**Research Data Management Policy at** **Faculty of Law, Economics, and Governance (LEG)
This document has been established by the BOS meeting 11 september 2019.**

**Version 13 augustus ( Adapted after advice from LISM Stijn Hoogervorst and Research Data Management Consultant Iza Witkowska)**

**Premises**

The goal of this policy document is to introduce the general requirements of data management and specific requirements and guidelines for researchers at LEG. All researchers are obliged to take good care of their research data through data management, which allows for proper storage and handling of research data. Special attention is required for research with privacy-sensitive and / or security-sensitive data.

**Law and regulations**

In the last few years, a number of laws and regulations have been introduced in the Netherlands and the EU that have a material impact on how researchers should handle research data. These include the Utrecht University [**Policy Framework**](https://www.uu.nl/sites/default/files/university_policy_framework_for_research_data_utrecht_university_-_january_2016.pdf) for Research Data, the [**Netherlands Code of Conduct for Research Integrity**](http://www.vsnu.nl/files/documents/Netherlands%20Code%20of%20Conduct%20for%20Research%20Integrity%202018.pdf)**,** and the requirements for privacy- and security-sensitive data as stated in the[**General Data Protection Regulation**](https://eur-lex.europa.eu/eli/reg/2016/679/oj) (GDPR). Additionally, UU and LEG aim to operate at the forefront of [**Open Science**](https://open-science.sites.uu.nl/wp-content/uploads/sites/321/2018/09/Utrecht-University-Open-Science-programme-draft-ver.-0.9.pdf) and towards FAIR principles (Findable, Accessible, Interoperable, Reproducible). These two approaches will increase visibility and impact of the researchers and their research.

**What is FAIR?**

Please, note that FAIR doesn’t necessarily equal Open Science or publishing of data. For the purpose of this document, the FAIR principles are defined as follows:

* **Findable** means making data and metadata discoverable for others preferably within computer systems.
* **Accessible** stands for storing and preserving data in a way that they can be easily accessed and/or downloaded with well-defined access conditions, whether at the level of metadata, or at the level of the actual data content.
* **Interoperable** means that data and metadata should conform to recognized formats and standards to allow them to be combined and exchanged.
* **Reusable** stands for making sure that data is ready to be used, replicated and/or combined in future research.

**What is a research project?**

For the purpose of this document, the definition of a research project includes projects gathering and using empirical data within the UU, whether or not it aims at publishing the work in a book, book chapter, academic journal, data journal, report and advice or public (data) repository. This applies to researchers and research master students.

**What is research data?**

Research data includes any source or information material used within research that leads to the results of the research project. It can, therefore, be qualitative and / or quantitative in nature and may include audio files, video files, interview transcripts, data spreadsheets, literature references, text documents, books, programming- / software- / analysis scripts, theoretical modelling scripts, etc. Any material or source that is required to reproduce the research falls under research data and data management.

Two main types of research data can be identified: primary and secondary research data. Primary research data are data generated by the research group itself by means of surveys, interviews, experiments, field research, simulations, theoretical modelling, etc.. Secondary research data are data collected by a third party, including institutions that specialize in data collection (e.g., CBS, ECB, Thomson Reuters).

**General Requirements for proper data management**

Every research project carried out by LEG researchers should respect minimum standards in terms of data management and implies that:

* Researchers write a [Data Management Plan](https://www.uu.nl/en/research/research-data-management/guides/data-management-planning) (DMP) for the project before the project starts. The DMP helps researchers plan storage needs, data security issues, data archival requirements, and privacy risks. Researchers are encouraged to use the [UU template](https://www.uu.nl/en/files/uudmptemplate20180518docx) for the DMP, which is also available via the web-based tool [DMPonline](https://www.uu.nl/en/research/research-data-management/tools-services/tool-to-create-your-dmp-online). The DMPonline tool guides a researcher through the DMP sections with explanations and general examples.
* [Yoda](https://www.uu.nl/en/research/research-data-management/tools-services/tools-for-storing-and-managing-data/data-management-solutions) is the LEG default research data storage and management solution developed within the UU. Researchers and research master students who work with sensitive data should use Yoda for their data storage. The costs of Yoda will be absorbed by the faculty. For other research projects, the use of alternative solutions (i.e. One-Drive, O- or U-drive) is allowed but should be motivated. Data should be stored digitally whenever this is possible. Local storage of data for analysis purposes is allowed with the necessary precautions (see Privacy and Security).
* Proper data management is needed also to make sure that:
	+ a research project is not jeopardized simply because one of the members abandons the research team;
	+ every LEG researcher involved in a specific project is able to verify which version of a dataset and which version of the code written to analyze data were used to produce the results presented in a (published) paper;
	+ all research data is securely stored and backed-up so that device theft or breakdown (computers, USB devices, etc.) does not result in loss of the data;
	+ research data is still findable, accessible, interoperable, and reproducible (FAIR) for10 years time.

