Guidelines for Research Data Management at ETH Zurich (RDM Guidelines)[1]

The Executive Board
pursuant to Article 4, Paragraph 1g and Art. 10, Paragraph 4b of the ETH Zurich Organization Ordinance of 16 December 2003 (RSETHZ 201.021)[2] as well as pursuant to Article 3, Paragraph 3 and Article 10, Paragraph 4 of the ETH Zurich Guidelines on Scientific Integrity (Integrity Guidelines) of 1 January 2022 (RSETHZ 414)[3] issues the following guidelines:

Preamble

ETH Zurich considers *Research Data* of high quality as an essential resource and *Research Data Management* as part of good scientific practice according to internationally recognised standards for scientific integrity. ETH is convinced that good data management enhances the robustness of research findings, facilitates research collaborations, makes research results more reproducible, and strengthens society's trust in science.

ETH Zurich therefore commits to creating a research environment that supports FAIR[4] *Research Data Management* by design. ETH Zurich engages on a national and international level to provide a sustainable and suitable infrastructure for *Research Data Management* to offer its researchers information and services that satisfy the highest standards. The ideal to be strived for is *Open Research Data* (*ORD*) — the practice of making *Research Data* publicly available, accessible and reusable while always respecting legal and ethical framework conditions ("as open as possible, as closed as necessary").

The Guidelines for Research Data Management (hereinafter "RDM Guidelines") specify further details of *Research Data Management* in accordance with Art. 10 Para. 4 of the ETH Zurich Guidelines on scientific integrity (Integrity Guidelines)[5]. They take into account strategic developments and activities on the national level and within the ETH Domain[6]. They serve to establish minimal requirements for all ETH Zurich members involved in scientific research and to define responsibilities. They can be complemented by departmental regulations which foster recognised *Community Standards* pertaining to *Research Data Management* and *Open Research Data* by design.

---

[1]   These guidelines owe in particular to "Open Research Data: Position of the ETH Domain, adopted by the ETH Board on 13/14 May 2020" and to the DLCM Research Data Management Template (Version 1.3).
[2]   RSETHZ 201.021 (https://rechtssammlung.sp.ethz.ch/Dokumente/201.021.pdf)
[3]   RSETHZ 414 (https://rechtssammlung.sp.ethz.ch/Dokumente/414.pdf)
[4]   Wilkinson, Mark et al., 2016, 'The FAIR Guiding Principles for scientific data management and stewardship, in: *Scientific Data* 3 (1): 160018. doi:10.1038/sdata.2016.18
[5]   RSETHZ 414 (https://rechtssammlung.sp.ethz.ch/Dokumente/414.pdf)
[6]   Swiss National Open Research Data Strategy, adopted by the Delegation Open Science (swissuniversities) on 23 April 2021, Swiss National Open Research Data Action Plan, adopted by the Delegation Open Science (swissuniversities) on 17 November 2021.

Chapter 1: General terms

Art. 1 Purpose

[1] Based on article 10, paragraph 4 of the ETH Zurich Guidelines on Scientific Integrity (hereinafter Integrity Guidelines)[7], the purpose of these guidelines is to define principles for *Research Data Management* (*RDM*) and *Open Research Data* (*ORD*) at ETH Zurich.

[2] The guidelines are complementary to the Open Access-Policy,[8] the Statutes of the ETH Zurich Ethics Commission,[9] the Policy for Information and Communications Technology,[10] and the Exploitation Guidelines.[11]

[3] Applicable national and international law always takes precedence (e.g. personal data protection, export control of dual-use technology, archiving, copyright).[12]

Art. 2 Scope

[1] These guidelines apply to all members (including academic guests)[13] of ETH Zurich involved in scientific research (hereinafter "researchers").

[2] The present guidelines complement and do not supersede regulations issued by Swiss, European and international funding agencies, which prevail in case of any discrepancies.

[3] In research projects financed by commercial and/or private third parties, the guidelines should be taken into account when negotiating conditions with partners and should be followed as far as possible within the contractual obligations.

[4] Within the context of academic research cooperation that extends beyond ETH Zurich, researchers of ETH Zurich involved are responsible for compliance with these guidelines.

Art. 3 Definitions

[1] *Research Data* are all data that the relevant scientific community accepts as necessary for validating research findings (based on Article 10 of the Integrity Guidelines).[14]

[2] *Metadata* are data that provide information about data.

[3] *Programming Code* signifies machine-readable instructions created in the context of a research project, for example, to analyse *Research Data*, to reproduce research findings from a given data set or to perform experiments.

