RESEARCH DATA MANAGEMENT

GUIDELINES

Document owner: Directorate
Title and Version: Research Data Management Guidelines

Table of Contents

1) Object

Human Technopole's overarching mission is to promote and contribute to improving human health and wellbeing, including healthy ageing. It pursues these aims by carrying out frontier research (a mix of fundamental and translational research) in the life sciences.

The HT operating model represents a combination of creating an internationally competitive research institute and being a contributor to the Italian research community through the provision of User facilities and training, as well as through broad academic and translational collaborations.

Rapid and open Research Data sharing strategically supports this mission by enabling research and accelerating translation. HT recognizes the relevance of records management for supporting the principles of quality and integrity in scientific research while being committed to pursuing the highest standards of data collection, storage, and preservation in accordance with the FAIR (findable, accessible, interoperable, reusable) principles.

This document lays out HT's guiding principles, expectations and assurances towards pursuing the highest standards in scientific data management. HT prioritises these principles in the Foundation's interactions with Researchers and infrastructure Users, including those of the HT National Facilities. A fundamental principle is that data and results from publicly funded research should be made publicly available to the greatest extent possible, ***as open as possible, as closed as necessary***.

2) Abbreviations, terms, definitions

2.1. Abbreviations

DMP : Data Management Plan

FAIR : Findable, Accessible, Interoperable and Reproducible

GDPR : General Data Protection Regulation

HT : Fondazione Human Technopole

RDM : Research Data Management

2.2. Terms and Definition

- DATA MANAGEMENT PLAN means a plan for the entire lifecycle of Research Data that outlines how data will be managed from the point of collection/receipt at the start of a research project all the way through to what will happen to the data once the project ends.

Data management plans assure that Research Data are accessible, traceable, available, authentic, citable, and that they adhere to clearly defined legal parameters and appropriate safety measures governing subsequent use.

- EXTERNAL USER of HT's National Facilities is a Researcher not affiliated with HT who accesses the facilities to perform a research project, or parts of it, with the support of HT staff.
- METADATA means information that describes significant aspects of a dataset. For example, this may include authors, title, date of publication, unique identifier, a description of what the dataset contains and usage licence. Metadata provides other Researchers with the information needed to understand and reuse the dataset as well as making the dataset more findable.
- PERSONAL DATA means any information relating to an identified or identifiable natural person ("Data Subject"); an identifiable person is one who can be identified, directly or indirectly, by reference in particular to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to his or her physical, physiological, genetic, mental, economic, cultural, or social identity.
- PROCESSING means any operation or set of operations which is performed upon personal data or sets of personal data, whether or not by automatic means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction.
- RAW DATA or Primary Data means any data that has not been processed, coded, formatted, or analysed for useful information.
- RESEARCH DATA means information in digital, computer-readable or paper-based format that is contained or presented in various ways including notes, facts, figures, tables, images (still and moving), audio or visual recordings; and which is collected, generated or obtained during the course of or as a result of undertaking research (which includes but is not limited to conducting laboratory experiments, conducting trials, observation, modelling, or analysis of data and derivation from existing evidence); and which is subsequently used by the Researcher as a basis for making calculations or drawing conclusions to develop, support or revise theories, practices and findings. These might be quantitative information or qualitative statements collected by the Researcher in the course of their work. Data may be raw (e.g., direct from measurement or collection) or derived from primary data (e.g.,

cleaned up, or processed). Research Data can also be derived from existing sources where the rights may be held by others, according to the respective usage license.

- RESEARCH DATA MANAGEMENT means the organisation, storage and preservation of Research Data created during a research project. It covers the initial planning, data collection, processing, analysis, preservation, data sharing, secure disposal of data, and ensuring availability for reuse.
- (HT) SCIENTIST means HT staff who performs research-related activities at HT or any external Researcher performing their research at HT (including but not limited to Heads of Research Centre, Heads of Facility, Group Leaders, Senior Managers, Staff Scientists, Postdocs, PhD Students, Postgraduate Fellows, Internship students, Technicians, Lab Managers, Postdoctoral Associates, Affiliate Scientists, and Scientific Visitors – including e-Visitors).

