DMP and/or GDPR Registration

Introduction

Version: UT-DMP-GDPR_v3-13

Welcome to the UT tool for writing your data management plan (DMP) and the notification of processing of personal data in research in compliance with the General Data Protection Regulation (GDPR registration).

This DMP form has a generally accepted structure which complies to the policy of funders like NWO and ZonMw. The EU allows you to deliver a DMP based on this form as well.

When filling in the answers, please check data policies and guidelines of your research group, department, or faculty and/or the UT research data management policy. Also the website of these organization units may contain relevant information.

This DMP form can or must be reviewed. When your draft version is ready for review press "Submit for Review" to start/continue this review process. An e-mail message is sent to the Reviewer. The same holds for GDPR registration.

Choose your form

What do you want to do?

- (●) DMP
- ( ) DMP with GDPR Registration
- ( ) GDPR Registration

*When you will **process personal data*** in terms of the European General Data Protection Regulation (GDPR), choose option 2 or 3, the latter for instance in case you have already a DMP. UT bachelor or master students are not obliged to make a DMP but in case of processing personal data should do a GDPR registration (option 3). More information.*

A workflow of GDPR and research data (appropriate use of personal data in scientific research according to the GDPR) can be found on this website.

** **processing** means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated 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.*

***personal data** means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person;*

General

G1. Short title of the research

Risk Management in Digital Finance

G2. Full title of the research

Risk Management in Digital Finance

G3a. Name of the researcher

Baals, L.J.

If this is not the right name, please change it (use the same format)

G3b. Email Address of the researcher

l.j.baals@utwente.nl

If this is not the right email address, please change it

G4. Name of the supervisor or principal investigator (PI)

Osterrieder, J. ; Mes, M.

Please fill in as the following examples: Einstein, A. ; Tocqueville, A.C.H. de ; Waals, J.D. van der

G5. Name of research group/department

BMS-IEBIS

G6. Faculty

Behavioural, Management and Social sciences (BMS)

Please, type the abbreviation of your research group in format:
***Faculty-Research Group**; (e.g. EEMCS-HMI, ITC-ESA, BMS-HTSR)*

G7. Description of the research

Doctoral thesis research on risk management and mitigation in digital finance, with a particular focus on detecting and assessing risks in digital assets (e.g., NFTs and DeFi protocols) and peer-to-peer (P2P) lending markets. The research aims to identify pricing inefficiencies and systemic vulnerabilities using quantitative modeling. Methodologically, the study combines empirical data analysis, network-based modeling, and supervised machine learning techniques. The project will primarily use secondary financial and transactional datasets sourced from P2P platforms and blockchain data aggregators. No laboratory experiments will be conducted; however, algorithmic simulations and statistical modeling will be employed to generate derived datasets and evaluate risk patterns.

G8. Funding body

☐ NWO ☐ EU ☐ ZonMw ☑ Other

In case of NWO, EU and ZonMw, this form is in accordance with their policy. Please note that the ERC Data Management Plan is not a part of the Ethics Review. It is the responsibility of the Principal Investigator to inform the ERCEA Ethics Team of any ethics issues/concerns regarding the collection, processing, sharing and storage of data in relation to the project. The Principal investigator can also be asked to submit an Ethics Data Management Plan (Ethics DMP). For more information you can contact the UT Strategic Business Development office

Please, specify other funder(s)

Swiss National Science Foundation

G9. Grant number

100018E_205487

G10. Partner organization(s) and/or third party(ies)

NA

G11. Project start date

01/12/2022

Please fill in the (provisional) start date of the research project.

G12. Project end date

30/06/2026

Please fill in the (provisional) end date of the research project.

G13. Date written

31/05/2025

Choose the date you have finished this version of the DMP.

