
Greenland Ecosystem Monitoring - Data Management Plan

A Data Management Plan created using DMPTool

Title: Greenland Ecosystem Monitoring (GEM).

Creator: Jonas K. Roemer

DOI: <https://doi.org/10.5281/zenodo.6973634>

Funder: Ministry of Environment of Denmark. Danish Ministry of Climate, Energy and Utilities.

Project abstract:

Greenland Ecosystem Monitoring (GEM) is an integrated monitoring and long-term research programme on ecosystems and climate change effects and feedbacks in the Arctic. Since 1995 the programme has established a coherent and integrated understanding of the functioning of ecosystems in a highly variable climate, which is based upon a comprehensive, long-term inter-disciplinary data collection carried out by Danish and Greenlandic monitoring and research institutions. The GEM Programme put around 75 scientists in the field annually to collect data on ecosystem and climate change in Greenland. The database currently covers data from monitoring programmes from Zackenberg (1995-), Kobbefjord at Nuuk (2007-) and Disko (2017-). The well over 2000 parameters are freely available via the GEM Database and used by GEM participants and external scientists to produce scientific papers, scientific assessments, advisory reports, education materials and public outreach.

Last modified: 08-08-2022

Greenland Ecosystem Monitoring (GEM)

Data Collection

What data will you collect or create?

The data collected and shared are at a low granular level: the direct measurements and observations, not reanalysed or aggregated. They are however processed and quality assured to high standards before being shared, to provide a trustworthy data source.

The GEM project works with a structured form of observation data that is uniform across the entire project:

- Observations are collected in datasets. A dataset is created by a monitoring sub-programme. Datasets are organized in categories.
- A dataset has a number of observation rows, making up a timeseries of environmental data.
- An observation row is indexed by its date (and possibly time).
- An observation row can have one or more columns of measured environmental variables.

This very basic form of datasets - rows - columns containing timeseries, are delivered and shared in csv format, which is completely open and can be reused by anyone in their software of choice.

For reasons of long term storage, backup and searchability data are actually internally stored in a SQL Server relational database with dataflows implemented to transform to/from csv files.

How will the data be collected or created?

The data will be collected annually, at each of the GEM sites (currently Disko, Nuuk and Zackenberg), by:

- Scientists and research assistants doing observations and measurements during the fieldwork season.
- Automatic data loggers, drones and other means of obtaining observations.

At each GEM site, cross-disciplinary monitoring is taking place with some or all of a number of cross-disciplinary sub-programmes participating: BioBasis, ClimateBasis, GeoBasis, GlacioBasis, MarineBasis as well as a supporting sub-programme specialized in remote sensing.

For data collection, each sub-programme follows their detailed monitoring manual to ensure that data is collected and handled consistently. The monitoring manuals are available here: <https://g-e-m.dk/gem-publications-and-reports/manuals/>

The GEM sub-programmes are linked to international scientific networks and where possible, they follow accepted best practices and standards.

Documentation and Metadata

What documentation and metadata will accompany the data?

When delivering data to the GEM database, scientists are required to include metadata at two levels:

1. The element description metadata file is high-level metadata about the file / dataset scientists deliver: stating information such as sub-program, element group and ecosystem element and in which filename the data is located.
These columns must be present in the file: Program, ElementGroup, Location, Element, Latitude, Longitude, Description, Filename
2. The column description metadata file is detailed metadata about the columns in all of the data files scientists deliver: stating information such as datatype and detailed description of the data in the column and how it was measured. Metadata entries must exist for all columns in all data files.
These columns must be present in the file: Program, ElementGroup, Location, Element, Columnname, Unit, Datatype, FieldDescription and DatasetDescription

One purpose of the metadata is to give sufficient information for users who download the open data, to understand and use it. Another benefit of the detailed metadata is that it can be transformed to various specific metadata standards, so the GEM data can be indexed and found in other Arctic data portals and repositories.

Ethics and Legal Compliance

How will you manage any ethical issues?

In general, none of the monitoring data in the GEM project concerns people and as such is free from conflicts with GDPR. Also, there is no sensitive or restricted data involved.

It is a term of the funding, that all monitoring data are preserved, shared and made openly available.

