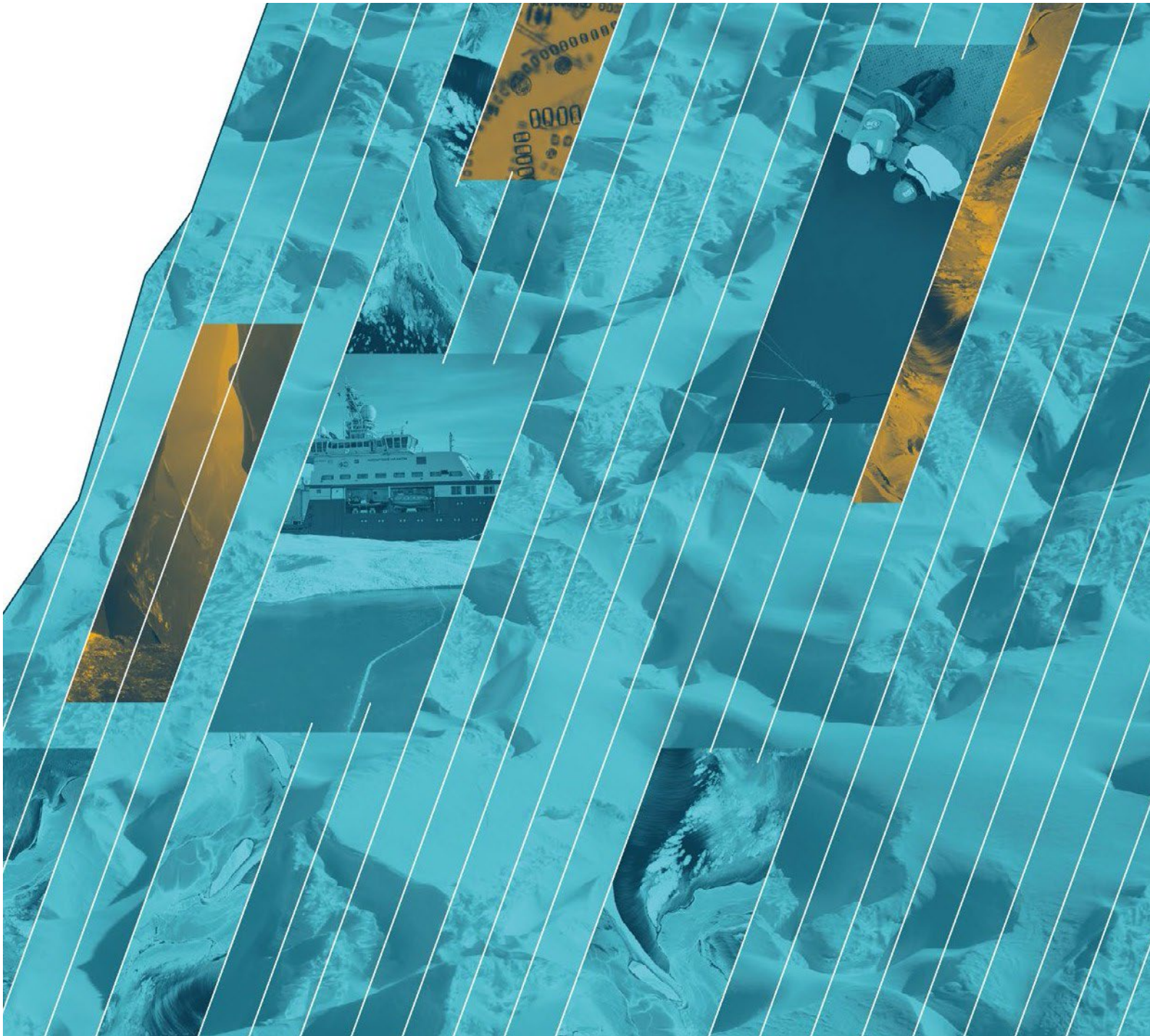


iC3

Research Data Management Plan



iC3 Report Series, Volume 01 (2025)

Title of the report: iC3 Research Data Management Plan

Year: 2025

List of authors: Sarti, Fabio^{a*}; Wadham, Jemma^a; Winsborrow, Monica^a; Assmy, Philipp^{a,b}; Langebroek, Petra^{a,c}; Souster, Terri^a

Authors' address (Institution):

^a iC3: Centre for ice, Cryosphere, Carbon and Climate, Department of Geosciences, UiT The Arctic University of Norway, 9037 Tromsø, Norway

^b Norwegian Polar Institute, Tromsø, Norway

^c NORCE Norwegian Research Centre, Bjerknes Centre for Climate Research, Bergen, Norway

*Correspondent author's email address: fabio.sarti@uit.no

Keywords: data management plan, DMP, research data management plan, RDMP, FAIR principles, open data, data lifecycle, data storage, data sharing, data use, data preservation, metadata

iC3 Report Series, Volume 01 (2025)