G14. Name of the researcher(s), their institution/group and their roles and responsibilities in data management

Name	Institution/group	Role in RDM	RDM Responsibility

Lennart John Baals	UT	Executing main researcher	Full responsibility for writing, updating, and implementing the Data Management Plan (DMP), including data collection, documentation, storage, security, selection, preservation, and ensuring availability for reuse.
Jörg Osterrieder	UT	Co-Promoter	Supervision, review of the DMP, funding acquisition, and shared responsibility for data selection and long-term preservation in coordination with the executing researcher.
Martijn Mes	UT	Promoter	Supervision, coordination of research activities, review of the DMP, and shared responsibility for overseeing its implementation and long-term data archiving.

Name: Please fill in as the following examples: Einstein, A. ; Tocqueville, A.C.H. de ; Waals, J.D. van der

- *Role in RDM: Besides you, as the main researcher, also mention others who will play a role and have responsibilities in research data management, for example the Principal Investigator (PI), the supervisors or the lab technician.*
- *RDM Responsibility: As the main researcher you are responsible for at least writing, updating and implementing (data collection, data documentation, data storage, data security, data preservation and making data (publicly) available) the DMP. Your supervisors are responsible for at least reviewing the DMP and monitor its implementation, but most likely also long-term data preservation and archiving.*

Furthermore, indicate for each specific part of the DMP (e.g. data collection, data documentation, ...) who is responsible. In case of a collaborative research project, think of the RDM coordination role and responsibility. Check both project organization and research group policies.

G15. Laws, policies, contracts and agreements to comply with

- ☑ RDM policy UT / faculty / research group
- ☐ General Data Protection Regulation (GDPR)
- ☐ UT Research Ethics Policy
- ☐ Medical Research Involving Human Subjects Act (WMO)
- ☐ Contract (funding bodies, partners, third parties)
- ☑ Other

Please, specify other laws, policies, contracts and agreements to comply with

The SNSF requires a Data Management Plan (DMP) for funded projects in line with its Open Research Data policy. All data underlying publications must be shared in a FAIR-compliant manner unless legal, ethical, or confidentiality constraints apply.

In principle the handling of data in your research must comply with RDM policies of the UT, faculty and group. However, additional options may apply. When you process personal data your research must comply with the GDPR. Third-party contracts may deal with ownership, responsibilities and requirements about the handling of research data.

G16. Estimated total costs (€) involved in data management

€ 0,00

You can not enter a value here, as this will be automatically filled in as a total of the costs you add in the following sections.

G17. Name of reviewer

Deniece Nazareth and Minsi Li (BMS)

Date of review

The person in your faculty who can review your DMP.

Data collection

The answers to the following questions should be entered in the table below.

1. **Add for each set of data you will collect or generate a descriptive title and the type of data. Types of date are observational data, experiment data, simulation data and/or derived or compiled data. Also mention materials, such as lab notebooks, field diaries, informed consent, or algorithms, scripts, etc. Add physical data or materials, like samples, as well. Make separate items for personal data.** *Indicate whether it concerns secondary data (pre-existing data collected or generated by other people or organizations)*
2. **Which form will these types of data have (e.g. text, numbers, tabular data, survey data, models, software, audio, video, physical samples)?**
3. **Which file format will the types of data have (e.g. pdf, xls, doc, txt, rdf)?**
4. **Which software or tools are needed to create, process and/or visualize these types of data?**
5. *Personal data is set to "No" and can not be changed, because you have chosen not to use the "GDPR Registration" (see first question of this form).*

DC1. Descriptive title / type of data	DC2. Form	DC3. File format	DC4. Software/tools	DC5. Personal data
nft_prices_raw_2023-06_v1.1_ [secondary]]	tabular data	.csv	Python, R	○ Yes ◉ No
p2p_loans_raw_2020_v1.1_ [secondary]	tabular data	.feather	Python, R	○ Yes ◉ No
Bubble_detection_code_v1.1	Data cleaning file	.qmd	R	○ Yes ◉ No
Bubble_detection_code_v2.1	Code execution file	.qmd	R	○ Yes ◉ No
P2P_loan_analysis_code_v1.1	Data cleaning file	.pynb	Python	○ Yes ◉ No
P2P_loan_analysis_code_v2.2	Code execution file	.qmd	R	○ Yes ◉ No
P2P_loan_analysis_code_v3.3	Code execution file	.pynb	Python	○ Yes ◉ No
P2P_loan_analysis_output_tables	Derived data	.csv	Python, R	○ Yes ◉ No
P2P_loan_analysis_plots	Derived data	.png	Python, R	○ Yes ◉ No
P2P_loan_analysis_model_summaries	Derived data	.pdf	Python, R	○ Yes ◉ No
Bubble_detection_output_tables	Derived data	.csv	Python, R	○ Yes ◉ No
Bubble_detection_plots	Derived data	.png	Python, R	○ Yes ◉ No
Bubble_detection_model_summaries	Derived data	.pdf	Python, R	○ Yes ◉ No

