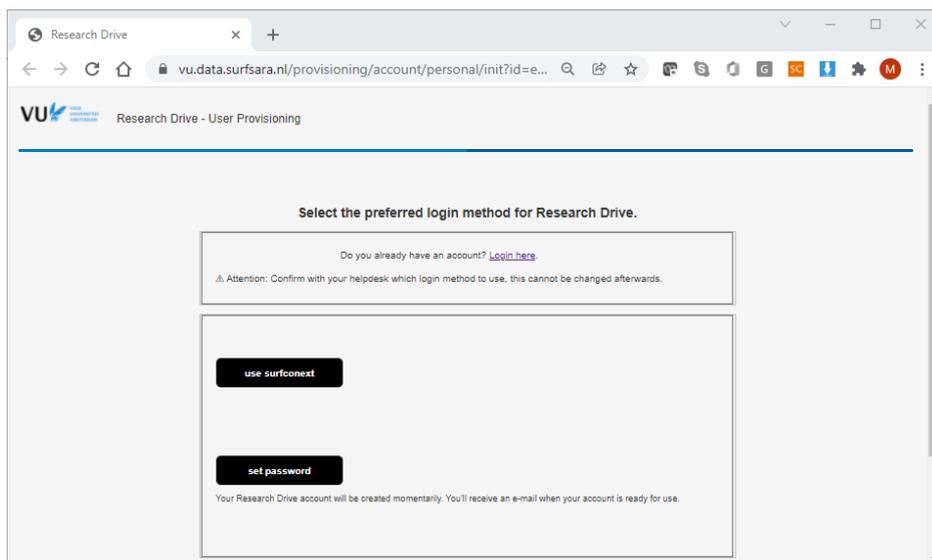


Figure 2. Select the “set password” option.

Step 4: Log in to the platform. You are requested to set up:

- A new password to replace the default provided.
- Set up two-factor authentication.

Please see the instructions on the Research Drive Wiki. Tutorials are provided to explain how a password is changed as well as how two-factor authentication is set up.

<https://wiki.surfnet.nl/display/RDRIVE/SURF+Research+Drive+wiki>

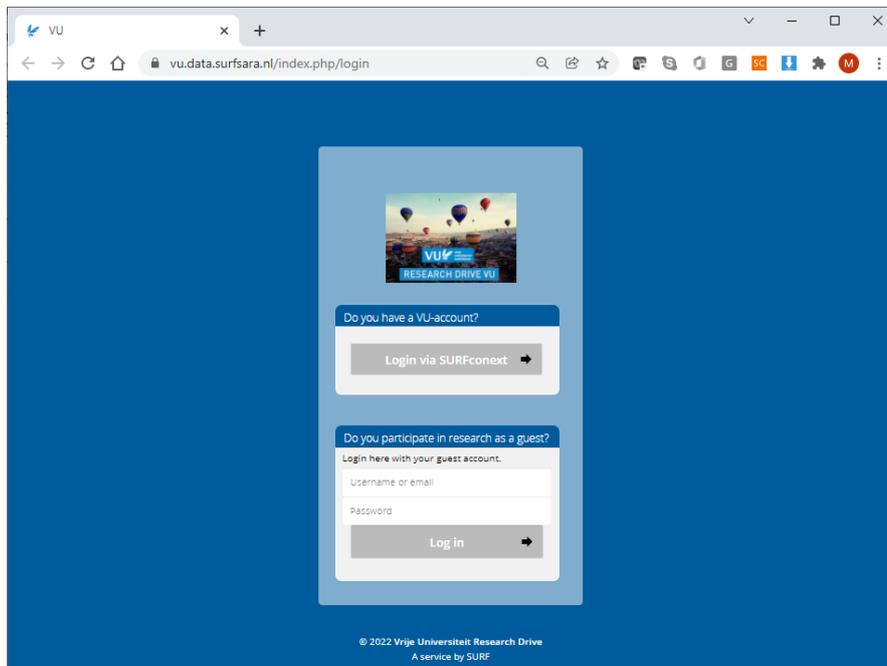


Figure 3. Log in with your guest account.

Step 5: Log in to the platform using your new password and two-factor authentication.