To be cited as: Sarti, F. et al. (2025). iC3 Research Data Management Plan. *iC3 Report Series, Volume 01*.
<https://doi.org/10.7557/30.8014>

© The authors. This report is licensed under the Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>)

ISSN: 2704-2065

Publisher: Septentrio Academic Publishing Tromsø Norway

Versions:

Version 0 (12.2024)
Comments First extended for internal use

Version 1 (02.2025)
Comments First published online

Table of Contents

Introduction.....	1
Centre’s presentation	1
Funding and organization	1
Research data management: responsibilities, rights and data ownership	1
Research data management plan: preliminary information	2
1 Research question	4
2 Data search/reuse	4
3 Data Management Plan.....	4
4 Description (Collection, Analysis, Data Generation, Storage).....	4
4.1 How new data will be collected.....	5
4.2 How new data will be generated.....	5
4.3 Analysis: protocols.....	6
4.4 Data quality control.....	6
4.5 Data Storage.....	6
4.6 Storage areas and features.....	6
4.7 Preparing data for Archiving	7
4.8 Data collection mapping	7
5 Archive (long term).....	8
5.1 Data archiving and sharing	8
5.2 Data availability	9
5.3 Data interoperability	9
5.4 Data re-use: licensing.....	10
6 Publication	10
6.1 Documental repositories	10
7 Ethics, codes of conduct, personal data and consent	11
Appendix A: sub-Data Management Plans.....	13
Appendix 1: iC3 expeditions	18
References.....	25

Introduction

Centre's presentation

iC3 Centre for ice, Cryosphere, Carbon and Climate (hereafter iC3) aims to fill a major research gap in polar science by quantifying the future impacts of ice sheet change on the Earth's carbon cycle over policy-relevant timescales. It achieves this by bringing together the complementary world-leading expertise of the UiT The Arctic University of Norway (UiT), the Norwegian Polar Institute (NPI), the Norwegian Research Centre AS (NORCE) and a network of collaborators in an unprecedented research effort spanning both the Arctic and Antarctic. Its integrated, interdisciplinary hub of experts studying the cryosphere, oceans, atmosphere and geosphere will close order-of-magnitude uncertainties in polar carbon budgets, addressing the hypothesis that changing ice sheets (and the associated cryosphere) profoundly affect the Earth's carbon cycle, with direct implications for human societies through feedback on our future climate and invaluable polar ecosystems. iC3 will address this major challenge by using excellent Norwegian infrastructure (Ny-Ålesund Research Station, Kongsfjorden, Svalbard; Troll Station, Antarctica; Research Vessels Kronprins Haakon and Helmer Hanssen; INES, Infrastructure for Norwegian Earth System modelling) and innovative technologies to collect and integrate novel data sets at both poles with state-of-the-art numerical models to assess future impacts at regional to global scales.

Funding and organization

The project is hosted at the Department of Geosciences (IG), UiT Faculty of Science and Technology (Nt-Fak) and it will span 10 years from 1st of July 2023 to 30th of June 2033. After completion of the first 5 years a mid-term evaluation will occur in year 6 and enable the second part (years 6 to 10). iC3 is a stand-alone Research Centre that is neither part of a larger project nor the continuation. External funding is provided by the Research Council of Norway (RCN) through its Centres of Excellence funding scheme (project number 332635¹). In-kind contributions, including working hours, equipment, and running costs, are provided by UiT, NPI and NORCE.

The centre comprises 5 Research Units (RU) and 2 Impact Teams (IT) focused on research, innovation, technology, and developing future leaders. The researchers affiliated with iC3 are from different institutions at national and international levels.

Research data management: responsibilities, rights and data ownership

Several responsibilities are identified by this plan and attributed to specific individuals.

- Responsibilities for DMP oversight and associated actions:
 - Wadham, Jemma; Director; jemma.l.wadham@uit.no

¹ [RCN Project Bank](#)

- Winsborrow, Monica; Director; monica.winsborrow@uit.no
- Sarti, Fabio; Data Manager; fabio.sarti@uit.no
- Souster, Terri; Project Manager; terri.souster@uit.no
- Responsibilities for data collection, documentation, archiving:
 - Research Unit and Impact Theme leaders
(with support from Data Manager and oversight by iC3 Directors)
- Responsibilities for published datasets:
 - first or corresponding authors
(with support from Data Manager)
- Administrative responsibilities
 - Souster, Terri; Project Manager; terri.souster@uit.no

Any other responsibilities not previously mentioned are defined in the 3rd party agreement.

This document also includes rules regarding a) data management rights and b) data ownership.