3) Scope

This document aims at defining the HT Guidelines in relation to Research Data Management. All HT Scientists and professionals involved in HT research-related activities, as well as individuals and organisations that access HT facilities and services, including National Facilities, must comply with the following Research Data Management Guidelines, irrespective of the source of their funding or area of research.

4) Regulatory references

FAIR Principles - GO FAIR (go-fair.org)

Open Science (europa.eu)

EMBL Internal Policy No. 71 “Open Science and Open Access”

UNESCO Recommendation on Open Science

EU General Data Protection Regulation

UKRI-020920-ConcordatonOpenResearchData.pdf

5) Background and Purpose

The purpose of these Guidelines is to:

5.1 illustrate Research Data Management principles for all scientists at HT and Users of HT services and facilities, including National Facilities, to ensure good research practice and procedures;

5.2 maximise the openness of outputs of HT's research, services, innovation, and technology development;

5.3 foster responsibility for Research Data management of the scientific community at large through the promotion of best practices;

5.4 make the assessment and reporting of science outcomes more accurate and transparent;

5.5 lay the foundations to ensure that Research Data are recorded, stored, retained, accessed, and disposed of securely and in accordance with all legal, statutory, ethical, contractual, and funding requirements.

6) Data sharing

6.1 HT is committed to sharing its Research Data broadly with the scientific community.

6.2 HT expects all outputs of research activities performed by its Researchers as well as external Users accessing the National Facilities and supported by public funds allocated to HT to be made publicly available as quickly as possible. This refers to research articles, data, software, protocols, and similar outputs.

6.3 Provided compliance with data protection and other applicable laws and regulations, HT expects Research Data and all data behind research articles (including raw data) to be made public and to adhere to the FAIR principles (Findability, Accessibility, Interoperability, and Reuse of digital assets) (https://www.go-fair.org/fairprinciples or DOI 10.1038/sdata.2016.18).

6.4 Research Data are normally made available no later than the time of public dissemination of the related scientific results. However, depending on circumstances the sharing of scientific data may be delayed (e.g., to legitimately seek protection of intellectual property rights) or excluded (e.g., if not allowed under data protection regulations).

6.5 HT aims to provide rapid access to datasets of use to the research community and will place these in publicly accessible repositories, when possible, and provided compliance with applicable regulations. Data should be deposited in the appropriate thematic (community) database as a top priority. If no appropriate community database exists, then

a general database may be used. HT will strive to adhere to data and interoperability standards to maximise access and ensure ease of integration with other global resources.

6.6 When useful for the research community and appropriate in terms of additional effort, and provided compliance with applicable regulations, a Researcher might choose, or be requested, to publish processed as well as raw datasets, along with the appropriate metadata (rather than, e.g., only subsets of data selected for discussion in a publication). As data pipelines include various levels of processing, the meaning of 'processed/raw' should reflect the current community best practices.

6.7 Provided compliance with data protection and other applicable laws and regulations, HT encourages the publishing of datasets that are not accompanied by a research article when the dataset has the potential for reuse.

6.8 HT expects all research projects to include a Data Management Plan. This applies to any work that is funded by grants, or is part of PhD and postdoctoral projects, or is intended to result in a scientific article. It also applies to all projects approved for access to National Facilities by external Users.

6.9 The above principles apply to all data, including those with restricted access due to specific regulations, or for ethical reasons. The principle "as open as possible, as closed as necessary" applies.

7) Personal data protection

7.1 Conducting life science research carries responsibilities to protect confidentiality and the privacy of research participants. Access to certain datasets will therefore be carefully managed and granted on a need-to-know basis and in compliance with applicable regulations as well as in a transparent manner to appropriately qualified Researchers to ensure best practices with regard to ethical aspects.

