INTERACT GIS Data Management Plan

In collaboration between INTERACT[1] - International Network for Terrestrial Research and Monitoring in the Arctic and SITES[2] - Swedish Infrastructure for Ecosystem Science, an information system called INTERACT GIS has been developed and made available for usage at INTERACT and SITES field stations. Interact GIS is a network and web-based service provider for currently more than 85 terrestrial research stations located in the arctic and northern boreal and alpine areas and for the researchers visiting these stations. A subpart of the network, SITES GIS, is focused on SITES research stations, which represent a variety of Swedish climate zones and ecosystems.

This data management plan has been created to layout the principles and strategy on how the Interact GIS project will manage data. When reading the document, keep in mind that it is the principal plan and not full documentation. The plan will be kept as a living document, and whenever the project reaches new achievements or a need for further work on data management is identified, new content will be added to appendices in this data management plan (see *Appendix)*. It is important for us to keep and show coherence between the plan and choices, initiatives and achievements reached.

The Interact GIS Management Organization commits to continuously review developments and, if necessary, update the plan at least annually. Revision history will be kept, see section *Versioning* in this document.

A general data management plan for all INTERACT projects has previously been published: D4.1 DATA MANAGEMENT PLAN. This, along with suggestions from https://dmptool.org and https://go-fair.org as well as existing documentation and discussions within Interact GIS has served as input and starting points in creating this specific data management plan for Interact GIS.

The most current version of this document, license and how to cite, can always be found here:

https://doi.org/10.5281/zenodo.7143021

[1] https://eu-interact.org/

[2] https://www.fieldsites.se/en-GB

Contents

INTERACT GIS Data Management Plan 1
The Data Managed in Interact GIS 3
Core Data 3
Vocabularies 3
People / Personal Data 4
Documentation and Metadata 4
Ethics and Legal Compliance 4
Development, Operations, Storage and Backup 4
Selection and Preservation 5
Data Sharing 5
Responsibilities and Resources 7
Interact GIS Management Organization 8
Appendix 9
Current State and Achievements Regarding the Topics in the Data Management Plan 9
Current State, Achievements and Further Work Regarding FAIR Goals 9
Versioning 10

The Data Managed in Interact GIS

Interact GIS is a network and web-based service provider for currently more than 85 terrestrial research stations located in the arctic and northern boreal and alpine areas, and for the researchers visiting these stations.

- ***Interact GIS manages*** detailed research infrastructure data about stations, facilities, activities and visitors´ applications.
- ***Interact GIS does not manage*** research project output data but seek to provide users with cross links to research data outputs in other data repositories. Another component of Interact, outside the scope of Interact GIS, is Interact Virtual Access (VA). Interact VA provides a platform that indexes research data from the Interact stations: https://dataportal.eu-interact.org/. Likewise, for the SITES stations on Interact GIS, research data can be found in the SITES data catalogue: https://data.fieldsites.se/portal/.

Different types of data is managed in the system, and conceptually they can be grouped as shown in Figure 1.