**Data management guidelines for primary and secondary data**

***Primary research data****.*

* The research data needs to be organized in a proper [folder and file naming structure](https://www.uu.nl/en/research/research-data-management/guides/storing-and-preserving-data) which allows the identification of the type of information enclosed in the specified folder or file. A general folder structure could include the folders: Data, Analysis, Literature, and Results. The folder structure will differ per project.
* Raw data should be safely stored in a unique master version and stored in a separate folder with restricted access as soon as they are collected and validated by the research group. Raw data is the research data initially gathered before combined, reshaped, adjusted, etc., for actual analyses.
* Subsequent official master versions of the dataset (e.g., the merged and reshaped dataset putting together primary raw data to perform analyses) should be created and stored on Yoda in correspondence to specific milestones. Versioning can be used to officially identify different versions of a dataset.
* Primary research data can often be privacy- and security-sensitive. When this is the case, researchers should comply with the security requirements illustrated below (at Privacy and Security).
* When a storage solution other than Yoda has been agreed upon that is not under UU management, the researcher is responsible to back-up the data as recommended by [RDM support](https://www.uu.nl/en/research/research-data-management/guides/storing-and-preserving-data).
* Researchers carry the full responsibility when sharing their data with external contacts. When data is being shared before publication or high sensitive information is shared with external contacts, a [data transfer or data processing](https://www.uu.nl/en/research/research-data-management/guides/legal-instruments-and-agreements) (or both) agreement needs to be set-up and signed by both parties.

***Secondary research data***.

* If the third-party data owner allows for archiving of the data, this should be the preferred option. Archiving the data is not required when this is not allowed and/or when the same data can in principle be recovered at any point in time by someone (within and outside the UU) with access to the original source (e.g., a valid account).
* When data is allowed to be stored, the same principles as with primary research data apply.
* Even if the actual data cannot be stored, researchers should always store the code used to analyze the data and the code used to extract the data (or a data description detailed enough to allow in principle a recreation of the analyzed dataset in the same way).

**Privacy & Security**

***Personal data*** is any information relating to an identified or identifiable natural person (data subject). This means information directly about someone or traceable to this person. E.g. names, ID number, location, IP-address, physical characteristics, economical characteristics, cultural characteristics, etc.

 ***Personal sensitive data*** is data consisting of racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, genetic data, biometric data, data concerning health or data concerning a natural person's sex life or sexual orientation. Processing of this personal (sensitive) data shall be prohibited, except for specific purposes and under certain circumstances. Research is one the possible exceptions that allows processing of special personal data. E.g. Personel data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person’s sex life or sexual orientation.

Personal (sensitive) data, regardless of whether they are used for a research project as defined in this document or not, are subjected to the GDPR. Researchers themselves are responsible for properly handling personal (sensitive) data during and after the research project. This type of data requires extra care and [the following six principles of the GDPR apply](https://www.uu.nl/en/research/research-data-management/guides/handling-personal-data): lawfulness, purpose limitation, data minimization, data accuracy, storage limitation and integrity & confidentiality.

* Whenever LEG researchers are about to process personal (sensitieve) data for their research need to make sure that their initial DMP carefully covers how they intend to handle the personal (sensitieve) data. In particular, a Data ProtectionIimpactAassessment (DPIA) should be carried out. This requirement is not required for cases where the third-party owner of the data maintains full control over them at every step of the process and ensures the safe and secure storage of the data. CBS data are a typical example of this.
* Researchers and research master students who work with personal (sensitive) data should use Yoda for data storage.
* Researchers are strongly advised against keeping copies of personal (sensitieve) data stored locally (including portable devices) or on cloud storage systems different from Yoda. If storing data locally on your PC is needed at a certain point (e.g., to analyze the data), researchers should make use of a type of encryption. Researchers can make use of the UU available software [BoxCryptor](https://www.uu.nl/en/research/research-data-management/tools-services/tools-for-storing-and-managing-data/storage-solutions#encrypt).
* The researchers should understand that third party storage and sharing solution may not have their servers within the EU and, therefore, is not (or may not be) GDPR compliant (allowing data to possibly be leaked or unwittingly shared).
* Anonymize the data as soon as possible in the research process and only store anonymized data. The GDPR does not apply to anonymized data.
* If personal (sensitive) data must be retained, for example because potentially needed in the future, a key-file mapping subjects to pseudonymized data should be preserved safely and in a separate folder with restricted access.
* Researchers should not travel with unencrypted personal (sensitive) data. Researchers should be aware that encrypted data are illegal in a number of countries (e.g., China) and that some countries may require you to give access to devices and data at for example at airports (such as the United States).
* It is not allowed to share personal (sensitive) data with contacts / researchers who are not (joint) controllers of data unless a data transfer agreement or data processing agreement (or both) is signed and present.
* When personal (sensitive) data is collected, informed consents are necessary of the participating subjects. The degree of information provided and type of consent that needs to be given, depends on the type of information gathered. However, a data transfer and processing agreement should be present if / in case of??.