[4] *FAIR Principles* are internationally recognised guidelines to improve the **F**indability, **A**ccessibility, **I**nteroperability, and **R**euse of digital assets. The principles emphasise machine-actionability (for details see Annex A).[15]

---

[7]   RSETHZ 414 (https://rechtssammlung.sp.ethz.ch/Dokumente/414.pdf)
[8]   RSETHZ 134 (https://rechtssammlung.sp.ethz.ch/Dokumente/134.pdf)
[9]   RSETHZ 413 (https://rechtssammlung.sp.ethz.ch/Dokumente/413.pdf)
[10] RSETHZ 203.21 (https://rechtssammlung.sp.ethz.ch/Dokumente/203.21.pdf)
[11] RSETHZ 440.4 (https://rechtssammlung.sp.ethz.ch/Dokumente/440.4.pdf)
[12] For example: Federal Act on Data Protection (SR 235.1) and its Ordinance, Human Research Act (SR 810.30), Clinical Trial Ordinances (SR 810.305); Art. 36 ETH Act (SR 414.110); Bundesgesetz über die Archivierung (SR 152.1); Goods Control Ordinance (SR 946.202.1); Reglement für das Archiv der ETH Zürich (RSETHZ 420.1 https://rechtssammlung.sp.ethz.ch/Dokumente/420.1.pdf).
[13] See General Terms and Conditions for Visitor Stays at ETH Zurich ("GTC Visitor Stay ETH Zurich"; https://rechtssammlung.sp.ethz.ch/Dokumente/515.2.1.pdf)
[14] For the sake of clarity: Intermediate data that are part of the measurement device are not covered by this regulation.
[15] Wilkinson et al., 2016. See also www.go-fair.org/fair-principles/

[5] *Open Research Data* (*ORD*) are *Research Data* that are FAIR and, in addition, publicly available, accessible and reusable at no cost. These guidelines understand *ORD* as an ideal to be strived for that might not be achievable in all cases (for justified exceptions see Art. 6.3).

[6] *Research Data Management (hereinafter "RDM")* is the systematic, transparent, and adequate handling of data across the complete data lifecycle using recognised services and infrastructures. *Research Data Management* planning includes measures to be undertaken at every step of the data lifecycle[16] to ensure transparency and reproducibility of research. Data management planning starts from the creation and/or collection of data, followed by their analysis, publication, storage, sharing, preservation and reuse. Together, all these steps are continuously documented in a Data Management Plan (hereinafter "DMP"). In addition, *RDM* can support fulfilling the *FAIR Principles* by distinguishing data of temporary or permanent relevance according to accepted *Community Standards* and thus providing a reliable basis for informed decisions on data deletion, temporary retention (usually a minimum of 10 years), or unlimited preservation.

[7] *Community Standards* are understood as both explicit, formalised standards and informal, but well established and widely accepted best practices within a community. A community can, e.g. comprise researchers sharing an interest in the same object of research or working with the same methods. Common *Community Standards* address, e.g. identification, citation and reporting of *Research Data* and *Metadata*. They reflect a community's consensus, at a certain point in time, on how reproducible and reusable research should be implemented.

[8] *Data Availability Statements* provide a statement about where data necessary for validating the research findings reported in a result publication can be found — including, where applicable, hyperlinks to publicly archived additional datasets analysed or generated during the research activity.

[9] *Research Group Leaders* are professors or heads of platforms and scientific facilities who hold budget and the overall responsibility for a research group, platform, or facility.

Chapter 2: Research Data Management

Art. 4   Planning of Research Data Management

[1] ETH Zurich expects its researchers to include *RDM* in the planning of their activities.

[2] The establishment of best practices for working with *Research Data* in line with *Community Standards* is required and should be agreed upon on the appropriate organisational level (see article 10 below)

[3] For research projects with temporal boundaries, a DMP (according to Art. 3, Para. 6 above) is expected not later than 6 months after the start of the research project. The DMP is an instrument for researchers themselves, to facilitate the implementation of good *RDM* practice in their projects as well as for planning storage and safeguarding of data according to Art. 7 below.

[4] DMP stipulations issued by funding agencies must be strictly adhered to. The responsibility for this lies with the grantee(s).

---

[16]   Data management Planning starts from the creation and/or collection of data, followed by their analysis, publication, storage, sharing, preservation and reuse. https://ethz.ch/services/en/service/a-to-z/research-data/data-management-planning.html

Art. 5 Data collection and processing

¹ Digital data that are in the process of being collected or processed must be stored such that data can be restored in case of hardware problems or human error. Using storage systems centrally supplied by ETH fulfils this criterion. If such data are too big for being stored redundantly in the available infrastructure, researchers must be able to recreate them using existing code and documentation when required.

² Access rights to the storage system must be specified and regularly updated, e.g. upon the start of a new project. The scientific needs of all researchers working with the data must be considered by the access right system. Unauthorised access must be prevented by suitable technical and organisational measures, such as ETH Zurich's identity and access management system.