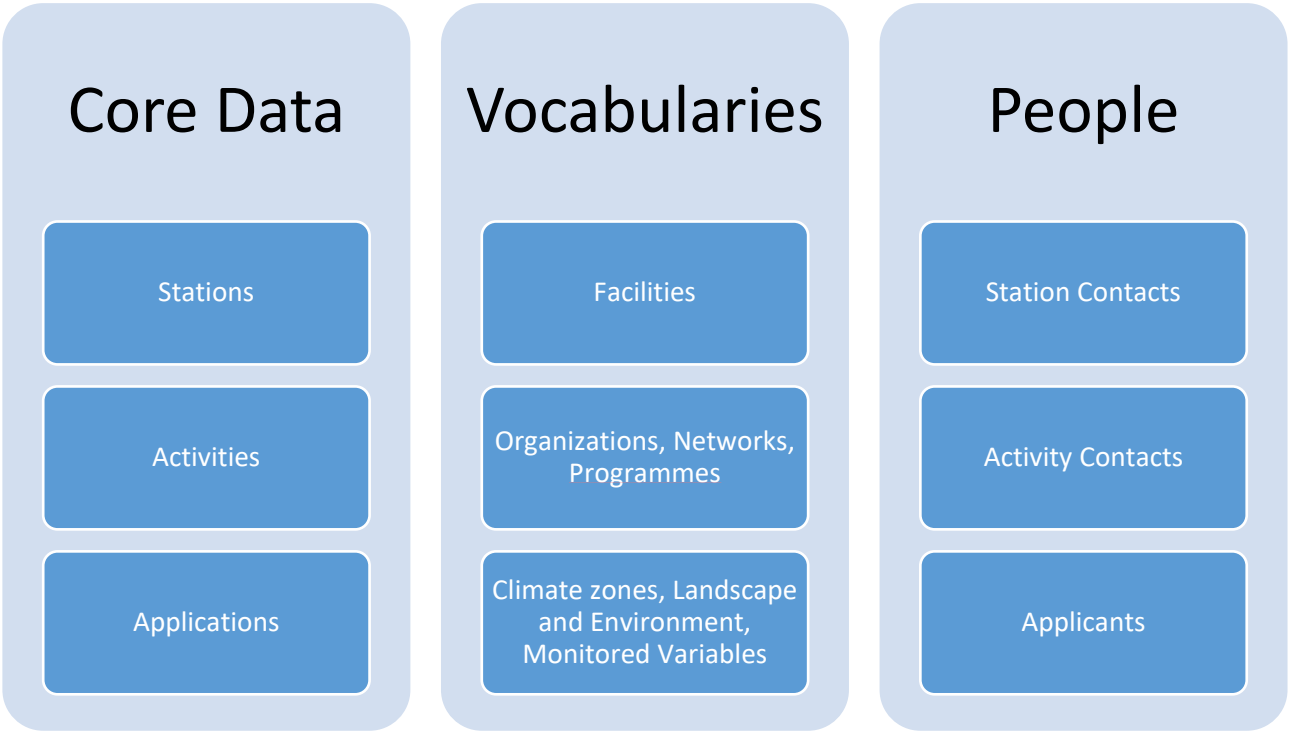

Figure 1. Types of data managed in Interact GIS. Core Data, Vocabularies, People / Personal Data.

Core Data

The primary content and what Interact GIS is all about: Detailed data about research stations, activities hosted on the stations (projects, courses and conferences) and handling visitors´ applications.

Vocabularies

Rich, controlled vocabularies are developed and used in Interact GIS to enhance and structure core data with organized categories and attributes.

People / Personal Data

Personal user data is managed according to different roles in the system: administrators, station managers, activity managers (applicants) and their activity participants. Personal information varies in detail for the different roles but must be handled as sensitive data and subject to GDPR regulations[3]. For instance, applicants' education and qualifications are stored in the system. All stations participating in the application module of the network must submit a signed GDPR form when joining Interact GIS.

Documentation and Metadata

The data in Interact GIS is from the very conceptualization of the project about building a data collection of detailed and structured descriptive data about research stations and their facilities and activities: We are not dealing with unstructured data, very different types of data, or data that is hard to understand without accompanying documentation.

With regard to documentation and metadata, Interact GIS works with:

- The research data community and especially other groups and data portals dealing with arctic research infrastructure to establish and adhere to common and well-documented standards for data and metadata.
- Sharing metadata and full datasets with other portals to make Interact GIS and its stations more findable for researchers and attract new users to Interact GIS.

Ethics and Legal Compliance

Interact GIS is committed to protecting personal data and following GDPR regulations. Users shall whenever they submit personal information to the system, be made aware of terms and conditions.

Any ethical and legal compliance issues will be discussed and dealt with in the project group, with legal advice sought from experts within the participating institutions when needed.

Development, Operations, Storage and Backup

The strategy is to place the responsibility for technical operations and development of Interact GIS with one institution partnering in the Interact GIS project. This institution must have the capabilities and skilled staff to operate, maintain and develop new features for the system.

Development of new features are discussed and initiated in the Interact GIS project group, in close cooperation with the development staff.

The Interact GIS project funds pay the responsible institution for the maintenance and development of the system.

Server location: Servers must be placed in a trusted center in an EU country, to avoid GDPR issues with international transfer of personal data.

Backup: Backup must be scheduled automatically according to a rolling backup scheme. In case of a database breakdown, it must be possible to restore quickly to a recent point (daily backups). In case of undetected errors that corrupts the database or causes loss of data over time, it must also be possible to restore to a less recent point in time. Preferred – keep daily backups rolling for one month, keep monthly backups rolling for 12 months, and keep annual backups permanently.