- a) During the project period, all iC3 members are granted access to data. However, the responsibility for managing and granting access to this data lies with the Research Unit and Impact Theme leaders as well as associated scientists, aided by the Data Manager and overseen by iC3 Directors.
- b) Data ownership in a research centre where multiple organizations collaborate is not a one-size-fits-all scenario and can be complex. A formal consortium agreement among the iC3 project partners outlines the formal data ownership agreement. If specific cases not covered by the agreement occur, data ownership is located at the parent organization of the most relevant Institution involved in such a case. (the relevant Institution will be identified by the committee iC3 Directors and Partner representatives).

Research data management plan: preliminary information

The requirement for a Data Management Plan (DMP) for RCN-funded projects was first introduced in Norway in January 2018, and all directly funded iC3 research activity follows this recommendation. The iC3 DMP's purpose is not simply to fulfill an obligation, but to provide clear guidance to iC3 members, to ensure that data is proficiently managed at all stages of the project and that it remains accessible, secure and sustainable in the long term. Furthermore, it will ensure better project implementation by providing a clear overview of all data management activities. To achieve these goals the DMP includes measures to comply with FAIR principles² and OpenScience³ by Sikt⁴ and to be in-line with principles and guidelines for management of research at

² <https://www.go-fair.org/fair-principles/>

³ <https://www.openscience.no/en>

⁴ <https://sikt.no/en/home>

UiT⁵. The template utilized to develop this document (provided by UiT/Library and inspired by Science Europe⁶) has been adapted to encompass all stages of the data management lifecycle, as shown in Figure 1 below (from "Research Data Management Life cycle" by The University of California, Santa Cruz, Data Management LibGuide⁷, which is licensed under CC BY-NC 4.0⁸):

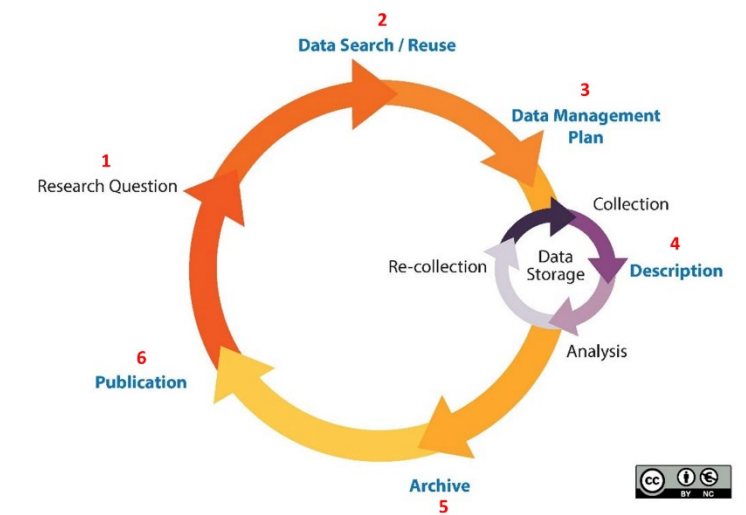


Figure 1: The Research Data Management Lifecycle

The iC3 DMP will be periodically reviewed and updated to ensure that it remains relevant and effective. As recommended by The Research Council and many other funders, the latest version of the plan is freely accessible online in Septentrio Reports (see title page for reference).

The use of sub-DMPs is allowed and expected in order to satisfy possible requirements for dedicated DMPs. This could include, for example, data infrastructures at external providers or PhD projects. In such cases, sub-DMPs will be implemented in accordance with Guidelines for follow-up of data handling plans⁹ and added in form of excerpts to the present document (see Appendix A); prior to the inclusion, sub-DMP contents will be evaluated in order to avoid contradictions between the two plans; cross citation between the documents will formally relate DMP and sub-DMPs.

Cost related consideration: when selecting infrastructures or services for data handling, such as computing, storage, and archive, iC3 has to consider the current price model offered by providers at local and national

⁵ [Principles and guidelines for management of research at UiT](#)

⁶ <https://www.scienceurope.org/our-resources/practical-guide-to-the-international-alignment-of-research-data-management/>

⁷ <https://guides.library.ucsc.edu/openscience>

⁸ <https://creativecommons.org/licenses/by-nc/4.0/?ref=chooser-v1>

⁹ [Guidelines for follow-up of data handling plans](#)

levels. These are user-contribution models that require reflections about how to maintain services beyond the project, when required. The same approach applies to any new services that iC3 may introduce for data management purposes.

1 Research question

Any type of research always begins with developing a research question by gathering relevant information from various sources (literature, data previously collected, and other materials), creating notes, planning, adopting methods, deciding on strategies, and improving protocols and routines. The result is a preliminary set of information that should be safeguarded carefully, especially within a context like iC3, where researchers are located across various institutions, and data management lacks visibility of the digital environment for the entire research team.

When it comes to information security and data protection, it is recommended that the following guidelines be adhered to: a) refrain from storing files on personal computers, b) utilize collaborative services for file sharing (such as Microsoft SharePoint), c) ensure that backup and versioning functionalities are active for said services, d) avoid exchanging data via hard drives, e) remember to assign a level of confidentiality to each document (e.g. open, internal, confidential; further information can be found in paragraph 4, Data storage).