DC6. In case you use secondary data which source will be used? *Be sure that also secondary data are added to the data collection table*

coinmarketcap.com
nonfungible.com
OpenSea.io
Bondora Lending (bondora.com)

Secondary data can be used from very different sources, such as data available in your own research group, from databases managed and offered by (inter)national institutes, e.g. statistical offices; commercial parties or clinical data from hospitals.

DC7. Are there any conditions or restrictions for storage, access, archiving, (re)use and sharing of the data (primary and/or secondary) imposed by a third party?

○ Yes ◉ No

This question is about rights and control regarding the research data. Although legally incorrect, this is often referred to as data ownership. In principle, when a set of research data is protected by intellectual property rights (e.g. copyrights or database law), such rights shall vest in the UT.

DC8. What will be the estimated total costs (€) involved in the collection, generation and/or use of data?

€ 0,00

Think of costs for acquiring, processing or analyzing the data or for getting informed consent. Use cost estimations which possibly are available in the research project budget. For more information, see this guide for estimating research data management costs. For costs of UT RDM services see this overview.

Data documentation

DD1. What standards will be used for metadata? If there is no discipline-specific metadata standard available for the data, you could use a general metadata standard (e.g. Dublin Core), and list additional metadata that will be used.

All datasets will be documented using structured metadata following the Dublin Core standard to ensure findability and reuse. A README file will accompany each dataset, providing information on content, structure, collection methods, and data formats. A data dictionary will describe all variables (e.g., definitions, units, types, and allowed values) to support interpretation. Version control will be applied to both data and scripts, with consistent naming conventions and a changelog. Preferred open file formats (e.g., .csv, .txt) and standardized terminology will be used throughout to ensure interoperability and long-term usability.

Data documentation should always contain basic details, called metadata, to make the data findable, including who created or contributed to the data, its title, date of creation and under what conditions it can be accessed. Metadata standards are for instance Dublin Core, DDI, TEI, EML, MARC or CMDI. Data repositories, e.g. 4TU.ResearchData or DANS, use metadata standards as well.

DD2. How will the data (primary and/or secondary) be documented during your research and for long-term preservation?

Throughout the project, all data will be documented using structured README files, a data dictionary, and versioned processing scripts. These will include information on data sources, methodologies, preprocessing steps, variable definitions, units of measurement, assumptions, and data formats. Metadata will be embedded where applicable (e.g., file headers or R/Python script comments). Secondary data (e.g., from NFT/DeFi platforms or P2P lending marketplaces) will include full provenance documentation, source URLs, access dates, and licensing terms, recorded in a separate provenance log. For long-term preservation, all documentation will be bundled with the datasets in preferred open formats and stored following the FAIR principles.

Data documentation should also include details on the methodology used, analytical and procedural information, definitions of variables, vocabularies, units of measurement, any assumptions made, and the format and file type of the data. Consider how you will capture this information and where it will be recorded, for example in a database with links to each item, a README file (guidance / template), file headers, code books, or lab notebooks. In case of secondary data, explain how data provenance will be documented.

DD3a. What directory-naming convention will be used?