³ In order to ensure long-term accessibility and reusability, the use of well-documented, non-proprietary file formats is recommended.

⁴ The structure and the processing steps of all *Research Data* must be digitally documented in order to ensure adherence to the *FAIR principles*. Where documentation includes a lab journal, Electronic Laboratory Notebooks (ELN) are recommended.

Art. 6 Publication of *Research Data* and *Programming Code*

¹ Publication

  a. *Research Data* and *Programming Code* that are considered as directly relevant for a result publication based on *Community Standards* must be published and deposited in a FAIR repository along with rich, openly available Metadata.

    (i) If there are limitations for sharing relevant raw data online because sharing is technically or economically not feasible, FAIR allows for publishing *Metadata* only which contain information on how raw data can be accessed if necessary.

    (ii) In the case of long-range data collection projects, the *Research Data* and *Programming Code* that are relevant for a result publication may be defined as a subset and may be aggregated.

  b. *Research Data* and *Programming Code*[17] are by default published at the same time as the associated results.
  c. All publications of research results must contain a *Data Availability Statement*.

² Repositories

  a. The ETH Research Collection is available as institutional repository for *Research Data* publication. If a recognised ETH-external FAIR repository[18] is used for depositing *Research Data* or *Programming Code*, it is recommended to register a metadata-only item in the ETH Research Collection linking to the dataset.
  b. Commercial platforms which are well-suited for specific purposes (e.g. code development) are usually not considered as FAIR repositories. In case such platforms are used, a copy of the version of *Programming Code* belonging to a result publication must be archived in a FAIR repository.[19]
  c. Commercial repositories should be avoided whenever possible.

---

[17] For the publication of *Programming Code* the Exploitation Guidelines of ETH Zurich may apply in addition: RSETHZ 440.4 (https://rechtssammlung.sp.ethz.ch/Dokumente/440.4.pdf)
[18] www.re3data.org
[19] It is, for example, possible to set up a connection between GitHub and Zenodo to ensure a permanent record of code which makes it citable in the sense of FAIR.

[3] Restrictions

Restrictions on sharing and publishing *Research Data* and *Programming Code* may be justified and, in some cases, necessitated by the following reasons:

a. constraints due to institutional provisions and/or applicable superordinate national and/or international ethical principles and/or legal regulations;
b. ongoing preparation of a patent application;
c. contractual obligations (e.g. to industry partners);
d. intended commercial use of a software linked to a data set;
e. intended use of *Research Data* and *Programming Code* for commercial purposes, e.g. sale.

[4] Licences

a. Researchers must not hand over exclusive rights to the publication and further use of *Research Data* and/or *Programming Code* to repositories.
b. ETH Zurich advocates the precise specification of usage rights for *Research Data* in the form of a Creative Commons Licence or a Public Domain Dedication (CC0). For *Programming Code*, established Open Source licenses[20] must be used.

[5] ETH Zurich supports the publication of *Research Data* and *Programming Code* as independent and citable academic achievements.

Art. 7   Storage and Safeguarding

[1] *Research Data* deposited in a repository according to the *FAIR principles* (see Art. 6) must be retained for extended periods of time, as a rule for a minimum of 10 years, unless legal or funder regulations or Community Standards require other periods. Researchers must make sure that the chosen repository observes the minimum and, where applicable, maximum storage periods.

[2] For data which are not explicitly mentioned in Art. 5 and 6 and which are not deposited in a repository, a minimum storage period of 10 years applies where technically and economically feasible. Such data must be safeguarded at ETH Zurich if they are required for justifying the origin of published data in cases where those data might be challenged (see Art. 11 of the Integrity Guidelines).[21]

[3] For practical and cost reasons, such data can be moved, e.g. to inexpensive storage with ETH Zurich's IT Services together with information on responsible contact persons and required retention periods. If for such data redundant storage in the available infrastructure is technically or economically not feasible, researchers must be able to fully recreate them using existing code and documentation if required.

---

[20]   https://opensource.org/licenses/alphabetical
[21]   RSETHZ 414 (https://rechtssammlung.sp.ethz.ch/Dokumente/414.pdf)

Chapter 3: Responsibilities

Art. 8   Responsibilities of researchers

[1] The departments:

    a. inform new researchers about the RDM Guidelines that are in force at ETH and, if applicable, at the level of the department; and

    b. define processes for the case where a *Research Group Leader* leaves ETH or retires in order to prevent the loss of valuable data and to make sure that archiving periods are observed.