2 Data search/reuse

iC3 researchers may seek data that has been previously collected or processed by others in order to integrate these into their studies. In such cases, it is important to a) check descriptions and features of data sources at registries such as re3data¹⁰, b) verify the compliance of data repositories with Open Science principles, c) if applicable, follow terms for data reuse, and d) cite data sources. If such data is not available in repositories, efforts to secure and make them available to the public will be undertaken.

3 Data Management Plan

Aims, scopes and sources of this DMP have been detailed in paragraph **Research data management plan: preliminary information**.

4 Description (Collection, Analysis, Data Generation, Storage)

The fourth part of the data lifecycle “Description” is essential to the DMP as it outlines the research stage in its most active phase. This involves collecting, analyzing, processing and ultimately storing the data.

¹⁰ <https://www.re3data.org/>

4.1 How new data will be collected

A significant proportion of the data managed by iC3 consists of newly collected information from scientific expeditions. These expeditions may be cruises or field campaigns arranged or co-arranged by the Centre. Additionally, other datasets will be obtained during non-iC3 led expeditions (e.g. led by collaborators and contractors). Such research activity will take place over the entire Centre period and allow iC3 to collect a wide and diverse range of marine and terrestrial datasets from remote polar regions. These will include, but are not restricted to, glaciological, biogeochemical, oceanographic, biological and geological datasets. Data will be collected and generated using standard conventions commonly recognized in the relevant scientific community. Where no clear standards exist, these will be developed in collaboration between iC3 scientists and the Data Manager. Appendix 1 outlines each expedition with specific emphasis on data. Exhaustive reports will be accessible to all after a necessary publishing workflow (refer to Chapter 6, Publication), detailing strategies and procedures for data collection and preparation.

Data loss risk during iC3 expeditions may be classified as high and poses a critical issue. Reliable IT infrastructures are provided by research vessels to ensure data safety during scientific cruises, but data transfer to final storage areas (refer to chapter 4.3: Storage areas and features) solely relies on portable hard drives or similar media. Moreover, researchers in remote field locations often depend on laptops and hard drives. This document urges researchers to promptly upload data to secure storage facilities immediately after each expedition.

4.2 How new data will be generated

An equally considerable part (in size) of the data managed in iC3 is process data, especially for Global Carbon prediction, which utilizes models applied to data previously collected or made available by the Centre. The need for High-Performance Computing infrastructures (HPC) has arisen within iC3 and a working group consisting of researchers and IT specialists has been established with the goal to identify potential solutions and associated budget. The selected service is distributed by Sigma2¹¹, also known as NRIS (Norwegian research infrastructure services). The infrastructure consists of clusters of servers with high processing power and large memory capacity servers. These are conveniently interconnected and work in parallel to perform massive calculations efficiently, allowing large dataset processing at incredibly high speeds. The infrastructure is offered as a service, to which researchers apply for computer time (CPU hours). Data storage, software support and data archiving are complementary services included. Further information is available on the HPC webpage¹².

¹¹ <https://www.sigma2.no/>

¹² <https://www.sigma2.no/service/high-performance-computing>

As recommended by Sigma2, an ad hoc DMP has been issued to better fulfill the infrastructure requirements. An extended excerpt of it is enclosed with this document as Appendix A where methods for data processing and the entire data cycle for this section are illustrated.

4.3 Analysis: protocols

Chapter for future uses

4.4 Data quality control

Chapter for future uses

4.5 Data Storage

Data storage concerns information security of raw data and datasets during later stages of processing, elaboration and analysis. Confidentiality, data integrity and accessibility are crucial at this stage, along with additional considerations such as data sharing and finalizing data for the subsequent phase of archiving.

Confidentiality: this means that data is accessible only to authorized users. At iC3, researchers are required to classify files as open, internal, confidential, or strictly confidential. In this data storage phase, the most appropriate classification for research data is internal; for personal data instead, a higher level of confidentiality will be applied: more details will be delivered at chapter **7 Ethics, codes of conduct, personal data and consent**). Data integrity and accessibility will be addressed in the following paragraph 4.6.

4.6 Storage areas and features

Although the distribution of iC3 researchers across different organizations would suggest the adoption of a centralized storage facility, a less invasive storage policy will be implemented at the outset. Reaffirming that storage of data on private laptops, stand-alone hard drives, or external storage devices including USB sticks is not recommended by any organization, Research Units and Impact Themes are free to use preferred storage services or those recommended by their respective Institutions.

When handling iC3's data, the Centre recommends that its members use robust data storage services, preferably managed by IT Departments, which feature automatic backup and protection against unauthorized change and deletion (Integrity). Often, these services also enable collaboration and data sharing among other groups or individual external users (Accessibility). For secure file sharing of large datasets, Filesender by Sikt¹³ is an easy way to transfer files to one or more recipients by using a web browser. The service boasts positive tests for files up to 1000 GB and it is available to researchers based in Norwegian institutions who can either send or grant access to anyone by sending a Filesender voucher.