Any ethical, environmental impact and legal issues related to the collection of data, are discussed and settled at the two annual GEM coordination group meetings. Participants from all sub-programmes, the GEM secretariat as well as station managers are present at these meetings.

How will you manage copyright and Intellectual Property Rights (IP/IPR) issues?

All scientists contributing to the database, are aware that the data will be open and shared under specific terms of use. The data are owned by the GEM project consortium.

Any dataset downloaded from the GEM database is shared under a Creative Commons License: Creative Commons, Attribution-ShareAlike (CC-BY-SA). <https://creativecommons.org/licenses/>

This means that the datasets can be downloaded and used freely by anyone, under these terms:

- The GEM project must be credited+cited and link to the dataset provided.
- Any derivative products of the data must be published openly under the same terms and with credit to GEM retained.

When downloading the data, users always get pdf's with the terms of use and how-to-cite, in a zip package along with the actual datafiles.

Publishing of data to the database will not be unnecessarily delayed from the time of collection. Usually data are collected during one year, quality assured by the end of the year and then published before may 1. the following year.

Storage and Backup

How will the data be stored and backed up during the research?

During collection and quality assurance of the data, responsibility for storage and backup is with the individual sub-programmes and scientists.

After delivery to the GEM database, the following backup procedures apply:

The data is stored on a SQL Server managed by Aarhus University, IT department. Other research databases are stored on this server as well. This is a trusted repository, placed in Denmark (EU member) and dedicated to store important research data.

A Service Level Agreement (internal document) has been established with the IT department, specifying responsibilities, response times in cases of problems, as well as backup frequency:

- Daily backup every day of the week.
- Full backup once a month, retained in a 12 months archive
- Annual backup retained permanently.

The backups are stored on the Aarhus University network drive, which also is backed up and secured.

As a special measure, to ensure the backups are actually working, the GEM data manager and the AU Arctic DataCentre manager (who is responsible for other databases on the same server) performs restore tests twice a year, to validate that the operational databases can be restored from the backups.

How will you manage access and security?

There is no need to implement high levels of security regarding access to the data, since there is no restricted or confidential data stored.

Before users can download the GEM open data, they are required to register and login on the <https://data.g-e-m.dk> website. The website runs on an Umbraco CMS platform, and users are kept in the built-in member system.

We treat users data with care, and all communications between users and the website is over secure https/ssl encrypted protocol. We do not share users emails or names with third parties, we do not store clear-text passwords, and we keep no detailed personal data.

We log the history of downloads pr. registered user to keep statistics. The website operates with minimal technical cookies required for the website to work.

Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

Within the GEM sub-programmes and disciplines, the scientists have identified ecosystem elements that are of interest for long term monitoring.

Since the goal of the GEM project is to gather long time-series of climate and ecosystem monitoring data, the plan is to keep and share all data that has passed quality assurance and has been submitted to the database.

For this reason, the database grows significantly each year, as new data is added.

Foreseeable uses of the data include:

- Analysis, discoveries and publications carried out by scientists participating in the GEM project.
- Preparatory analysis by visiting scientists to the GEM sites.
- Data material in high school and university courses.
- Supporting data in related research projects.
- Visualization of measurements of key ecosystem variables and climate change.

What is the long-term preservation plan for the dataset?

The plan is to keep the data long term (no planned end of life), in its existing form or eventually by including it in a larger repository if the GEM project ends at some point.

Currently there are no infrastructure charges incurred to the project from the hosting of the database and the open data website. This is covered by Aarhus University and is part of the services the central IT department provides.

As long as the GEM project continues to be funded, money/time is set aside for the GEM Data Manager to handle the import of new data and improve on the data management procedures. The amount allocated approximately equals 300 hours / year.

When needs for significant development of new features or changes to the software and database arises, further funding will be sought.

Data Sharing

How will you share the data?

After collection and quality assurance, data will be shared via the website <https://data.g-e-m.dk> where anyone can register to download data.

We are working continuously to best meet the FAIR data principles (Findable, Accessible, Interoperable, Reusable) and scientific community standards and expectations. This is an ongoing process of improvements.

