

SIOS Core Data (SCD)



Ilkka Matero¹, Dariusz Ignatiuk¹, Øystein Godøy^{1,2}, Bartłomiej Luks³, Roland Neuber⁴, Vito Vitale⁵, Shridhar Jawak¹, Christiane Hübner², Inger Jennings¹, Heikki Lihavainen¹

1 SIOS Knowledge Centre, Longyearbyen, Norway

2 Norwegian Meteorological Institute, Oslo, Norway

3 Institute of Geophysics, Polish Academy of Sciences, Warsaw, Poland

4 Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Potsdam, Germany

5 Institute of Polar Sciences, National Research Council of Italy, Bologna, Italy

Corresponding author: Ilkka Matero, Ilkka.matero@sios-svalbard.org

ORCID number: 0000-0002-2333-9010

Keywords: SIOS core data, SIOS Data Management System, FAIR data

DOI: <https://doi.org/10.5281/zenodo.5751970>

1. Introduction

Svalbard Integrated Arctic Earth Observing System (SIOS) is a regional observing system for long-term measurements in and around Svalbard addressing Earth System Science (ESS) questions. SIOS integrates the existing distributed observational infrastructure to generate added value for all partner organisations beyond what their individual capacities can provide. SIOS brings together observations in a coherent and integrated observational programme that will be sustained over a long period. Only in this way can the inherent coupling processes in this regional-scale Arctic system and its connections with the Earth System at large be addressed adequately. In brief, SIOS facilitates addressing key scientific questions in ESS with a sustained measurement programme.

The SIOS Data Management Service (SDMS)¹ is one of the key services provided by SIOS. SDMS enables data submission, discovery, access, use and preservation of SIOS-relevant datasets and metadata across contributing data centres (operated by partners) and relies on the principles of distributed data management. The distributed

data management approach means that the contributing data centres host all the datasets, and SDMS provides access to them through regularly harvested access metadata. In addition, information is harvested from third party data centres that have information of relevance for the SIOS community. Among the main products and services of SIOS are data useful for addressing identified key research questions in Earth System Science. A subset of the data, the SIOS Core Data (SCD), comprises the data needed to scientifically assess the state of the environment in Svalbard.

The core observational programme carried out by SIOS partners provides the research community with systematic long-term observations and allows for integration of upcoming new methodologies and techniques, as well as research questions. The SIOS observational programme – and hence also the core data – are continuously monitored and if necessary updated based on the recommendations from annual SESS reports, the SIOS infrastructure optimisation report and interaction with the calibration/validation and modelling communities.

2. SIOS data

The scientific themes that guide the observations and the optimisation of the observing system of SIOS are:

- Energy and mass exchange;
- Combined effects of human perturbations;
- Effects of Global Environmental Change on organisms, populations, and ecosystems.

To address these themes, a comprehensive dataset of state variables allowing the diagnostics for Global Environmental Change is mandatory. Data products addressing the identified key research questions in ESS are a core product and making them accessible

is a central service of SIOS. The geographical extent of the SIOS region is limited, and thus ESS monitoring and observation activities need to be limited to regionally accessible and relevant variables which can be expected to change over timescales of years to decades. There is another aspect of data from the SIOS perspective: SIOS members must follow the SIOS data policy², which adheres closely to the FAIR guiding principles. A central element of this data policy is that all data are open and accessible for everyone. SIOS data also include higher-order products (level 3+, as described by NASA³).

1 <https://sios-svalbard.org/Data>

2 https://sios-svalbard.org/sites/sios-svalbard.org/files/common/SIOS_Data_Policy.pdf

3 <https://earthdata.nasa.gov/collaborate/open-data-services-and-software/data-information-policy/data-levels>

SIOS data are defined as all the ESS data available in the SIOS Data Access Portal⁴, including, e.g., solitary measurements and third-party data, see Table 1. SIOS data must be accompanied by

discovery metadata, making them findable and accessible. Where possible, predefined observation protocols (e.g. from ACTRIS⁵, GAW⁶, WMO⁷ etc.) should be included.