¹³ <https://sikt.no/tjenester/filesender#> (only Norwegian) or <https://filesender.sikt.no/>

Paired to the previously described HPC service (see Par. 4.2), a storage area at the national infrastructure NIRD Data Storage¹⁴ is available for modelling activities. NIRD Data Storage is specifically designed for data used actively, it is designed to allow intensive I/O operations and to support high-performance computing.

Data integrity is safeguarded by daily snapshots, built-in redundancy, error-correcting mechanisms, and user-controlled backups to protect valuable datasets from human errors and disruptions.

4.7 Preparing data for Archiving

Preparing data for archiving is a process carried out along all phases of any project that adheres to a data management plan. It should start at an early stage, before collecting data, with the purpose of ensuring that data can be understood by anyone, even multiple years after publication. The list presented below is a collection of good practices to be persistently followed during the research process: they will benefit the final archiving phase where data must be deposited and thoroughly described:

- Naming:
 - iC3 convention: apply the prefix iC3_yy_## (yy=years; ## progressive number 01 to xx) to expedition, file, folder and sample names;
 - other conventions are allowed (e.g., when participating in expeditions led by other organizations)
 - names: descriptive but up to 25 characters, no special characters, underscores instead of spaces, dates in international dating convention YYYY-MM-DD;
 - naming strategy: by choosing a structure like date and description or description and date or numbering filenames;
 - keep folder names in file names;
 - observe all the rules consistently and describe them in ReadMe files;
- Documentation to help understanding and re-using data:
 - prepare human readable introduction to datasets in form of ReadMe files containing project general information, content overview, methodology, information about data;
- Selecting data to provide a comprehensive set of data:
 - include raw and processed data;
 - include discarded data (collected and elaborated but that do not support the hypothesis)

4.8 Data collection mapping

Web map (under discussion)

¹⁴ <https://www.sigma2.no/data-storage>

5 Archive (long term)

Long-term preservation of research data is essential for iC3 data legacy. It ensures transparency, reproducibility, accessibility and sharing of all data collected or generated by the Centre during its lifespan. Furthermore, it will enhance iC3's overall impact within the scientific community and increase the number of citations.

5.1 Data archiving and sharing

To allow the necessary possibility of choice, a set of three repositories have been identified for iC3 long term data archiving. DataverseNO¹⁵, which is one of the over 100 installations worldwide that use the Open-source research data repository software provided by The Dataverse Project¹⁶. DataverseNO is a national infrastructure where UiT has its institutional collection, it is aligned with FAIR Guiding Principles for scientific data management and stewardship, and it is CoreTrustSeal-certified as a trustworthy and sustainable repository. No costs are associated with the service. In DataverseNO, each dataset receives a Persistent and Unique Identifiers (PID) in the form of a Digital Object Identifier (DOI) and a ready-to-use reference for publications. Data preparation and curation will be supported and supervised by curators at UiT/Library to ensure correct documentation, standardization and licensing. Datasets archived in the repository will become discoverable by search engines such as B2FIND – EUDAT¹⁷, Bielefeld Academic Search Engine (BASE)¹⁸, DataCite Search¹⁹, Google Dataset Search²⁰, Oria²¹ (Findability) or findable by other services like Open Polar²², a web map-based service jointly developed by UiT and NPI about open scientific results in Arctic and Antarctic regions. DataverseNO also features version control, which records all changes made to the dataset after initial publication, safeguarding data integrity.

In addition to DataverseNO, the Norwegian Polar Data Centre (NPDC)²³ is recommended. NPI manages and operates the repository, which is also well aligned to FAIR principles. No costs are associated with the service,

¹⁵ <https://site.uit.no/dataverseno/>

¹⁶ <https://dataverse.org/>

¹⁷ <https://b2find.eudat.eu/>

¹⁸ <https://www.base-search.net/>

¹⁹ <https://datacite.org/>

²⁰ <https://datasetsearch.research.google.com/>

²¹ https://bibsys-almaprimo.hosted.exlibrisgroup.com/primo-explore/search?vid=UBTO&lang=en_US

²² <https://openpolar.no/>

²³ <https://data.npolar.no/about>

each dataset receives a Persistent and Unique Identifiers (PID) in the form of a Digital Object Identifier (DOI) and a ready-to-use reference for publications.

The third repository for data archiving has been specifically selected to align with modelling HPC data infrastructure: NIRD Research Data Archive²⁴ by Sigma2. The archive is the largest archive in Norway (December 2023), it is free of charge and currently assures compliancy with FAIR principles. However, an extensive renovation project has been carried out by the provider with the aim to rise and consolidate FAIR compliancy at a higher level: as a result, The Next Generation Research Data Archive will be set in production by autumn 2024. Given the early stage of model-based HPC infrastructure, Research Data Archive description will be postponed to a subsequent version of the present document.