I will use a consistent, hierarchical directory-naming convention based on project components and data status. The top-level structure will include folders such as raw_data/, processed_data/, analysis/, results/, and documentation/. Folder names will include relevant identifiers where needed (e.g., p2p_loans_2020_raw/, nft_timeseries_clean/) to distinguish data sources and stages. This structure will be aligned with any conventions agreed upon in my research group and maintained throughout the project for reproducibility and clarity.

There is no single recommended way to name your folders/directories, but you should name your files consistently. Check the file-naming system with your colleagues in the group. More practical information can be found here.

DD3b. What file-naming convention will be used?

Files will be named consistently using lowercase letters and underscores, incorporating key elements such as data source, content type, version, and date (e.g., p2p_loans_raw_2020_v1.1.feather, nft_prices_clean_2023-06.csv). Scripts and outputs will follow a similar structure (e.g., model_results_logit_v2.R). File names will avoid spaces and special characters, and versioning will be indicated using v1.1, v2.1, etc. This convention supports clarity, automation, and reproducibility, and aligns with practices in my research group.

There is no single recommended way to name your files, but you should name your files consistently. Check the file-naming system with your colleagues in the group. More practical information can be found here.

DD4. How will you handle version control (tools, software, procedures) to maintain all changes that are made to the data?

Version control will be handled through a combination of manual versioning. Raw data will be stored read-only and never overwritten. Processed and analysis-ready datasets will include version identifiers in their filenames (e.g., p2p_loans_clean_v1.1.feather, nft_prices_daily_v2.1.csv). For code, documentation, and data processing scripts, markdown code files will be used to track all changes, with commits annotated clearly. A changelog in Markdown format will document major updates to data processing steps, file structures, and metadata. This setup ensures transparency, reproducibility, and traceability throughout the research lifecycle.

Version-control mechanisms should be established and documented before any data is collected or generated. Read more about version control and specific tools.

DD5. What will be the estimated total costs (€) involved in the documentation of data?

€ 0,00

In case of labour intensive metadata creation and documenting data files, hiring of personnel may be needed. Use cost estimations which possibly are available in the research project budget. For more information, see this guide for estimating research data management costs. For costs of UT RDM services see this overview.

Data storage

Use the table below to answer the the following questions for each type of data you collect or generate. If you want to add or change an item, please return to the table Data Collection.

1. **What storage medium will you use for the master files of the data?**
2. **In case of non-UT central storage media, what will be the backup frequency and location?**

It is UT policy to store the original files (master files) from which you make work copies of the research data on UT network file servers, such as the 'Home directory' or the 'Project and organization directory'. Data files on UT network servers are stored in the UT data centre and backed up daily. Have a look at the UT options for storing your research and the backup procedure. For questions and more information, contact the ICT Account manager in your faculty.
If you are, by third-party agreement, bound to store the master files on another medium, be aware of adequate backup frequency and location. Storing master files of data on laptops, stand-alone hard drives or portable storage devices such as USB-sticks, is not in compliance with the UT data policy.

Type of data	DS1. Storage Medium	DS2. Backup Frequency	Backup location
nft_prices_raw_2023-06_v1.1_ [secondary]]	○ UT Network Storage ◉ Non UT Network Storage *Cloud services such as Surfdrive or OneDrive (with @utwente.nl login) are non UT network storage* Please specify other storage: OneDrive of central research group	Monthly	Cloud (OneDrive)
p2p_loans_raw_2020_v1.1_ [secondary]	○ UT Network Storage ◉ Non UT Network Storage *Cloud services such as Surfdrive or OneDrive (with @utwente.nl login) are non UT network storage* Please specify other storage: OneDrive of central research group	Monthly	Cloud (OneDrive)
Bubble_detection_code_v1.1	○ UT Network Storage ◉ Non UT Network Storage *Cloud services such as Surfdrive or OneDrive (with @utwente.nl login) are non UT network storage* Please specify other storage: OneDrive of central research group	Monthly	Cloud (OneDrive)
Bubble_detection_code_v2.1	○ UT Network Storage ◉ Non UT Network Storage *Cloud services such as Surfdrive or OneDrive (with @utwente.nl login) are non UT network storage* Please specify other storage: OneDrive of central research group	Monthly	Cloud (OneDrive)
P2P_loan_analysis_code_v1.1	○ UT Network Storage ◉ Non UT Network Storage *Cloud services such as Surfdrive or OneDrive (with @utwente.nl login) are non UT network storage* Please specify other storage: OneDrive of central research group	Monthly	Cloud (OneDrive)
P2P_loan_analysis_code_v2.2	○ UT Network Storage ◉ Non UT Network Storage *Cloud services such as Surfdrive or OneDrive (with @utwente.nl login) are non UT network storage* Please specify other storage: OneDrive of central research group	Monthly	Cloud (OneDrive)
P2P_loan_analysis_code_v3.3	○ UT Network Storage ◉ Non UT Network Storage *Cloud services such as Surfdrive or OneDrive (with @utwente.nl login) are non UT network storage* Please specify other storage: OneDrive of central research group	Monthly	Cloud (OneDrive)