[3] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02016R0679-20160504&qid=1532348683434

Selection and Preservation

Once a station or an applicant has been accepted into the Interact GIS system, its data is made publicly available on the website. Stations are responsible for and can update their data; Likewise, applicants also create and update their data. As such, quality and completeness of data are expected to evolve over the span of a stations and applicants continuous use of the system.

Interact GIS aims to preserve data long term and principally operates towards a forever horizon. Should Interact GIS at some point be unable to continue operations, the data must be deposited as open data in a suitable data repository.

GDPR ensures that users can have their personal information removed upon request. A pseudonymization scheme is applied to preserve the consistency of the database over time while still allowing to remove person referable data.

Data Sharing

INTERACT GIS aims towards the goal to make data easily findable and available to users as well as to relevant indexing portals and search engines and share well defined data with the arctic research infrastructure community.

Referring to the section: The Data Managed in Interact GIS, Figure 1, there are three conceptual groups of data: Core data, Vocabularies and People. Data sharing primarily concerns the first two, while people data cannot be openly published or shared due to GDPR regulations.

Open data and the FAIR principles are an important guidance for Interact GIS. In terms of FAIR, Interact GIS plays two roles, see Figure 2:

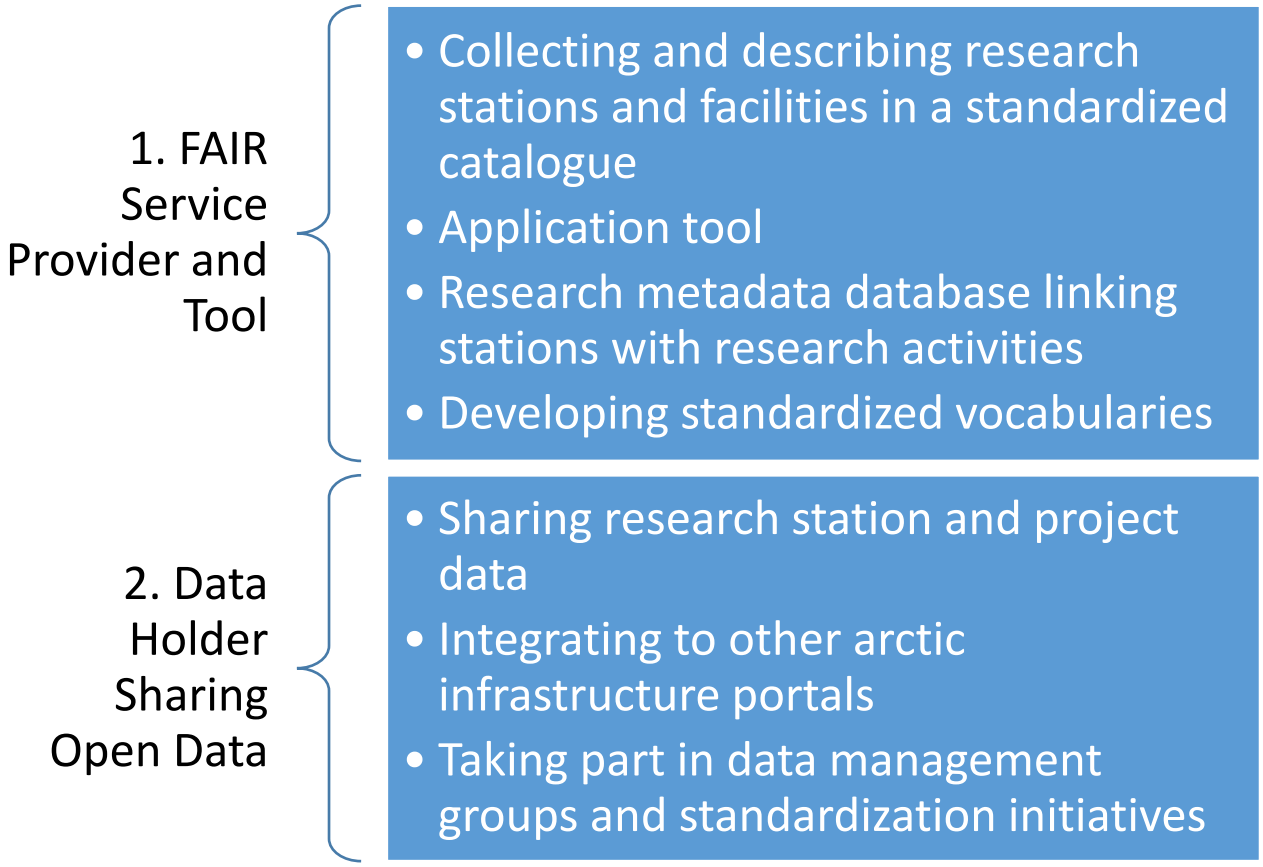

Figure 2. Interact GIS plays two roles in relation to open data and FAIR principles. 1: Interact GIS is a FAIR service provider and tool, helping research stations to become more visible and manage and share their data in a structured way. 2: Interact GIS holds the collected data and according to FAIR principles wish to share this data further and engage in the community of research data and other arctic infrastructure portals.

Table 1 below shows: The FAIR goals and methods to fulfill them as seen by the general Interact Data Management Plan and go-fair.org. In the appendix, Table 3 shows: How Interact GIS addresses the FAIR goals and plans to work on them in the future.

Goal	**Methods recommended (Interact general DMP)**	**Methods recommended (https://go-fair.org)**
Findable	Strategy: Exchangeable and discoverable metadata approach, sharing with other portals. Metadata standard suggested: ISO 19115 – Geographic Information – Metadata Use Controlled Vocabularies: for instance GCMD and discipline specific vocabularies. Assign globally resolvable persistent identifiers to data.	F1. (Meta)data are assigned a globally unique and persistent identifier F2. Data are described with rich metadata (defined by R1 below) F3. Metadata clearly and explicitly include the identifier of the data they describe F4. (Meta)data are registered or indexed in a searchable resource