It is possible that researchers at iC3 may occasionally have preferred archiving needs. In such instances (to be regarded as exceptions), it is recommended to use search engines (e.g. re3data) to assess the FAIR compliance of the chosen repository.

5.2 Data availability

Even if iC3 adopts an Open Data Policy, some data may have access restrictions (or embargos, e.g., data under use in doctoral dissertations for max 4 years after collection or under other exceptions regulated in the consortium agreement). Restricted datasets are announced at the relevant expedition recap in Appendix 1. For data linked to a scientific publication, data will be made available no later than the time of publication. Any other data that may be of interest for research will be made available at the end of the project.

5.3 Data interoperability

All data will be converted to the appropriate persistent formats (preferred) to ensure long-term interoperability. The following table summarizes the most relevant conversions:

type	extension recommended	extension not recommended
text	.txt	.docx
slides	.pdf/A	.pptx
spreadsheet	.txt	.xlsx
image	.tiff, .png, .jpeg	.gif, .raw, .bmp
audio	.wav, .mp3	.m4a, .wma
video	.mpeg	
database	.xml/txt	.accdb

²⁴ <https://archive.sigma2.no/pages/public/search.jsf>

data	.csv
oceanographic data	.netCDF

Table 1: Conversion table original to persistent filenames

If non-persistent file formats do not have a corresponding persistent file format, documentation will be provided in the ReadMe file on how the files can be accessed.

5.4 Data re-use: licensing

Free and open research data sharing, which is the ultimate goal of archiving data (software and codes included), increases data re-use. In this context, licensing plays an essential role in sharing data effectively, preventing misuse of data and ensuring that the data creators' rights are protected, while promoting a culture of collaboration. In line with RCN, the recommendation is to use standard Creative Common (CC)²⁵ license.

6 Publication

The last phase of a data lifecycle usually involves scientific publications, where research findings must be freely released to researchers, businesses, the public sector and to the wider society. This is in accordance with the RCN rules: “In line with Plan S²⁶, the Research Council requires full and immediate open access to all scientific publications from the projects we fund. The articles should be made available without delay (embargo) and with an open license allowing reuse of the publication”²⁷. iC3 additionally intends to apply the statement to results under the categories reports and presentations.

6.1 Documental repositories

To achieve the commitment, and in addition to opportunities to publish Open Access offered by national agreements with publishers, two different repositories have been identified where authors are required to upload their results:

- Cristin²⁸ (Current Research Information System In Norway) in combination with local Institutional repositories (Munin²⁹ for UiT, Brage³⁰ for NPI and NORCE) for articles in peer-reviewed journals (independently by the Open Access options Gold or Green), doctoral dissertations and presentations (oral or poster)

²⁵ <https://creativecommons.org/share-your-work/cclicenses/>

²⁶ Plan S: <https://www.coalition-s.org/>

²⁷ <https://www.forskningsradet.no/en/research-policy-strategy/open-science/publications/>

²⁸ <https://app.cristin.no/>

²⁹ <https://munin.uit.no/>

³⁰ <https://sikt.no/tjenester/brage-lokale-vitenarkiv#>

- Septentrio Academic Publishing (the UiT's Institutional service provider for open access publishing) for results under the category reports

Both the repositories enhance findability of iC3 results, assure their accessibility, provide terms of reuse (it does not apply to presentations) and are free of charge. Guidelines (for internal use) have been issued to support authors in post-publication duties.

As for end of 2024, it has been announced that the combination Cristin/Institutional Repositories will be soon replaced by the Norwegian Open Research Archive (NVA)³¹ delivered by Sikt, the Norwegian Agency for Shared Services in Education and Research

There is a strong recommendation that iC3 members use ORCID³² as the primary method for digital researchers' identification. iC3 encourages researchers to register or update their profiles in ORCID, and to make publicly available those that already exist. The recommendation has implications for the core phases: Storage; Archiving; and Publication since ORCID is used by key services selected for data managing purposes at iC3 such as the previously described DataverseNO, NVA and Septentrio. The use of ORCID increases the effectiveness of search engines and is, therefore, beneficial for the visibility of both the Centre and individual researchers.

7 Ethics, codes of conduct, personal data and consent

iC3 recommends associated researchers to become familiar with the principles and good practices outlined by partner organizations at the local level (Principles and guidelines for management of research at UiT)³³, and at national and international levels. In their guidelines and code of conducts, all partner organizations deal with aspects related to ethics in data collecting and handling; the most pertinent for the iC3 community is Guidelines for Research Ethics in Science and Technology issued by The Norwegian National Committee for Research Ethics in Science and Technology³⁴.