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

We have already reached several important milestones:

- Schema.org metadata is implemented to allow google and arctic data repositories to crawl and index our datasets. We are using the extension [science-on-schema.org](https://github.com/ESIPFed/science-on-schema.org) (<https://github.com/ESIPFed/science-on-schema.org>)
- A DOI issuer agreement has been established with DataCite via Aarhus University and the Danish DeIC initiative. Every dataset we publish is assigned a DOI to make it easy to link to and cite properly. As a part of this, we publish metadata to DataCite (DataCite MDS 4.3 standard) and our datasets are findable in the DataCite catalogue: <https://commons.datacite.org/doi.org?query=client.uid:dk.gem>
- Collaboration and integration with Interact Virtual Access (VA), an EU funded arctic data portal, is in place. This makes our datasets findable here as well: <https://dataportal.eu-interact.org/>

Are any restrictions on data sharing required?

In general, none of the monitoring data in the GEM project concerns people and as such is free from conflicts with GDPR. Also, there is no sensitive or restricted data involved.

It is a condition set by the funding agencies, that all monitoring data are preserved, shared and made openly available.

Any ethical, environmental impact and legal issues related to the collection of data, are discussed and settled at the two annual GEM coordination group meetings. Participants from all sub-programmes, the GEM secretariat as well as station managers are present at these meetings.

All scientists contributing to the database, are aware that the data will be open and shared under specific terms of use. The data are owned by the GEM project consortium.

Any dataset downloaded from the GEM database is shared under a Creative Commons License: Creative Commons, Attribution-ShareAlike (CC-BY-SA). <https://creativecommons.org/licenses/>

This means that the datasets can be downloaded and used freely by anyone, under these terms:

- The GEM project must be credited+cited and link to the dataset provided.
- Any derivative products of the data must be published openly under the same terms and with credit to GEM retained.

When downloading the data, users always get pdf's with the terms of use and how-to-cite, in a zip package along with the actual datafiles.

Publishing of data to the database will not be unnecessarily delayed from the time of collection. Usually data are collected during one year, quality assured by the end of the year and then published in May the following year.

Responsibilities and Resources

Who will be responsible for data management?

Data Management really involves all participants in the project:

1. In each sub-programme, the programme manager is responsible for the structuring, storage, collection, quality analysis and delivery/sharing of their datasets. The procedures are described in details in the monitoring manuals. <https://g-e-m.dk/gem-programme-managers-and-logistics/>
2. The GEM Data Manager (Jonas K. Rømer, jkr@bios.au.dk) is responsible for: a) Receiving, validating and importing the datasets to the database, communicating with the sub-programme managers and making the data openly available. b) Supporting and maintaining the database and the open data website, the systems they run on, handling technical questions and error reports from users. *The hardware / infrastructure and operating systems are maintained and delivered as a service (IaaS) by Aarhus University central IT department.* c) Participating in relevant fora and meetings, to follow technological developments and better integrate our data and initiatives with other arctic data portals. d) Continuously improving the data flows and the open data website.
3. A database group, with interested participants from across the sub-programmes and the GEM secretariat, supports the data manager and helps

prioritize and guide the development initiatives. The group typically meets twice a year, before the coordination group meetings and other than that, everyone can bring out questions or issues when needed.

4. Twice a year, the GEM project hosts a coordination group meeting with participants from the GEM secretariat, all GEM sites and all sub-programmes. The data manager presents the state of the database, progress of the yearly data import, and demonstrates new developments since the last meeting.

What resources will you require to deliver your plan?

The annual budget, funded by annual fees from the participating institutions, goes towards ~300 hours of work for the GEM Data Manager. This time is prioritized, with the primary goal of receiving, validating and importing the annual data, improving the data flows and metadata and keeping the systems running.

New initiatives are important to keep up with the expectations of users as well as ongoing developments, standards and requirements from the arctic data community at the larger scale. These initiatives need to be funded individually, when needed.

Currently there are no infrastructure charges incurred to the project from the hosting of the database and the open data website. This is covered by Aarhus University and is part of the services the central IT department provides. *We have started storing binary data, and this can lead to increased storage costs that the GEM project eventually may have to pay for.*