SIOS

SVALBARD INTEGRATED ARCTIC EARTH OBSERVING SYSTEM

Data Management Plan

Versions

Version	Date	Comment	Responsible
0.5	2022-01-10	Updated data centres contributing (not finished yet) and some minor editing.	Øystein Godøy
0.4	2020-06-17	Transferred into ASCIIDOC and updated contributing data centres.	Øystein Godøy
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

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 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.

3. FAIR data

3.1. Making data findable, including provisions for metadata

The SIOS Data Management System (DSMS) relies on documentation of datasets through Discovery metadata (requirements are outlined in the SDMS technical manuals). These discovery 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. The information model used by SIOS in the central catalogue is the MET Norway Metadata Format Specification which is compliant with ISO19115 and GCMD DIF, and further extends these in certain areas.

NOTE	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 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.

NOTE	Mapping of vocabularies are still work in progress.

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 of SIOS consumes discovery metadata exposed as GCMD DIF and ISO-19115 records (using GCMD keywords). Support for DCAT and schema.org is in progress. In addition SDMS can generated discovery metadata directly from NetCDF-CF files that are exposed on OPeNDAP, preferably using THREDDS catalogues. These data however, have to contain global attributes according to the Attribute Convention for Dataset Discovery (ACDD). The elements required by SDMS are indicated in https://adc.met.no/node/4.

The catalogue is exposed through a human interface which is embedded in the SIOS website and through a

machine actionable interface which is available at https://sios.csw.met.no.

WARNING	The machine interface is still under development and may lag behind the information found in the human interface.

3.2. Making data openly accessible

Data are made available by data centres contributing to the SDMS. No data is handled centrally in the first version, only services towards data. Initially a wide range of access solutions and formats will be supported, but for specific datasets like the SIOS Core Data (SCD), harmonised encoding and access mechanisms are required[1]. 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 described in the SIOS Interoperability Guidelines. These relies on internationally accepted open standards 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.

The SIOS Data Policy promotes free and open data. As a general rule, discovery metadata are only provided where actual data are available online, although exceptions may occur.

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.

NOTE	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.

3.3. Making data interoperable

The current situation for the legacy datasets of SIOS is that there is no common level of interoperability at the data level. The primary focus of SIOS is to establish interoperability at the discovery metadata level. This is in good progress and the results is viewed in the Data Access Point embedded in the SIOS website. At the data level, interoperability is in progress for SCD which are prioritised. The technical requirements for both are outlined in the SDMS Interoperability guidelines. The reason why SCD are prioritised for this is that proper interoperability is required to ensure seamless integration of SCD data from various providers. This in support of Earth System Science.

NOTE	SDMS is promoting the use of NetCDF following the Climate and Forecast Conventions and Darwin Core Archives for SCD. This ensures standardised structures and semantics.

For the interoperability aspect of SIOS to further evolve will require substantial support through dedicated resources. A very good example of what can be achieved was the SIOS Core data Curation call of 2021 that focused on increased harmonisation of the way SCD are published.

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, ENVRIPlus and WMO are followed.

3.4. Increase data re-use (through clarifying licenses)

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.

IMPORTANT | SIOS promotes the usage of the Creative Commons Attribution 4.0 license on all data.

4. 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 data centres which often have FAIR elements implemented at the discovery metadata level. At the dataset level very few data centres comply with the FAIR principles.

Activities within the SIOS Data Management System (SDMS) are coordinated through the SDMS Working Group which every year plans for the work to be undertaken and the resources needed in a work plan that is submitted to the SIOS Board of Directors and finally approved by the General Assembly. The SDMS WG has members from most of the partner data centres contributing and reports to the SIOS Board of Directors and eventually the General Assembly. The SDMS WG has approximately 20 members[2], not all are active.

The SDMS WG is chaired by Stein Tronstad and Marcin Wichorowski, the SIOS Data Manager acts as secretary for the group and the SIOS Data and Integration Officer acts as liaison with the scientific community, in addition to the relations already established internally by partners.

The partner data centres involved in development of SDMS, and their respective members of the SIOS Data Management System Working Group are listed below.

Name	Country	URL	Responsible	Comment
Arctic Data Archive System	JP	https://ads.nipr.ac.jp/	Hironori Yabuki	Discovery metadata interoperability interfaces are available. This data centre is regularly harvested.

Name	Country	URL	Responsible	Comment
Arctic Data Centre	NO	http://adc.met.no/	Øystein Godøy	Discovery metadata interoperability interfaces are available. Data are served according to the Interoperability Guidelines. This subsystem is connected to the WMO Information System and the data centre is regularly harvested.
AWI/PANGAEA	DE	http://pangaea.de/	NA	Discovery metadata interoperability interfaces are available. This subsystem is a member of ICSU World Data System and is regularly harvested.
IGPAS	PL		Jakub Przewalski	Discovery metadata are embedded in the datasets exposed, discover metadata interfaces are under implementation. Interfaces are currently undergoing testing.
IOPAN	PL	https://geo1.iopan.pl	Marcin Wichorowski	Discovery metadata interoperability interfaces are available, integration of data sources according to the Interoperability Guidelines is in progress. The data centre is regularly harvested, but ingestion is still under testing.

Name	Country	URL	Responsible	Comment
Italian Arctic Data Center (IADC)	IT	https://metadata.iadc.cnr.i	Angelo Viola Giulio Verazzo	Discovery metadata interoperability interfaces are available, integration of data sources according to the Interoperability Guidelines in in progress. The data centre is regularly harvested, but ingestion is still under testing.
NILU	NO	http://ebas.nilu.no/	Markus Fiebig	Discovery metadata interoperability interfaces are available. Data are served according to the Interoperability Guidelines. This subsystem is connected to the WMO Information System and is regularly harvested.
Norwegian Marine Data Centre	NO	http://www.nmdc.no/	Arnfinn Morvik	Discovery metadata interfaces are available. This subsystem is a national e-infrastructure in Norway and is regularly harvested.
Norwegian Polar Institute	NO	http://data.npolar.no/	Stein Tronstad	Discovery metadata interoperability interfaces are available. This data centre is regularly harvested.

Name	Country	URL	Responsible	Comment
Nansen Environmental and Remote Sensing Center	NO	http://metadata.nersc.no/	Torill Hamre	Discovery metadata interoperability interfaces are available and data are served according to the interoperability guidelines. This data centre is regularly harvested.
University of Silesia	PL	http://ppdb.us.edu.pl/	Łukasz Małarzweski	Discovery metadata interoperability interfaces are available, integration of data sources according to the Interoperability Guidelines is in progress. This data centre is regularly harvested.

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, and once lost, cannot be regenerated.

5. Data security

Most of the data generated by the SIOS related infrastructure are open. SIOS is working to establish secure connections between data centres and data consumers to ensure that correct decisions can be made using data. However, data from third parties will also be made available, for these data there is limited room for SIOS to ensure integrity and security of data.

NOTE SIOS promotes usage of secure communication at all end points connected to SDMS. The central node utilises secure HTTP, but not all contributing data centres support this yet. This is expected to evolve during the implementation.

Initially SIOS relies on the 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.

In order to further improve the integrity of data served, the information model used for discovery metadata has room for checksums on datasets, however there is no universal approach to conveying this as it is today and for data served through web services it doesn't help.

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

6. Ethical aspects

Ethical aspects are handled according to the SIOS Data Policy. On a general basis SDMS is primarily handling non sensitive data, but SIOS follows the principle of "as open as possible, as closed as necessary".

7. Other

None known yet.

---

[1] The SCD specifications are still under development.

[2] The number of members fluctuates.