Type of data	DS1. Storage Medium	DS2. Backup Frequency	Backup location
P2P_loan_analysis_output_tables	○ UT Network Storage ◉ Non UT Network Storage *Cloud services such as Surfdrive or OneDrive (with @utwente.nl login) are non UT network storage* Please specify other storage: OneDrive of central research group	Monthly	Cloud (OneDrive)
P2P_loan_analysis_plots	○ UT Network Storage ◉ Non UT Network Storage *Cloud services such as Surfdrive or OneDrive (with @utwente.nl login) are non UT network storage* Please specify other storage: OneDrive of central research group	Monthly	Cloud (OneDrive)
P2P_loan_analysis_model_summaries	○ UT Network Storage ◉ Non UT Network Storage *Cloud services such as Surfdrive or OneDrive (with @utwente.nl login) are non UT network storage* Please specify other storage: OneDrive of central research group	Monthly	Cloud (OneDrive)
Bubble_detection_output_tables	○ UT Network Storage ◉ Non UT Network Storage *Cloud services such as Surfdrive or OneDrive (with @utwente.nl login) are non UT network storage* Please specify other storage: OneDrive of central research group	Monthly	Cloud (OneDrive)
Bubble_detection_plots	○ UT Network Storage ◉ Non UT Network Storage *Cloud services such as Surfdrive or OneDrive (with @utwente.nl login) are non UT network storage* Please specify other storage: OneDrive of central research group	Monthly	Cloud (OneDrive)
Bubble_detection_model_summaries	○ UT Network Storage ◉ Non UT Network Storage *Cloud services such as Surfdrive or OneDrive (with @utwente.nl login) are non UT network storage* Please specify other storage: OneDrive of central research group	Monthly	Cloud (OneDrive)

DS3. In case of storing master files (also) on other media than the UT Group/Project drive (P-drive or UniShare), what are the reasons of this?

The time series data for NFT and DeFi assets is stored (in addition to UT-managed environments) on a secure OneDrive cloud folder hosted and maintained by my research group. This setup facilitates collaborative access, cross-institutional work, and version synchronisation with external partners. Although OneDrive is not part of UT's official network storage, weekly backups are scheduled, and access is limited to project team members. Upon finalising the data, the master files will be migrated to the UT Project directory to comply with long-term preservation policies.

If (master files of) data must be (also) stored on servers of external parties, please refer to the contract or agreement.

DS4. If other storage media for work copies and/or temporary storage of the data files will be used, specify this here.

Work copies and temporary data files may occasionally be stored on local devices (e.g., laptop or external SSD) and in cloud environments such as OneDrive, primarily to enable offline work, collaboration, and syncing across devices. These copies are always derived from the master file and are used for short-term processing only. Confidential or sensitive data will be pseudonymised if stored outside UT-managed systems, and all devices are encrypted and password-protected. Temporary copies are regularly deleted or synchronised with the master version stored in the UT infrastructure.

Although not recommended, copies of data files can be kept on remote, cloud and/or portable storage. Please, keep in mind confidentiality and security requirements.