This will give access to the main folder structure. Where access is required to specific folders, through for example inclusion of your user name (email) in a specific group, this will need to be requested from the data managers.

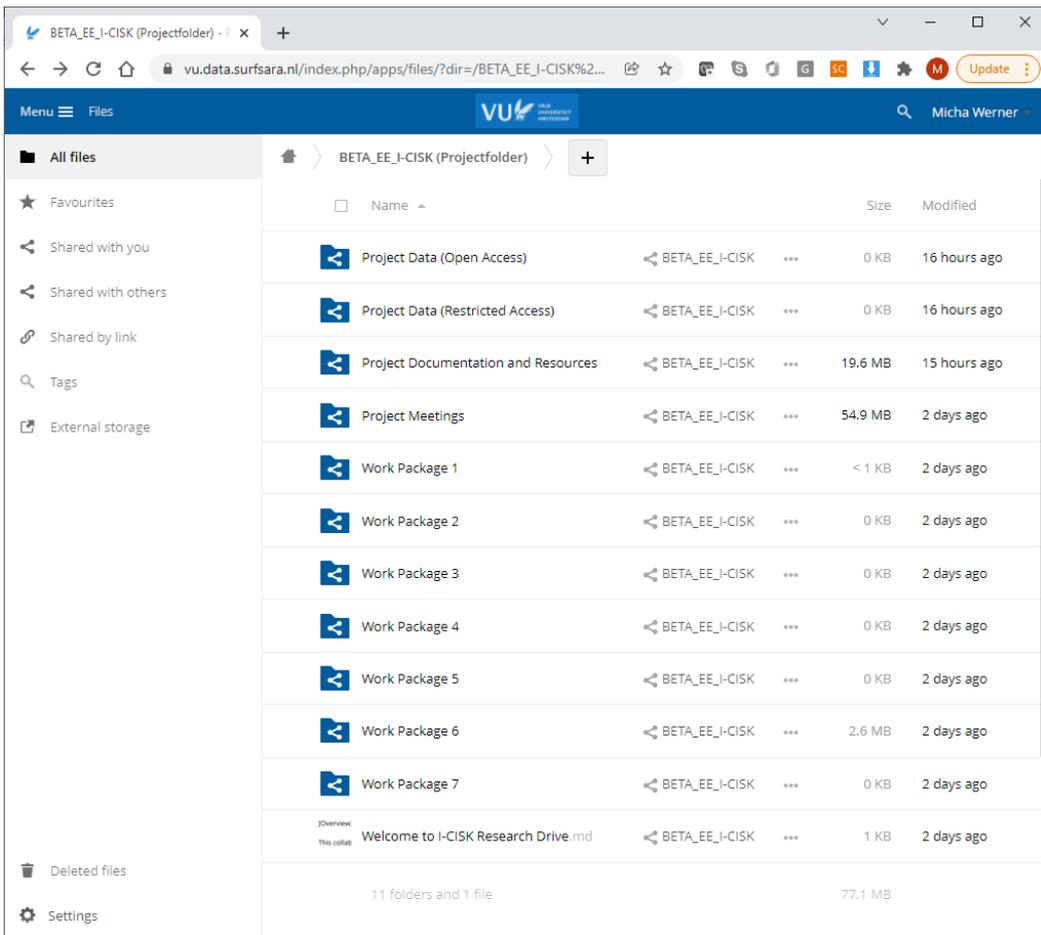


Figure 4. I-CISK Research Drive main folder structure.



I-CISK

HUMAN CENTRED CLIMATE SERVICES

Colophon:

This report has been prepared by the H2020 Research Project “Innovating Climate services through Integrating Scientific and local Knowledge (I-CISK)”. This research project is a part of the European Union’s Horizon 2020 Framework Programme call, “Building a low-carbon, climate resilient future: Research and innovation in support of the European Green Deal (H2020-LC-GD-2020)”, and has been developed in response to the call topic “Developing end-user products and services for all stakeholders and citizens supporting climate adaptation and mitigation (LC-GD-9-2-2020)”. This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 101037293.

This four-year project started November 1st 2021 and is coordinated by IHE Delft Institute for Water Education. For additional information, please contact: Micha Werner (m.werner@un-ihe.org) or visit the project website at www.icisk.eu

