



SIOS

SVALBARD INTEGRATED ARCTIC
EARTH OBSERVING SYSTEM

Technical documentation

SIOS Data Management Plan



Versions

| Version | Date | Comment | Responsible |
|----------------|-------------|--|-----------------------------|
| 0.3 | 2017-04-19 | Minor modifications following internal review. | Kim Holmen Øystein Godøy |
| 0.2 | 2017-04-18 | Correction of typos. | Øystein Godøy |
| 0.1 | 2017-03-03 | Draft for discussion | Øystein Godøy |

Table of Contents

| | | |
|-----|---|----|
| 1 | Admin details | 5 |
| 2 | Data summary | 5 |
| 2.1 | Making data findable, including provisions for metadata [fair data] | 6 |
| 2.2 | Making data openly accessible [fair data] | 7 |
| 2.3 | Making data interoperable [fair data] | 7 |
| 2.4 | Increase data re-use (through clarifying licenses) [fair data] | 8 |
| 3 | Allocation of resources | 8 |
| 4 | Data security | 10 |
| 5 | Ethical aspects | 10 |
| 6 | Other | 10 |

1 Admin details

| | |
|---------------------|--|
| Project Name | Svalbard Integrated Arctic Earth Observing System |
| Funding | Described in the SIOS Statutes. Core activities are covered by membership fees and in-kind contributions, additional activities are supported by external funding through joint proposals. |
| Partners | Alfred-Wegener-Institute (DE), British Antarctic Survey (UK), National Research Centre of Italy (IT), Finnish Meteorological Institute (FI), Institute of Geophysics - Polish Academy of Sciences (PL), Institut Polaire Francais (FR), Norwegian Meteorological Institute (NO), Nansen Environmental and Remote Sensing Center (NO), National Institute of Polar Research (JP), Norwegian Polar Institute (NO), Polar Research Institute of China (CN), Stockholm University (SE), The University Centre in Svalbard (NO), University of Helsinki (FI). |

2 Data summary

SIOS is a regional observing system for long-term measurements in and around Svalbard addressing Earth System Science questions. SIOS integrates the existing distributed observational infrastructure and generates added value for all partners beyond what their individual capacities can provide.

SIOS brings observations together into a coherent and integrated observational programme that will be sustained over a long period. Within SIOS, researchers can cooperate to access instruments, acquire data and address questions that would not be practical or cost effective for a single institution or nation alone.

By bringing many types of observations together and asking questions about how these are influenced by each other SIOS generates new insights about the Svalbard region's role in the Earth system. Thus, SIOS offers unique opportunities for research and the long-term acquisition of fundamental knowledge about global environmental change.

SIOS focuses on processes and their interactions between the different spheres, i.e. biosphere, geosphere, atmosphere, cryosphere and hydrosphere. The core observational programme of SIOS provides the research community with systematic observations that are sustained over time, yet dynamic enough to be adapted as new methods and questions from society appear.

SIOS sets an example for how to systematically construct observational networks in the Arctic and how joint efforts provide added value to the user community. The services benefit the international polar research community as a whole and make SIOS a unique international meeting place for developing the science of long-term monitoring in Polar Regions.

SIOS is based on existing data collection, gap analyses and mitigation of gaps identified from a scientific perspective. This implies that legacy data comes in a wide range of data formats and types. Part of the work is to harmonise the data collection, but this has to be done in a stepwise manner. Initially, the main focus is on harmonising new data and some core datasets.

The existing datasets collected by the partners at the infrastructure locations described by the

Research Infrastructure overview and by the instruments identified is already handled by existing data management systems operated by the partners. The SIOS Data Management System connects these individual systems to a unified virtual data management system.

The main part of the data comes from permanent instrumentation located at SIOS Infrastructure locations in Ny-Ålesund (several research stations), Longyearbyen, Hornsund as well as a number of locations scattered around the Svalbard islands (e.g. meteorological stations) and in the surrounding ocean (e.g. Fram and Hausgarten observatories and regular sections or profiles).