DS5. What are the estimated total costs (€) for storage of the data, both on UT network servers and other locations?

€ 0,00

When you need to store more than 10 GB of data on UT network servers, ask the ICT Account manager in your faculty for information about costs.

Data security

DSec1. In case of an incident, explain how the data will be recovered.

In the event of an incident, data recovery will be performed using the most recent backup stored in our research group's OneDrive cloud environment, which is backed up weekly. OneDrive provides version history and file recovery features, allowing for quick restoration of previous versions or deleted files. The time between the incident and the last valid backup is typically no longer than seven days. Recovery access is managed by the research group lead, and in case of technical issues, support from the institutional IT services (e.g. LISA) may be requested. If a breach involves personal data, the incident will be reported and handled according to the UT Data Breach Procedure within the required 72-hour window.

Refer to data-recovery procedures of your backup services or facilities, such as how long it will take before you can use the data again, and the time between the incident and the last valid backup. The UT backup service for employees gives you more information. When applicable, refer to specific security measures or protocols.

DSec2. Do (part of) the collected or generated *NON-personal* data have a confidential character?

○ Yes ◉ No

Data may be confidential, due to political sensitivity, trade interests (including patentability) or public security

DSec3. During the research, who decide(s) about access to and (re)use of the data?

During the research, access to and reuse of the data will be jointly determined by the principal researcher (myself) and my supervisor. This ensures appropriate oversight in line with project goals and data sensitivity. In case of uncertainty or external requests, decisions will align with institutional data management policies and, where applicable, funder or ethical guidelines.

For instance, the researcher, the principal Investigator, the supervisor or head of the research group. Sometimes, institutions (e.g., the lab, the university or even the funder) should control access to the data.

DSec6. After the research, who decide(s) about access to and (re)use of the data?

After the research, responsibility for data access and reuse will lie with my supervisor, who manages the Areda bucket for our research group. This ensures long-term oversight and compliance with the UT research data management policy. Any future reuse will follow institutional guidelines and applicable contractual or ethical constraints.

In contrast to during the research, data access control after the research is often managed differently. In accordance with UT research data management policy datasets should be archived in Areda. Please, mention the person(s) who manages the Areda bucket of your research group. In case of a contract with third parties, other person(s) having access to the data may be indicated.

DSec9. What will be the estimated total costs (€) involved in data security?

€ 0,00

In case of labour intensive metadata creation and documenting data files, hiring of personnel may be needed. Use cost estimations which possibly are available in the research project budget. For more information, see this guide for estimating research data management costs. For costs of UT RDM services see this overview.

Data preservation

In this section you are asked to think about preserving (a selection of) data which is collected, generated or used in your research. Proper data preservation is always needed, in line with the Netherlands Code of Conduct for Research Integrity.

Apart from preserving research data it is UT RDM policy to make them, if possible, publicly available in a trusted data repository, such as 4TU.ResearchData and DANS. This issue is covered in the next section.

*Note: In the UT RDM policy, latest version from 2018, it is stated that data should be **archived** in a trusted data repository. This policy will be adapted to the recent practice described above.*

Use the table below to answer the following questions for each type of data you collect or generate. If you want to add or change an item, please return to the table Data Collection. In case of a new item, don't forget to add information, also in the table of Data Storage section. Be aware that secondary data should not be added in this table. You can give details about preservation of secondary data in question 5.

1. **Which of the data will be selected for preservation and in which location?**
2. **Which file formats will be used for long-term preservation and accessibility?**
3. **After the end date of the project, how many years will the selected data be preserved?**
4. **What must be done with the data after the end of the preservation period?**

Selected data can best be preserved in Areda, the UT Archive for Research Data, unless explicitly prohibited by a third party (contract). If possible, use non-proprietary (open) persistent formats, such as pdf, txt, rif, dot, tif, flac, csv or xml. More information about preferred formats you can find at DANS or 4TU.ResearchData. Preserve the data for at least 10 years (in compliance with the UT data policy) unless there are legal or contractual conditions for choosing another preservation period. In some cases, data files have to be destroyed when the data are not needed anymore for the purpose for which it has been processen, mostly when containing sensitive personal data.