Table 1. Data definition in SIOS

Type of SIOS data	Policy	Category	Comment
Single field experiments	Transparency (data have been collected)		SIOS data if available in the SIOS Data Access Portal
Historic data records, including long-term measurements, but discontinued	Transparency (data exist). Submitted at owner's discretion		SIOS data if available in the SIOS Data Access Portal
SIOS Access call, SESS report, and other higher-level data	Available after the project		SIOS data if available in the SIOS Data Access Portal
Long-term observations	Transparency (data exist). Submitted at owner's discretion		SIOS data if available in the SIOS Data Access Portal
SIOS Core Data (SCD)	Transparency (data exist and available online)	Core	Fulfils SIOS criteria of scientific requirements, data availability and >5 years collecting commitment (see section 2.1.1-3 in this chapter). Available in the SIOS Data Access Portal
SIOS Core Data - Candidate (SCD-C)	Will become available online within one year	Core data candidate	Fulfils the same criteria as core data but data is not yet available online

2.1. SIOS core data

The SIOS core data⁸ comprises the data needed to scientifically assess the state of the environment in Svalbard. The first set of SCD variables is based on the Essential Climate Variables (ECV) as defined by the Global Climate Observing System (GCOS)⁹. The Svalbard-relevant variables were selected according to recommendations from SIOS' Scientific Optimisation Advisory Group (SOAG). In addition, SIOS-KC conducted a series of interviews with scientists in modelling and remote sensing calibration and validation (cal/val) communities to identify the most relevant variables for the first set of SCD.

SIOS core data are a dynamic set of key variables. Researchers can suggest new variables through

SESS reports or even by directly addressing SOAG, which would evaluate whether the variables would qualify as SCD candidates. The core data are provided by SIOS members, who commit to provide them on a regular and long-term (more than five years) basis. All SCD follow predefined file formats and are properly associated with appropriate metadata.

The three criteria described in the following subsections must be met for data to qualify as SCD, as defined by a task force organised by SOAG. The criteria are based on standards of scientific excellence in the Earth Science System, in SIOS' 'legal' framework and SIOS data policy. In 2020, the first set of SCD were defined to optimise the resources contributed by the SIOS research community. An updated list of variables for SCD

⁴ <https://sios-svalbard.org/metsis/search>

⁵ The Aerosol, Clouds and Trace Gases Research Infrastructure, <https://www.actris.eu/about>

⁶ Global Atmosphere Watch Programme, <https://community.wmo.int/activity-areas/gaw>

⁷ World Meteorological Organization, <https://public.wmo.int/en>

⁸ https://sios-svalbard.org/sites/sios-svalbard.org/files/common/CoreData_Documentation.pdf

⁹ <https://gcos.wmo.int/en/home>

has been prepared in 2021 and is waiting for approval and new mapping by SIOS partners. The SCD definition process promotes compliance with FAIR guiding principles¹⁰ (see figure in summary) for scientific data management and stewardship for key datasets.

SIOS Core Data Candidates (SCD-C) are data that fulfil the criteria outlined above but are not yet available online. Contributing members have committed to making the SCD-C available through the SDMS within a year of their qualification as SCD-C. SIOS provides support for its member institutions in transforming SCD-C to SCD.

2.1.1. Scientific requirements

To qualify as SCD, the variable must be critical to answer the key research questions as defined in the SIOS infrastructure optimisation report, and further updates in SESS reports. The requirement for temporal and spatial coverage varies between variables, and potentially depending on the scientific question. This should be considered in the scientific requirement (optimisation). Connection with GCOS ECVs and other Essential variable schemes, as for example Essential Ocean Variables, and marine Essential Biodiversity Variables, can provide guidelines and criteria for selection and prioritisation.

2.1.2. Data availability

SCD must be available through SDMS and accessible in the SIOS Data Access Portal. SCD candidates should be made available as soon as possible and at latest one year after data collection.

Data must be described with sufficient discovery and use metadata, which include the necessary information for finding the datasets of interest and the formatting of the data, respectively. Where possible, existing measurement and calibration protocols (e.g. WMO, GAW, BSRN¹¹, ACTRIS and ICOS¹²) should be used to collect SCD to secure comparability. Instrument inter-comparisons are highly recommended. SIOS will work on promoting data exchange and observation protocol harmonisation, and sustain and facilitate intercomparison campaigns.

2.1.3. Members' commitment

For SCD, there must be a commitment from the data-providing institute to maintain the measurement infrastructure setup and data production for at least 5 years, as well as making the data available through SDMS. Even though SIOS membership is based on a non-binding agreement, there are strong incentives for members to sustain their credibility in the system. There are cases in which the data delivery fails regardless of the members' commitment to delivery, for example in case of failures of key instruments that are mounted on weather stations or satellites. In these cases, it is expected that the data providers do their best within their means to ensure continuation of the measurements.

An example of a commitment in the SIOS 'legal' framework could be a letter indicating intentions of the member about data offered to SIOS, a period of validity of the offer, and at least a tentative plan for frequency of delivery (following rules fixed in criterion 2).