The primary data centres to interface initially include:

- Alfred Wegener Institute/Pangaea (DE)
- Italian Arctic Data Center/CNR (IT)
- Institute of Marine Research (NO)
- National Institute of Polar Research/ADS (JP)
- Norwegian Meteorological Institute
- Norwegian Polar Institute
- Polar Research Institute of China

The total amount of data is yet not known as the Scientific Optimisation Advisory Group is identifying the core datasets.

The data is vital to the understanding of Earth System processes in and around Svalbard.

2.1 Making data findable, including provisions for metadata [fair data]

The SIOS Data Management System relies on documentation of datasets through Discovery metadata (technical guidelines under development). These metadata are maintained by the data centres contributing to SIOS. The central node of SIOS is harvesting these metadata on a regular basis and ingests them in a central catalogue which is used to support users through the SIOS Web Portal.

SIOS promotes the implementation of Persistent Identifiers at each contributing data centre. Some have this in place, while others are in the process of establishing this. SIOS is not creating this centrally.

The diversity of data handled by SIOS implies that a full overview is not possible in the current situation. However, SIOS is promoting utilisation of standards. One example is encoding of geophysical data using NetCDF following the Climate and Forecast convention. There are however issues with many of the standards and these issues have to be addressed by the SIOS Data Management System Working Group as the system evolves.

The search model used is based on GCMD Science Keywords for parameter mapping. Not all data centres use these. Where possible, the central node will map between local vocabularies used by the data centres and the GCMD keywords until a more uniform annotation of data is

possible. That will however take time as large legacy systems are integrated in SIOS.

Versioning of data is the responsibility of the contributing data centres and SIOS formulates requirements in the technical guidelines. Some data centres have this in place, while others are working on this. Full implementation, will as other issues take time and depend on the resources available.

The central node can consume GCMD DIF and ISO19115 records (using GCMD keywords). Support for more formats is considered, but depends on resources.

2.2 Making data openly accessible [fair data]

All metadata will be available through a search interface integrated in the SIOS web portal. Some data may have access restrictions (as the data policy also opens for). These will be handled accordingly by the responsible data centre. In the long perspective Single Sign On is planned, but this is not feasible within the resources available initially.

Data are made available by the responsible data centre. No data is handled centrally in the first version, only services towards data. Initially a wide range of access solutions and formats will be used. The SIOS Data Management Working Group is however developing technical guidelines addressing this issue as well. Implementation will have to be stepwise and at different speeds for the different data centres.

Methods and software for accessing data is under development. However, the intention is to use internationally accepted open standards (as mentioned above) which have a wide range of open software tools available. The central system for dataset discovery is developed under a GNU General Public License version 2 or higher. Some parts are published under GNU Lesser General Public License version 2.1 or higher.

Metadata and data for the datasets are maintained by the responsible data centres, metadata supporting unified search is harvested and ingested in the central node. A dedicated GitHub area has been set up for sharing software, but much of the software is already open source and in well maintained repositories. The addresses of these repositories will be collected and published within the infrastructure Intranet (including the GitHub repository).

Datasets with restrictions are initially handled by the responsible data centre. Generally the metadata will be searchable and contain information on how to request access to the dataset.

2.3 Making data interoperable [fair data]

The current situation for the legacy datasets of SIOS is that there is no common level of interoperability at the data level. Some data centres support or have started to implement OPeNDAP and use metadata following the Climate and Forecast conventions.

At the metadata level, interoperability is better as many of the data centres do support GCMD DIF with the appropriate controlled vocabularies.

The interoperability aspect of SIOS will evolve and will require substantial support through dedicated resources.

A system for semantic translation of annotated discovery metadata content is under development, but many of the controlled vocabularies are not available in machine readable form, neither have they been compared with other vocabularies (whether within or external to the discipline in question). This is work in progress. Relevant efforts of Research Data Alliance, ENVRIPplus and WMO are followed.

Initially GCMD Science keywords will be used, but given the open data space of SIOS, this has to be evaluated continuously. Currently mapping between GCMD and CF keywords is supported (but needs to be updated), mapping from e.g. the Canadian Polar Data Catalogue keywords are also supported.

2.4 Increase data re-use (through clarifying licenses) [fair data]

The data policy promotes free and open data sharing. However, while the data policy outlines the general rules of sharing data, a license describes the specific requirements for a dataset. Within SIOS each dataset will have a license. The SIOS recommendation is to use [Creative Commons](#) attribution for data.