Accessible	Enable programmatic access to data for exploration and visualization from users tool of choice (Matlab, R, Python etc.). Open Data Access Protocol recommended - https://www.opendap.org/	A1. (Meta)data are retrievable by their identifier using a standardised communications protocol A1.1 The protocol is open, free, and universally implementable A1.2 The protocol allows for an authentication and authorisation procedure, where necessary A2. Metadata are accessible, even when the data are no longer available
Interoperable	Use self-explaining file formats. Use discipline specific standards and vocabularies.	I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation. I2. (Meta)data use vocabularies that follow FAIR principles I3. (Meta)data include qualified references to other (meta)data
Reusable	Use creative commons license, CC-BY or CC-BY-SA. When sharing data, provide information on data quality.	R1. (Meta)data are richly described with a plurality of accurate and relevant attributes R1.1. (Meta)data are released with a clear and accessible data usage license R1.2. (Meta)data are associated with detailed provenance R1.3. (Meta)data meet domain-relevant community standards

Table 1. FAIR goals and recommendations from Interact General Data Management Plan and from https://go-fair.org .

Responsibilities and Resources

INTERACT GIS has been developed on the initiative of the Swedish University of Agricultural Sciences (SLU), and with SLU as the principal investigator in the first phase of the project.

The development work has most recently received its primary funding from:

- The EU's Seventh Framework Programme and EU Horizon 2020 (via INTERACT)
- The Swedish Research Council (via Swedish Infrastructure for Ecosystem Science (SITES)) under the grant no 2017-00635.

The conceptualization of Interact GIS and the first development work was initiated in 2011. INTERACT (EU) and SITES has been primary funders via several grants since then.

The funding is given on a time-limited basis for a number of consecutive years. Renewed funding from the same or similar sources is likely but not guaranteed. In the event that at some point in the future funding cannot be obtained, INTERACT GIS can be kept operational (but not developed further), by charging a small fee of estimated less than 100 EUR pr. year from each station registered in Interact GIS.

SLU owns the intellectual property rights to INTERACT GIS, and the right to freely develop the INTERACT GIS software and offer it to users freely determined by SLU. This applies regardless of the number of open

components in the final software. SLU is also free to choose administrative and operational organizations for INTERACT GIS[4].

Interact GIS Management Organization

The establishment of an Interact GIS management organization has been initiated and has also led to the creation of this data management plan. The management organization shall support and tend to the different interests of the participants in the Interact GIS project, both funders, universities, technical development team and research stations. The management organization shall help guide the future of Interact GIS and the data it holds.