7.2 HT will never share the identity of study participants nor information that allows for their identification through means reasonably likely to be used. HT has strict measures in place to ensure that the privacy of study participants is protected. Depending on the circumstances, these measures may include pseudonymisation techniques, e.g., when full anonymisation is not possible.

7.3 When a research project requires collection of human data and subsequent processing by HT, such data are either anonymised or (where anonymisation is not possible or compatible

with the research purposes) pseudonymised. Said data are expected to be anonymised or pseudonymised by data providers or third-party collaborators prior to receipt at HT.

7.4 Where processing pseudonymised data, the information that allows to attribute the data to a specific individual ("re-identification key") will remain with the data provider or third-party collaborator (hence, HT does not hold nor have access to said information nor to other means reasonably likely to be used to re-identify individuals).

7.5 Collection, use, storage and sharing of personal data is covered by the General Data Protection Regulation (GDPR). In addition, HT has a dedicated "Policy on the Organisation and management model for the protection of personal data".

7.6 According to the GDPR, personal data shall follow the principles of (Art. 5 GDPR):

- "Lawfulness, fairness and transparency"
- "Purpose limitation"
- "Data minimisation"
- "Accuracy"
- "Storage limitation"
- "Integrity and confidentiality"

7.7 Ethical Approval

All research using personal data requires approval from an appropriate Ethics Committee. Researchers must also have participants' consent to collect, use, store and share their personal data.

7.8 Data storage, retention and destruction

7.8.1. HT scientists must be mindful of how they store all human data, regardless of whether it is covered by the GDPR, and should ensure that the established appropriate security measures are observed to limit access and prevent misuse.

7.8.2. HT scientists must ensure that all Research Data in digital and computer-readable form is stored securely in a durable format appropriate for the type of Research Data in question, including adequate metadata, and backed-up regularly.

7.8.3. Research Data shall be held for a minimum period of time as specified in the Data Management Plan, which shall not be less than one month after the end of

the research project. Research Data that is not deemed ‘significant’ do not need to be retained beyond the end of the research project.

7.8.4. The disposal and destruction of Research Data must be undertaken in accordance with the Data Management Plan. The agreed processes for the timing, manner and recording of Research Data disposal and destruction should be included in the Data Management Plan of the project and stored with other project information and documentation.

7.8.5. Data generated by HT National Facilities for external Users’ projects will be handled as specified in the project’s Data Management Plan and the Data Processing Agreement related to the User Access Agreement. It is important to notice that in this case the User will retain ownership of the data. The User will be able to freely download the generated data within the time span set forth as in 7.8.3, and HT will delete its copy after the time span set forth as in 7.8.4, in accordance with the data retention policy as defined in the project Data Management Plan and the Data Processing Agreement related to the User Access Agreement.

8) Software

8.1 For the purpose of these Guidelines, software includes:

- Code that is needed to reproduce a specific data analysis result (or demonstrate a method) reported in a scientific paper, ranging from small projects to large consortia. The objective of making open this type of software is to ensure transparency and reusability, which are required for reproducibility.
- Software that supports HT National Facilities services or the processing of data generated at HT facilities. The objective of making open this type of software is primarily transparency, while reusability is a secondary aspect.
- Broad purpose tools and methods. The objective of making this type of software open is primarily reusability.
- Code that is part of training material and/or is not in the above categories (e.g., exercises used in teaching courses).

8.2 HT expects all of the above types of software, whether used for services or research activities, to be Open Source by default, and made available in open/community software repositories.

8.3 The above does not apply to transient scripts that remain private and do not support publicly available data or protocols.

9) Enforcement

9.1 The matters covered in these Guidelines are of central importance to HT's mission and good standing, and to the future careers of HT scientists.

9.2 All members of HT staff are required to comply with these Guidelines in carrying out their duties for HT.

9.3 National Facilities external Users are required to comply with these Guidelines in managing data produced during the access.

9.4 HT personnel found to have violated these Guidelines may be subject to disciplinary action.