Personal data: data collected, generated or re-used (therefore stored or archived) during iC3 research activities do not contain personal data; no measures will be therefore adopted to handling sensitive data. Collection and storage of names, contact details and other personal information (e.g., food allergies) are instead necessary for organizing field trips and scientific cruise expeditions or to maintain research collaborations. In this regard,

³¹ <https://nva.sikt.no/>

³² <https://orcid.org/>

³³ [Principles and guidelines for management of research at UiT](#)

³⁴ <https://www.forskningsetikk.no/en/guidelines/science-and-technology/guidelines-for-research-ethics-in-science-and-technology/>

the legal foundation for managing personal data is based on point (f) of Article 6(1) of the GDPR, which states that processing is essential for the legitimate interests of the controller. These legitimate interests include conducting events efficiently. Information of this kind may be stored until the conclusion of the Centre, then deleted. Processing information about food allergies or other health related information instead refers to point (a) of GDPR article 9 (2) t. Information of this kind must be deleted as soon as no longer necessary. Researchers responsible for field trips, scientific cruise expeditions or other research-related events should assure personal data handling, storage and deleting.

Consent: images and/or film recordings and/or audio recordings showing participants in research activities such as (but not limited to): outreach; group photos; scientific activities; and shooting, must have consent by signature and indicating the dates applicable. For this purpose, a specific form has been issued by the Centre Communications Advisor.

Appendix A: sub-Data Management Plans

The following pages present excerpts of Data Management Plans issued for specific purposes. Updating occurs in the event of new DMPs but not more often than twice a year.

High-performance Computing infrastructure at SIGMA2

Preliminary data plan for iC3 Computing and Storage

(Sarti & Petrini, 2024)

Version	1
Template	Sigma2 Data Management Plan
Last modified date	2024-05-08 12:34:09Z
Last modified by	Fabio Sarti (fabio.sarti@uit.no)
Last checked OK	2024-05-08 12:34:09Z
Editors	Fabio Sarti (fabio.sarti@uit.no)

General Project Information

This section covers general details about your project.

Please provide the name of your project.

iC3 - Centre for ice, Cryosphere, Carbon and Climate - Research Unit 5

Please provide a description of your project

The impact of retreating ice sheets on the Earth's future carbon cycle and ocean ecosystems is a major unknown. Scientists still disagree on whether the polar regions will emit or absorb carbon over the coming decades, and we know very little about how life in the oceans will alter with ice sheet change. We are addressing this major gap in scientific knowledge by pulling together data gathered by iC3 researchers and comprehensively mapping nutrient and carbon fluxes in Greenland and Antarctica. We are modelling the complex interactions between ice, land, ocean, atmosphere and the various organisms they contain. We are assessing whether the retreat of ice sheets and their direct and indirect effects on marine systems will act as accelerator or brake on global warming, and how they will impact the precious ocean ecosystems on which human lives and livelihoods depend. Our aim is to provide national and international initiatives, including the IPCC and IPBES, with a reliable way to assess how the retreat of polar ice sheets will affect Earth's future climate and marine ecosystems.

Which academic subject(s) does your project belong to?

Earth Science/Earth System Science

Please provide the name of the project principal investigator

Fabio Sarti, Data Manager iC3: Centre for ice, Cryosphere, Carbon and Climate - Department of Geosciences – UiT, The Arctic University of Norway

Please provide the funding sources for this project.

This project is core for Research Unit 5 (RU5), but may extend to other research units of iC3. It pertains to addressing the question: What is the sensitivity of Earth's carbon cycle to shrinking ice sheets? at iC3: Centre for ice, Cryosphere, Carbon and Climate - Department of Geosciences – UiT, The Arctic University of Norway. Information about the Centre itself are available at Research Council of Norway Project Bank:

<https://prosjektbanken.forskningsradet.no/en/project/FORISS/332635?Kilde=FORISS&distribution=Ar&chart=bar&calcType=funding&Sprak=no&sort=Matematikk+og+naturvitenskap&source=FORISS&projectId=159137>

Who will be the Data Officer for your project?

Fabio Sarti, Data Manager iC3: Centre for ice, Cryosphere, Carbon and Climate - Department of Geosciences – UiT, The Arctic University of Norway

Does your project have the appropriate resources for the management of your data?

iC3: Centre for ice, Cryosphere, Carbon and Climate benefits of a 50% Data Manager position. Data management costs for this project (infrastructures, personnel) are therefore allocated in the Centre budget and ultimately covered by the funder.

Data

This section covers the data that your project will create or use.

Please describe how your project will create and/or reuse data.

Input data is essentially open-source data available in SIGMA2 shared clusters. The project will create data as the output from numerical models, specifically coupled Earth System Models/Ice Sheet, ocean and bio-geochemical models. Model output will be re-used mostly through postprocessing (ncv and cdo operators, shell, NCL, and python scripts) or as boundary conditions/restart files for other simulations.

Please describe how you will manage the intellectual property rights and ownership of your data.

