
Research Data

Definition of terms

Data

- **Any information collected, observed, generated, or produced to verify or reproduce research results**
- documents
- tables
- audio/visual recordings
- images
- photographs
- questionnaires
- interview transcripts
- software
- laboratory journals
- field notes
- samples
- standard data formats are utilized for various types of content, such as: Text: XML, PDF/A, HTML, JSON, TXT, RTF; Images: JPG, GIF, TIFF, PNG; Video: MPEG, AVI, MKV; Audio: WAVE, MP3, FLAC
- a structured collection of research data is referred to as a **dataset**

Metadata

- **Data about the data or information that describes the attributes of a dataset, facilitating its identification, retrieval, and management in the future**
- respondent data
- time and place of data collection
- explanatory notes
- diaries
- model informed consent
- license conditions
- different formats can apply, ranging from free text to structured machine-readable content (some fields or data repositories may have specific requirements)

Data Steward

- **A person who offers support for research data management at the faculty or research team level**
- **Faculty Data Steward**
- Responsible for ensuring that faculty management of research data aligns with university and international standards. It serves as a bridge between the data community and faculty researchers, providing broader support as needed
- Responsibilities include assisting in the development of data management plans, selecting appropriate (meta)data formats, identifying suitable repositories, and recognizing any barriers to research data publication
- Does not fulfil the role of a data analyst, does not perform disciplinary analysis of research data
- **Project Data Steward**
- Responsible for developing a data management plan, storing, securing, backing up and sharing research data, creating metadata and final upload of data into the repository
- It is recommended to allocate 0.1 or 0.2 FTE for the Project Data Steward
- The position is only available for some projects (GA CR, OP JAK)

Research data management

The dataset of research projects has a specific life cycle.

- **Data creation:** data collection and storage, generation of part of the metadata
- **Data processing:** digitization, validation, anonymization and storage of data, generation of additional metadata
- **Data analysis:** interpretation and analysis of data, preparation of publication
- **Data protection:** backup, format migration, documentation
- **Data sharing:** access control, copyright and licensing
- **Data reuse:** new research, partnerships, teaching and learning

When naming files, please follow these guidelines:

- Include the date in the format YYYYMMDD
- Avoid using special characters such as !, @, #, &, %, *, and \$
- Include the initials of research participants
- Assign a unique code for anonymous respondents
- For tabular data: provide a one-line description of each column, starting from cell A1; clearly name and describe each sheet; do not use colors to convey information and refrain from linking cells

Data varies by category and accessibility selected:

- **Public data:** accessible to anyone without restriction
- **Internal data:** only for internal use by a loosely defined group of people (correspondence, minutes of meetings, internal rules and regulations)
- **Discreet data:** meant for the internal use of a specific group of individuals, requiring regulation or protection either by law (GDPR) or by contract/license. This includes economic and personal data of a private nature, such as ID card numbers and birth numbers
- **Sensitive data:** intended strictly for the internal use of a well-defined group of individuals, requiring regulation or special protection, either by law or by contract/license (health data, personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, biometric data processed to identify a person)
- **Open data:** access is usually through electronic data repositories; data must be provided in a manner that allows for further use from both a technical and legal standpoint; access, use, reproduction and dissemination must be free of charge
- **Access with embargo:** the data administrator must specify the public access date for the dataset in the repository
- **Restricted access:** the data administrator will specify the conditions under which access is granted; the data admin shall not charge any fees for granting access
- **Closed access:** enforced for reasons of commercial confidentiality or intellectual property protection; the data may be stored in the repository without public access

It is recommended that access to data should adhere to the FAIR principles, which stand for:

- ****Findable**:** Provide machine-readable metadata along with a unique identifier, such as a DOI (Digital Object Identifier)
- ****Accessible**:** Ensure that the data is openly accessible, preferably through a dedicated repository
- ****Interoperable**:** Use standardized terminology to describe the data to facilitate interoperability
- ****Reusable**:** Implement appropriate licensing to allow for the research data to be reused

None of the FAIR principles requires data to be open or free, but they emphasise the need for clear and transparent conditions for access and reuse. Therefore, FAIR data does not need to be open; it only requires an assigned license.

The principles applied to industry standards follow the principle: ***As open as possible, as closed as necessary***.

Data Management Plan (DMP)

A document that summarises the different phases of research data management during and after a project.