 **Data ownership and reuse rights**

* [Utrecht University retains ownership rights of the research data created by researchers of Utrecht University, unless otherwise agreed upon.](https://www.uu.nl/sites/default/files/university_policy_framework_for_research_data_utrecht_university_-_january_2016.pdf) The researcher has the primary responsibility for management and quality of the data. The researcher is allowed to hold a copy of the research data, unless otherwise agreed upon.
* A researcher should clarify who the copyright holder of the datasets is, especially when existing data is used or when there is a collaboration with external parties. Copyrighted output from research could include spreadsheets (and other forms of originally selected and organised data), publications, books, reports and computer programs. Consult the [Copyright Information Office](https://www.uu.nl/en/organisation/copyright-information-office) for more information on copyright on publications and books.

 **Publishing**

The Open Science Program of Utrecht University Research promotes sharing data and making research data accessible for other rearchers and third parties. Research data can be published openly or with restricted access, or metadata can be published without publishing the actual research data. However, researchers will retain full discretion on whether to make their work publicly available, available upon request, or not available. It is advised to always describe metadata of the research and make it publicly available (only access to the metadata, not the data itself) to improve findability and impact of the research and improve career statistics. Please remember that:

* Personal research data can never be published in its original form. Personal data needs to be anonymized or aggregated before data can be published.
* Publishing can be done easily in Yoda. Upon publication in Yoda, research data and /or metadata will receive a Digital Object Identifier (DOI) and consequently be findable by other researchers through search machines and data search engines (such as [DataCite](https://search.datacite.org/)).

**Post-publication data storage**

Long-term data archiving should be done digitally.

* **Data format**. Research data should be stored in preferred [data formats](https://dans.knaw.nl/en/deposit/information-about-depositing-data/before-depositing/file-formats) which guarantee long-term preservation, meaning that the files can be opened and used in the future. Data formats associated to commonly used software do not necessarily guarantee long-term preservation due to frequent changes in software and format updates.
* **Storage period.** Research data should be stored after the research is published or the research project has ended for a period of at least 10 years. This is also the default option in Yoda. Research data acquired from third parties which do not allow archiving of the data, adhere to the policies of the third party.
* **Metadata.** Proper metadata should be associated to the final data stored in Yoda. The minimal requirement is the Yoda metadata scheme which complies with the Datacite V4 standard. Metadata can be used to indicate that the data exist without actually sharing the data.

**PURE.** If possible, (meta)datasets should get linked to the publication and registered in PURE. To link publications in PURE, an [ORCID](https://orcid.org/) should be created. An ORCID provides a persistent digital identifier that distinguishes you from every other researcher and, through integration in key research workflows such as manuscript and grant submission, supports automated linkages between you and your professional activities ensuring that your work is recognized. At this moment it is currently possible to register publications in PURE. The registration of data sets in PURE can now be done manually. With the link between Yoda and Pure, this process will soon become easier.

**Data management support**

Extensive support is available at any step of data management. This support can range from practical help (helping with folder structure, completing metadata, filling in DMP, etc.) to advising (on GDPR guidelines, use of software, data formats, etc.). If in doubt about data management topics/issues and in need of clarification, researchers are encouraged to contact:

 ***Support at the Leg Faculty level***

- The Research Support Office (RSO) at LEG at rso.rebo@uu.nl

- The Local Information and Security Manager (LISM) at s.j.hoogervorst@uu.nl

- The Facultaire Ethische Toetsingscommissie (FETC) at FETC.REBO@uu.nl

***Support at the university level***

- Research Data Management (RDM) support at info.rdm@uu.nlinfo.rdm@uu.nl

- Website: [Research Data Management Support](https://www.uu.nl/en/research/research-data-management)