The project partners hold intellectual property rights to the data and this list will be maintained and provided to potential users of the data. The project will ensure that published data are free from restrictions and abide by the CC By 4.0 or other international licenses for open access to data. Existing observational data (possibly complemented with new datasets collect by the project partners) will be used in the project as model input and for output validation. Appropriate permissions to access and use the data will be secured and documented, and correct attribution will be given when the findings and data are disseminated. The data will be incorporated into the project's analysis processes and will abide by the access constraints placed on the data.

Describe how you will ensure compliance with legislation and institutional regulation?

The data produced by the project will not result in data that may threaten personal or national security if accessed and access to the data produced by this project will not be in conflict with judicial regulations. Data is created as a result of a consortium effort and open and free access is the norm in the climate modelling community and in iC3. Therefore, we consider compliance with institutional regulations not an issue.

Please describe any ethical issues that may affect your data.

No issues anticipated

Documentation and Metadata

This section covers the information that will help you and your colleagues and other researchers to find and reuse your data.

What metadata and documentation do you plan to provide with your data?

The data will mainly be stored in netcdf format with metadata describing the origin, model and experiment following community standards (CMOR)

What data quality measures will you use for your data?

The datasets are a) compared against other models and observations (including those produced within other Research Units in iC3) and b) quality-checked

Storage

This section covers how you will store your data.

Where will you store your data?

NIRD

Please provide your NIRD project ID

NS11022K; Application granted, and resources allocated

Which NIRD services do you intend to use (please select all that apply)?

Computing Resources, Data Storage, easyDMP - Data Planning, NIRD Toolkit and Research Data Archive.

How much data do you plan to store each year from 2020-2024?

Storage forecast:

- 2024: 10 TiB, backup = 0%
- 2025: 20 TiB, backup \leq 25%
- 2026: 30 TiB, backup \leq 75%
- 2027: 35 TiB, backup \leq 75%

- 2028: 0 TiB, backup = 0%

What will you primarily use the storage for?

Computing (including HPC) input and output, Sharing data and Backup.

How much data do you intend to transfer to/from the Computing platform?

More than 50TB.

Please briefly describe how you will ensure the safety of your data.

Except during the first year, which will be dedicated to set up, refine and test the infrastructure, an increasing amount of data will be backed up at NIRD in the years to come. Specifically, output will be quality-checked, and only key variables and output (e.g., CMIP6 standard variables, restart and forcing files) files will be backed up. Access to data will be limited to up to 5 users for the entire lifespan of this project. The PI (Data manager) will grant permission and keep track of users' information.

End of project

This section covers the end of your project where your findings on your data have been published.

Do you plan to make some/all of your data available to others?

Yes, all/some

How will you make your data accessible?

NIRD Archive and Other (please specify in More Information below)

Note: Smaller data sets related to publications may be deposited at other repositories like: Open Research Data (<https://dataverse.no/dataverse/uit>)

Zenodo (<https://zenodo.org>)

Describe how you will select the data for reuse and ensure it can be reused?

Criteria to identify data for reusing are described at the following step 5.1.3 Licensing: most of the output will be licensed under the Creative Commons Attribution Non- Commercial 4.0 International license Documentation: data will be generated following the coupled Model Intercomparison project (CMIP) protocol. Additional read-me files will integrate the set of information necessary for re-use.

When will your data be available for reuse?

As soon as the data has been analysed.

Note: Model forcing files and restart files will be made available for re-use immediately after simulations are analyzed. Model output will be made re-usable to partner and collaborators immediately after the analysis, and portions of the data will be shared publicly after publication.

Apart from a possible embargo period, are there any other restrictions on the reuse of your data?

The scientific output resulting from interpreting the data have to be published in scientific articles before the data becomes available.

Please provide any additional information you think is relevant to your plan.

No additional relevant information provided

Appendix 1: iC3 expeditions

The following pages present information about iC3-funded expeditions (from the most recent) with particular focus on datasets. Updating occurs twice a year.

Expedition iC3-24-02

iC3 name (naming convention)	iC3-24-02-GoNorth
Expedition Name	GoNorth24 - KH24-259
Chief Scientist (or PI)	Risebrobakken, Bjørg/Laberg, Jan Sverre
Expedition Type	Cruise
Research Vessel	RV Kronprins Haakon
Year	2024
Date From/To	29.08-19.09
Region/Area	Independence Fjord, Hagen Fjord; Greenland
Departure	Longyearbyen, Norway
Arrival	Longyearbyen, Norway
Note	GoNorth 2024 (https://gonortharctic.no/)
Report	

DATA COLLECTED/GENERATED

Water sampling (CTD) -> biogeochemistry

Sediment sampling (GC) -> pore water and gas

References

Sarti, F., & Petrini, M. (2024). *High-performance Computing infrastructure at SIGMA2. Preliminary data plan for iC3 Computing and Storage*. Deposited at Sigma2