- **What does the DMP contain?**
 - Administrative details (title, research team, provider, abstract)
 - Data collection (methods, formats, volumes, software)
 - Data organisation (quality control, documentation, identifiers)
 - Data storage (security, access, backup)
 - Data disclosure (metadata, licenses, embargoes)
 - Ethical and legal issues of research data
 - Research data management costs (APC fees, project Data Steward, in-kind costs)
- **How to create a DMP?**
 - Creating a DMP can be done in several ways: you can use a traditional word processing application or even write it out by hand on paper
 - However, it is recommended to utilize online tools that guide you through pre-made questions to streamline the process
 - [FAIR Wizard CUNI](#) : a tool facilitating the creation of DMPs, provided for staff and students of Charles University
 - The FAIR Wizard provides a guided experience through various research data management pathways using a tree-based questionnaire format
 - The university's Open Science Support Centre has developed a comprehensive [step-by-step guide](#) to help users familiarize themselves with the tool. Additionally, faculty support has prepared a mock sample of a specific DMP for reference

- **Why create a DMP?**

- Many internal and external funders require a Data Management Plan in various forms, including but not limited to: [GA CR](#) , [TA CR](#) , [Horizon 2020](#) , [Horizon Europe](#) , [PRIMUS](#) , [OP JAK](#) , and the [EXCELES programme](#)
- DMP can also have practical benefits: it helps anticipate potential issues, reduces the risk of data loss, and facilitates data sharing, which ensures continuity in long-term research
- On a broader scale, a data policy should establish standards for the replicability and integrity of research

Data repositories

An online platform for storing, publishing and preserving data, associated metadata and documentation.

- **multi-disciplinary repositories** publish data from any scientific field
- **subject repositories** are preferred
- **institutional repository** is currently under development at the level of Charles University

The selection of an appropriate repository depends on the type of data. Trustworthy repositories are characterized by offering open access, assigning persistent identifiers, utilizing standardized and machine-readable metadata, allowing datasets to be licensed, and obtaining certification.

- **Zenodo**: a general-purpose, open repository developed by CERN and supported by the European Commission through the OpenAIRE project; total files size limit per record is 50GB (max 100 files)
- **National Data Repository**: general-purpose repository, operated by CESNET, pilot mode
- **Harvard Dataverse**: a general repository operated by Harvard University, allowing up to 1 TB of files to be uploaded
- **Czech Social Science Data Archive**: a subject repository, operated by the Institute of Sociology of the CAS, has no maximum file size
- **LINDAT/CLARIAH-CZ**: subject repository for linguistic data and tools, operated by the Institute of Formal and Applied Linguistics at CU, has no maximum file size
- **Re3data.org**: the registry offers an overview of existing international data repositories
- **OpenDOAR**: database of open repositories

Legal aspects of research data

Act No. 130/2002 Coll., [on support for research and development from public funds](#) , defines research data in § 2(2)(o): **"information, excluding scientific publications, in electronic form, which is collected or produced in the course of research or development and is used as evidence in the research or development process or which is generally accepted by the research community as necessary to validate the findings and results of research or development"** („informace, s výjimkou vědeckých publikací, v elektronické podobě, které jsou shromažďovány nebo vytvářeny v průběhu výzkumu nebo vývoje a jsou používány jako důkazy v procesu výzkumu nebo vývoje nebo které jsou obecně akceptovány výzkumnou obcí jako nezbytné k validaci zjištění a výsledků výzkumu nebo vývoje“).

For the purposes of providing financial support under the Act, research data therefore means data in digital form. The University's Centre for Open Science Support has a broader view of research data, including non-digital data.

Act No. 130/2002 Coll. was amended in 2022 ([No. 241/2022 Coll.](#)) in an effort to implement into Czech law **Directive (EU) 2019/1024 of the European Parliament and of the Council**, [Open Data and Reuse of Public Sector Information](#) . The amendment to the Act in question establishes an obligation, for projects supported by public funds, to publish information on how research data is managed. It also imposes:

- **§ 9(1)(l)**: the provision imposes a general obligation to include in the grant agreement a section setting out how research data will be managed by the beneficiary
- **§ 9(1)(m)**: the provision imposes that research results and research data shall not be made public only in justified cases that may include, for example, third-party data, sensitive data on human research participants or trade secrets
- **§ 12(1)**: according to the opinion of the University Centre for Open Science, the paragraph establishes the obligation to publish information about research data in **IS VaVal**, " [their metadata, not the data themselves](#) "
- **§ 12(3)**: the beneficiary is obliged to review at least once a year for five years after the end of the grant whether the justified cases for non-disclosure continue
- the newly inserted **§ 12a(1)**: the beneficiary is obliged to provide research data free of charge on request which „are not protected under the laws governing the protection of the results of copyright, inventive or similar creative activities“ („nejdou chráněna podle zákonů upravujících ochranu výsledků autorské, vynálezové nebo obdobné tvůrčí činnosti“).
- Data obligations under **§ 12a** do not apply to projects announced or supported before **1 September 2022**

There is no standardized system for legal protection of research data. [The Open Science Support Centre](#) offers a more comprehensive overview.

