
INSTITUTIONAL POLICY FOR DATA MANAGEMENT

Approved 15th of February 2021

This document describes a set of **policy** guidelines for Data Management at IGC. IGC is committed to open, rigorous, and reproducible science and therefore the guiding light of Data Management at IGC will be the FAIR (Findable, Accessible, Interoperable, and Reusable) and RRI (Responsible Research and Innovation) principles. IGC considers all publicly funded research data a public good, produced in the public interest, and should be made available in a timely manner, with as few restrictions as possible, but taking into account the intellectual property of its researchers.

OWNERSHIP OF DATA

All data generated by the researchers at IGC is owned by IGC and IGC will remain the custodian of such data should the researcher leave the Institute. Exclusive rights to data generated at IGC shall not be granted to third-parties, unless mandated by the funding agency or other contracts.

RESPONSIBILITY OF DATA MANAGEMENT

The principle investigator of the research project will be the main responsible for data management while the research is ongoing. This responsibility can be delegated on others (lab manager, data steward). In the case of research performed in cooperation with other institutions, all parties shall agree on a common data management plan.

DATA MANAGEMENT PLANS

All research performed at IGC shall be governed by its own project-specific DMP, submitted at the time of the grant submission. In case the funding body does not require it, one shall be provided to IGC regardless. Preliminary DMPs can be initially submitted, but should be regularly updated as the project unfolds, considering the realities of the project. Project specific DMPs will be specified by the responsible researcher in accordance to these institutional guidelines. For all projects dealing with personal data or human samples, the DMP is a required component of ethical review. Special attention should be given to personal data and compliance to General Data Protection Regulations (GDPR).

DMPs should specify specific criteria for the identification of significant data (data that will be shared and retained – see below). These should at least include:

- Data that substantiates published research findings
- Unrepeatable observations
- Experimental results that would be impossible or expensive to reproduce

DMPs should specify the manner in which data will be stored (file formats), annotated/documentated (metadata) and shared (repositories).

Grant applications should budget for the associated costs of data storage and management for the duration of the project.

STORAGE & MANAGEMENT

All data belonging to research projects at IGC shall be stored on the institutes designated storage infrastructure. All digital data produced within the scope of a research project shall be stored in this space, except in cases where this is manifestly impossible, and in accordance with other legal obligations (for example, GDPR). Non-digital data that is unsuitable for digitisation, should be stored securely, catalogued and indexed, and a note should be made in the project's space with this information. All data should be accompanied of description metadata and documentation, allowing its identification and effective reuse.

In the interest of reproducible and transparent science, all additions, deletions and alterations to datasets shall be documented. Ideally, datasets will be content versioned and changes will be properly documented and justified. Significant research results, deriving from research data will reference the particular version of the dataset that was used.

Standardized naming conventions for data files shall be adopted per project. These should include a project acronym, date, version number and other descriptors of the data.

SHARING OF DATA

Sharing of data should follow the FAIR (Findable, Available, Interoperable and Reusable) principles. To this end:

- Metadata will be provided by the responsible researcher at the time of data submission to facilitate discovery.
- Data to be made publicly available shall be deposited in a public repository and associated with a persistent identifier (eg DOI).
- Data shall be made available preferentially in open file formats. In the cases in which this is not possible, software that can be used to access the data will be named (including where it can be obtained and which version was used).
- Protocols used to collect the data shall be provided alongside the data.
- Clear descriptions of the data shall be provided. Ideally, a standardized/controlled vocabulary should be used to describe the data entities.
- Any custom code for analysis or generation of the data, including visualizations tools, shall be made available alongside the data, properly documented. Whenever possible, procedures for the reproducibility of these components, should be made available (listing of packages/libraries used, including version numbers, at the minimum).
- Data should be made available using the least restrictive license possible. The recommended licence for the sharing of data is CC-BY-NC, allowing re-use for any non-profit purpose with attribution.

EMBARGO PERIODS

All significant data to be shared should be made publicly available within 2 years after the project's end. This does not apply to data supporting published findings, for which the data should be made available at the time of publication.

RETENTION POLICY

Research data deemed significant will be preserved by IGC for a period of at least 10 years. This period can be extended in case of contractual obligations, manifest public interest, patent applications, or in cases where the results became contentious or disputed within the initial 10 years.