[2] *Research Group Leaders* are responsible for good *RDM* practice in their research groups, platform, or facility. This implies in particular:

    a. ensuring compliance with *Community Standards*, the present ETH's RDM Guidelines and, if applicable, departmental as well as *RDM* guidelines established by other organizational units to which the group, platform or facility relates (e.g. institute, laboratory, non-departmental research unit);

    b. ensuring access to suitable infrastructures (whether provided ETH-internal or – external); and

    c. taking decisions about access rights for unpublished data as well as for group members leaving ETH Zurich.[22]

[3] *Research Group Leaders* are free to define further roles and responsibilities regarding *RDM*.

Art. 9   Responsible administrative units

[1] ETH Zurich provides the basic infrastructure, information and services that facilitate and where possible enable *RDM* according to the present guidelines.

[2] ETH's central IT Services as well as IT support groups in the departments and all organisational units (e.g. institute, laboratory, non-departmental research unit) support researchers with processing and management of *Research Data* and *Programming Code* (see Annex B for details).

[3] The ETH Library supports researchers in the planning of *RDM* as well as in the publication and long-term archiving of data (see Annex B for details).

[4] ETH Zurich may require financial expenditure to be covered by research groups for usage of infrastructure and services and for individual technical solutions provided.

[5] The Office of Research is the main contact point for questions of policy development and supports the departments, organisational units (e.g. institutes, laboratories, non-departmental research units) and research groups, platforms, facilities wishing to compile and edit their own guidelines.

[6] The ETH Ethics Commission is the main point of contact for ethical questions about the publication of *Research Data*. The Legal Office may be involved for specific legal questions.

---

[22]   Cf. Art. 11 Par. 2 ETH Zurich Integrity Guidelines: RSETHZ 414
(https://rechtssammlung.sp.ethz.ch/Dokumente/414.pdf).

Chapter 4: Final provisions

Art. 10   Further regulations

Departments, other organisational units (e.g. institute, laboratory, non-departmental research unit) or research groups, platforms and facilities as well as thematic clusters and competence centres consisting of different Departments and/or organisational units can issue further regulations that complement the present guidelines. These regulations may be more specific than but not in contradiction to the RDM Guidelines. Further regulations respect relevant *Community Standards*. They reflect the perspectives and needs of all academic career levels.

Art. 11   Commencement

The present guidelines enter into force on 1 July 2022.

Zurich, 7 June 2022   ETH ZURICH

On behalf of the Executive Board

The President: Joël Mesot

The Secretary General: Katharina Poiger Ruloff

Annex A: FAIR Principles[23]

    a. **F**indability:
- Research Data and Metadata are assigned a globally unique and persistent identifier.
- Research Data are described with rich Metadata.
- Metadata clearly and explicitly include the identifier of the Research data they describe.
- Research Data and Metadata are registered or indexed in a searchable resource.

    b. **A**ccessibility:
- (Meta)data are retrievable by their identifier using a standardised communications protocol. The protocol is open, free and universally implementable and allows for authentication or authorisation procedures, where necessary.
- Metadata are accessible, even when the Research Data are not openly accessible or no longer available.

    c. **I**nteroperability:
- (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- (Meta)data use vocabularies that follow FAIR principles.
- (Meta)data include qualified references to other (meta)data.

    d. **R**euse:
- (Meta)data are richly described with a plurality of accurate and relevant attributes:
  - (Meta)data are released with a clear and accessible data usage license.
  - (Meta)data are associated with detailed provenance.
  - (Meta)data meet domain-relevant community standards.

Annex B: Support, services and infrastructures for Research Data Management at ETH Zurich

Date of latest revision: 23.05.2022

[1] ETH's central IT Services as well as IT support groups in the departments and institutes support researchers with processing and management of data and code by providing:

    a. redundant storage infrastructure featuring different storage classes and tailored to various user needs;

    b. general infrastructure for professional RDM in running research projects (ETH Research Data Hub, ETH Research Data Node);

    c. individual and / or customized RDM solutions in line with recognized best practices (with cost contribution by the research group);

    d. a versatile and scalable computing infrastructure (HPC and virtual servers); and

---

[23]   According to Wilkinson et al., 2016. See also www.go-fair.org/fair-principles/

    e. a data platform for management and processing of confidential and strictly confidential Research Data (Leonhard Med);

    f. training and consulting including courses on RDM as well as reproducible and efficient data processing (together with ETH Library); and

    g. other relevant IT services.

[2] The ETH Library supports researchers in the planning of RDM as well as in the publication and long-term archiving of data by providing:

    a. training and consulting for RDM including DMP writing (together with IT Services) as well as for related topics such as Open Access;

    b. DMP reviews (together with IT Services);

    c. the ETH Research Collection as a general-purpose repository for publication of Research Data according to the FAIR principles;

    e. a DOI registration desk for persistent identification of digital objects; and

    f. the ETH Data Archive as a long-term archive for data (including those first deposited in the Research Collection).