Ethical aspects of research data

Data protection involves multiple areas where dialogue is essential. It relies on an assessment of several key issues:

- Will informed consent from participants be necessary?
- Are there obstacles to making data accessible to other researchers?

- How will discrete and sensitive data be managed to ensure secure storage?
- Who will be responsible for storing the data, and who will have access to it during the project?
- How long will the data be retained after the project concludes?

You can obtain legal advice from the relevant contacts at the [Open Science Support Centre](#) or the [Centre for Knowledge and Technology Transfer](#). The [Committee for Ethical Research](#) within the Faculty of Humanities is responsible for ethical review. The [Information Technology Department](#) offers solutions for suitable network and cloud storage options.

The Committee for Ethical Research was established at the Faculty of Humanities by the **Dean's Measure No. 10/2018**, Statute of the Committee for Ethical Research of the Faculty of Humanities of Charles University, following the University-wide **Rector's Measure No. 74/2017**, [Statute of the Commission for Ethics in Research of Charles University](#). The procedure for faculty acceptance of requests for ethical review is regulated in § 5 of the relevant Dean's Measure.

Additionally, the Faculty Commission, in cooperation with the [Faculty of Humanities Main Filing Room](#), allows for the storage of sensitive data, original signed informed consents, and research data in their secure workplace.

Other (less recommended) types of storage include:

- **portable media** (flash drives, memory cards, CDs)
- **local disks** (computers, laptops)
- **network storage** hosted on CU infrastructure (OneDrive)
- **cloud storage** operated by external entities outside the CU infrastructure (Sharepoint)

Variants of faculty storage and their appropriate use are indicated in the table below:

	Physical storage	Network drives (internal)	Cloud solutions (contract)
Data Categories	public, internal, discrete, sensitive	public, internal, discrete, sensitive	public, internal
Storage capacity	not specified	up to 100 GB per user	up to 5 TB per user
Backup	Main Filing Room	OIT	ÚVT

Faculty and university contacts

Faculty and university support can be contacted if you have any questions.

FHS UK

- **Martin Mišúr**: Data Steward, datasteward@fhs.cuni.cz
- **Miriam Vojtíšková**: Open Science Coordinator, miriam.vojtiskova@fhs.cuni.cz
- **Tomáš Renner**: Secretary of the Research Ethics Committee, tomas.renner@fhs.cuni.cz
- **Roman Sukdolák**: Main Filing Room Administrator, roman.sukdolak@fhs.cuni.cz
- **Alena Matuszková**: Library Director, alena.matuszkova@fhs.cuni.cz

UK

- Consultation regarding copyright issues, openlaw@cuni.cz
- Data Protection Officer, gdpr@cuni.cz
- Commercialisation and Intellectual Property, research.data@cuip.cz
- Technical (ICT) support, openict@cuni.cz
- Computer Science Centre, office@uvf.cuni.cz
- Open Science Support Centre, researchdata@cuni.cz

Important documents

[Act No. 130/2002 Coll., on support for research and development from public funds](#) (in Czech)

[Act No. 241/2022 Coll., the Act amending Act No. 106/1999 Coll., on free access to information, as amended, Act No. 123/1998 Coll., on the right to information on the environment, as amended, and Act No. 130/2002 Coll., on support for research, experimental development and innovation from public funds and on amending certain related acts \(Act on support for research, experimental development and innovation\), as amended](#) (in Czech)

[Directive \(EU\) 2019/1024 of the European Parliament and of the Council, Open data and re-use of public sector information](#)

[Dean's Measure No. 10/2018, Statute of the Research Ethics Committee of the Faculty of Humanities, Charles University](#) (in Czech)

[Rector's Measure No. 74/2017, Statute of the Research Ethics Committee of Charles University](#) (in Czech)

Useful links

[The Open Science Support Centre](#)

FAIR Wizard
How to make your data FAIR
GA CR and Open Science
How to handle data safely