According to the current version of the Data Policy:

SIOS data delivered in a timely manner means delivery of data without un-due delay. Any delay, due or un-due, shall not be longer than one year. Discovery metadata shall be delivered immediately.

Referring to the SIOS Data Policy, SIOS is promoting free and open access to data. Some data may have constraints and may be available to members only initially. If SIOS is to be terminated, data are still maintained by the contributing data centres and availability of data depends on the resources and priorities of those data centres.

Initially SIOS relies on data quality assurance processes implemented at each contributing data centre and by the Principal Investigators involved. Concerning the scientific quality of SIOS output, this depends on the data used and is monitored by the Science Optimisation and Advisory Group. The Science Optimisation Advisory Group will recommend actions or evaluations to be performed by the Data management working group.

As long as SIOS operates the intention is to make observations available at any time, for simulations or analysed products, the norm will be 10 years, but subject to the decision of the Science Optimisation Advisory Group.

3 Allocation of resources

In the current situation as the SIOS infrastructure is being developed, there is no overview of the total costs for making SIOS data FAIR. SIOS relies on existing infrastructures which in most cases already have FAIR elements implemented at the metadata level. At the dataset level very few data centres comply with FAIR principles. Activities within the SIOS Data Management System Working Group in 2017 are addressing the cost issue and will deliver a detailed work plan and budget for 2018.

The following data centres are involved, and respective members of the SIOS Data Management System Working Group are listed as well.

| Data centre | N ¹ | URL | Contact | Comment |
|-----------------------------------|----------------|---|-----------------|---|
| Arctic Data Archive System | JP | https://ads.nipr.ac.jp/ | Hironori Yabuki | Subsystem under development, Metadata interoperability interfaces are under development. |
| Arctic Data Centre | NO | http://arcticdata.met.no/ | Øystein Godøy | Subsystem integrated in demonstrator. Metadata interoperability interfaces are available. This subsystem is a WMO Information System Data Collection and Production Center. |
| AWI/PANGAEA | DE | http://pangaea.de/ | Angela Schafer | Subsystem integrated in demonstrator. Metadata interoperability interfaces are available. This subsystem is a member of ICSU World Data System. |
| British Antarctic Survey | GB | https://www.bas.ac.uk/data/ | Helen Peat | Subsystem can easily be integrated. Metadata interoperability interfaces are available. |
| Italian Arctic Data Center (IADC) | IT | http://arcticnode.dta.cnr.it/cnr/index.php | Angelo Viola | Subsystem can easily be integrated. Metadata interoperability interfaces are available, although not necessarily activated yet. |
| NILU | NO | http://ebas.nilu.no/ | Markus Fiebig | Subsystem is missing standardised metadata interoperability interfaces. These need to be implemented. |
| Norwegian Marine Data Centre | NO | http://www.nmdc.no/ | Terry Hannant | Subsystem is currently under development, but can be integrated in demonstrator when interfaces are implemented. This subsystem is national infrastructure in Norway. |

¹ Nation.

| Data centre | N | URL | Contact | Comment |
|---|----|---|----------------|--|
| Norwegian Polar Institute | NO | http://data.npolar.no/ | Stein Tronstad | Subsystem integrated in demonstrator. Metadata interoperability interfaces are available. |
| Norwegian Satellite Earth Observation Database for Marine and Polar Research (NORMAP) | NO | http://normap.nersc.no/ | Torill Hamre | Subsystem is currently under development, but metadata interoperability interfaces are available. This subsystem is national infrastructure in Norway. |

In the current situation SIOS has no overview of the costs of long term preservation of data. As mentioned, SIOS is built upon existing data streams and data centres. Thus these data are not only used for SIOS and would be preserved anyway.

Concerning the value of data preservation, observations are priceless for understanding our environment.

4 Data security

Data security relies on the existing mechanisms of the contributing data centres. Currently SIOS has no recommendations in this context, but such recommendations are likely to be made later.

The central node relies on secure HTTP traffic, but not all contributing data centres support this yet. This is expected to evolve during implementation.

5 Ethical aspects

Ethical aspects are handled in the SIOS Data Policy.

6 Other

Currently not applicable.