The Interact GIS Management Organization will meet with a suitable frequency, and form a team of various roles:

- System Operator (Institution responsible for development, operations, storage and backup)
- Daily administration team (Interact, SITES, System Operator, super user from a research station)
- Head / secretary of the management organization
- Technical adviser(s), on data management, FAIR and open data, development.

[4] This is an unclear and not ideal state. We are working for an agreement with SLU to release Interact GIS under a clear and free open source license and will update the plan accordingly when this happens.

Appendix

Current State and Achievements Regarding the Topics in the Data Management Plan

Topic	Highlight
The Data Managed in Interact GIS	Administration functionalities enables the creation and management of all types of core data via the web interface by different user roles.
Documentation and Metadata	A detailed data model and vocabularies designed within the project helps collect well documented data. The data model and vocabularies are internal but comprehensive, and can be mapped to other standards when suitable ones are found. The data model and vocabularies are not yet published, but available upon request. Interact GIS is collaborating directly with European Polar Board (EPB) and SAON CON Polar Observing Assets Working Group regarding standardization and metadata. https://www.europeanpolarboard.org/ and https://www.polarobservingassets.org/ .
Ethics and Legal Compliance	A detailed personal data processing declaration has been created, with statements on all GDPR related issues. Users are shown this and acknowledge it whenever they submit personal data: https://interact-gis.org/Files/INTERACTGIS_PersonalDataProcessingDeclaration.pdf
Development, Operations, Storage and Backup	The technical operations, storage and backup as well as ongoing development is placed with ICT Services and System Development at Umeå University (UmU ITS), Sweden. Source control is managed using Azure DevOps Server on Premises. The website and database servers are hosted at Umeå University.
Selection and Preservation	
Data Sharing	See Table 3…
Responsibilities and Resources	
Interact GIS Management Organisation	

Table 2. What we have done to meet the objectives set forth in the data management plan.

Current State, Achievements and Further Work Regarding FAIR Goals

Goal	Current achievements	Further targets / ideas
Findable	An integration to the PolarDex portal has been developed, that makes INTERACT GIS data findable via https://polardex.org . Very detailed vocabularies developed within the interact community covering a number of categories: Disciplines, Climate zones, Landscape and environment, Monitored variables.	Expose discoverable metadata: schema.org / science-on-schema.org, ISO 19115 xml, OGC Observations and Measurements endpoints. Metadata entry for each station. Define datasets clearly, assign URLs, DOIs or other globally unique, persistent and resolvable identifier. Publish the vocabularies as Interact GIS controlled vocabularies. Map to other controlled

		vocabularies when possible.
Accessible	The website https://www.interact-gis.org allows users to explore all data on stations, facilities, projects, networks. Advanced and intuitive filtering options allows users to quickly find and view the data that's relevant for them. An API has been developed that allows programmatic access to all INTERACT GIS data: https://interact-gis.stage.its.umu.se/swagger/ui/index (preliminary url, not launched yet). The api follows the OpenAPI standard and is documented using swagger tools, making it easily understandable and usable.	Allow users to export search results and station data as csv or other simple solution. Keep developing the API that's already been started. Document it on the website so users learn about it. Share the data as Open Geospatial Consortium (OGC) services (web feature services WFS).
Interoperable	The format returned from OpenAPI is JSON, a well-known and self-explaining format that can be read by humans as well as machines. The data returned is richly described with a detailed data model, and has many relevant attributes. It uses explanatory attribute names.	Interact is participating in research networks working to standardize the exchange of arctic observation facilities and research projects metadata. Once this has been formalized as a schema definition, the OpenAPI results should refer to and validate against this standard. Other standards and schemas should be considered as well, whenever possible - data should reference a standard schema.
Reusable	Data is richly described with a detailed data model and has many relevant attributes. It uses explanatory attribute names.	We are working to agree on and determine a creative commons license, so users can be informed about on what terms they can reuse the data. The Interact GIS datasets shared must be defined more exactly and assigned globally unique persistent identifiers.

Table 3. FAIR goals, current achievements and further targets / ideas.

Versioning

Current version: v1.0 published September 2022

Version History:

2022-10-04: v1.0. First finalized, reviewed and published version.