¹⁰ <https://doi.org/10.1038/sdata.2016.18>

¹¹ Baseline Surface Radiation Network, <https://bsrn.awi.de/>

¹² Integrated Carbon Observation System, <https://www.icos-cp.eu/>

3. The state of SIOS core data

The state of the SCD is mapped regularly through reviewing the SCD variable coverage and number of datasets available through the Data Access Portal. As part of this, the mapping effort also provides a review into the development of SIOS in general. The most recent full mapping was completed in November 2020 and has since been updated twice, in December 2020 and May 2021, with the mapping results available online¹³. Out of the 26 member institutions, 21 have indicated that they can deliver SCD and SCD-C datasets. Eighteen of these 21 have committed to delivery of SCD (up from 16 member institutions committed to delivery of SCD over last year). Note that not all SIOS members conduct ongoing monitoring in or around Svalbard, which obviously affects the number of institutions that can provide SCD.

SCD have been divided into 4 categories, between which there is no balance in the number of variables: ATMOSPHERE (30 variables), CRYOSPHERE (11 variables), TERRESTRIAL (1 variable), OCEANS (9 variables). The most obvious gap is in the TERRESTRIAL category and is due to the collaboration with COAT under the SIOS INFRANOR project¹⁴. When the COAT database is integrated into the SDMS and SIOS receives

information about key monitoring activities, this gap will be filled.

Currently 51 SCD variables¹⁵ have been identified in collaboration by SOAG, the Research Infrastructure Coordination Committee and additional scientific experts. Another 21 variables pending approval by SOAG were identified in the SESS reports and SCD workshops. Of the 51 SCD variables, 29 currently have datasets available through SDMS. The currently available SCD datasets are based on what has been made available and continuous delivery of datasets has been committed to. The list of variables is dynamic, and regularly reviewed and added to as part of SIOS development. It should however be noted that data being available for a SCD variable does not necessarily mean that the variable can be considered thoroughly covered in time and space, but rather that at least one dataset of this type exists; the temporal and spatial resolution varies between datasets. Both the coverage and number of SCD variables having data available are expected to increase over the next year, as there are currently 37 SCD-C datasets. These 37 SCD-candidates include data on 11 SCD variables that are currently not available through SDMS.

4. Core Data development efforts

Several recurring and ongoing efforts in SIOS facilitate and support the development and refinement of SCD availability. The most relevant actions are briefly described in this section.

An SCD seminar was organised in 2020 during the Polar Night Week in Longyearbyen. The workshop aimed to familiarise the community with the idea of SCD and present the results of work from SOAG and the task force on the definition of SCD and measurement protocols. In addition, a list of variables proposed as SCD was presented.

Discussion and work on SCD with groups interested in sharing and harmonising data as SCD continued at subsequent SCD workshops in November 2020 and January 2021. During the first workshop, the SCD identification and mapping process and the results were presented. The discussion and work covered issues related to harmonisation and standardisation of measurement protocols. The subsequent workshops continued on the topics already undertaken and discussed the list of new variables that should be added to the SCD list.

¹³ <https://sios-svalbard.org/CoreData>

¹⁴ <https://sios-svalbard.org/InfraNor>

¹⁵ https://sios-svalbard.org/sites/sios-svalbard.org/files/common/CoreData_Appendix_v1.pdf

SIOS facilitated two one-off efforts called ‘SCD curation calls’ in 2020 and 2021. These calls allowed for member institutions to apply for SIOS support for transformation of SCD-C to FAIR data. Eight such projects were carried out by member institutions. The project deliverables are undergoing final review during the writing of this report and based on initial assessment the outcomes of the

projects were mostly successful in transforming SCD-C into SCD.

The SDMS Working Group has recently initiated a task force with the aim of supporting harmonisation of SCD. Together with SIOS-KC, the task force will create a plan for harmonisation of the metadata and measurement standards for core data.

5. Recommendations for the future

- Facilitate transformation of SCD-candidates to SCD and verification of previously reported SCD-candidate variables
- Prioritise defining and harmonising measurement protocols and data protocols for SCDs
- Do an annual evaluation of variables on the SCD list to ensure their significance and reusability
- Activate hidden data from multi-year monitoring efforts that are currently not available in any database that meets the FAIR data principles
- Share knowledge, expertise, and experience of the SCD definition process in international projects

6. Acknowledgements

This work was supported by the Research Council of Norway, project number 291644, Svalbard

Integrated Arctic Earth Observing System – Knowledge Centre, operational phase.