Type of data	DSP1. To be preserved (in)	DSP2. Preservation format	DSP3. Preservation period	DSP4. Post preservation action
nft_prices_raw_2023-06_v1.1_ [secondary]]	○ No ◉ Areda ○ Other	CSV (.csv)	10 yrs	◉ To be destroyed ○ To be decided ○ Other
p2p_loans_raw_2020_v1.1_ [secondary]	○ No ◉ Areda ○ Other	Feather (.feather)	10 yrs	◉ To be destroyed ○ To be decided ○ Other
Bubble_detection_code_v1.1	○ No ◉ Areda ○ Other	R markdown (.qmd)	10 yrs	◉ To be destroyed ○ To be decided ○ Other
Bubble_detection_code_v2.1	○ No ◉ Areda ○ Other	R markdown (.qmd)	10 yrs	◉ To be destroyed ○ To be decided ○ Other
P2P_loan_analysis_code_v1.1	○ No ◉ Areda ○ Other	Jupyter notebook (.pynb)	10 yrs	◉ To be destroyed ○ To be decided ○ Other
P2P_loan_analysis_code_v2.2	○ No ◉ Areda ○ Other	R markdown (.qmd)	10 yrs	◉ To be destroyed ○ To be decided ○ Other
P2P_loan_analysis_code_v3.3	○ No ◉ Areda ○ Other	Jupyter notebook (.pynb)	10 yrs	◉ To be destroyed ○ To be decided ○ Other
P2P_loan_analysis_output_tables	○ No ◉ Areda ○ Other	CSV (.csv)	10 yrs	◉ To be destroyed ○ To be decided ○ Other
P2P_loan_analysis_plots	○ No ◉ Areda ○ Other	PNG (.png)	10 yrs	◉ To be destroyed ○ To be decided ○ Other
P2P_loan_analysis_model_summaries	○ No ◉ Areda ○ Other	PDF (.pdf)	10 yrs	◉ To be destroyed ○ To be decided ○ Other
Bubble_detection_output_tables	○ No ◉ Areda ○ Other	CSV (.csv)	10 yrs	◉ To be destroyed ○ To be decided ○ Other
Bubble_detection_plots	○ No ◉ Areda ○ Other	PNG (.png)	10 yrs	◉ To be destroyed ○ To be decided ○ Other

Bubble_detection_model_summaries	○ No ◉ Areda ○ Other	PDF (.pdf)	10 yrs	◉ To be destroyed ○ To be decided ○ Other

What will be the estimated total volume of the selected data to be preserved in Areda?

Amount of estimated total volume	Units of total volume
10-100	GB (GigaByte)

DSP5. Are there secondary data to be preserved?

◉ Yes ○ No

How will this be arranged in such a way that these secondary data are preserved and accessible for verification and/or reproducibility?

Secondary data obtained from sources CoinMarketCap, NonFungible.com, OpenSea.io, and Bondora.com will be archived, referenced and documented in detail, including the access dates and retrieval methods. This ensures that the data sources remain accessible for verification and reproducibility, in accordance with their respective terms of use of each data vendor.

Whether or not you are allowed to preserve secondary data depends on the terms of use. In either case, explain how accessibility to the data you used will be arranged after the project.

DSP6. For the sake of proper verification and reuse, what preparations will be made before preserving the data?

Before preservation, the data will undergo final cleaning and quality checks to ensure consistency and completeness. All scripts used for preprocessing and analysis (e.g., time series adjustments, resampling, filtering) will be documented and included. File formats will be converted to open formats such as .csv and .txt for tabular data and .Rmd or .py for code to support long-term accessibility. Metadata and a README file describing variable definitions, time ranges, data sources, and transformation steps will be added. Since no personal or sensitive data is included, anonymization is not required.

In some cases, data need to be cleaned, processed further, checked, etc. before being preserved. Also digitization or file conversion may be needed. When personal data is used anonymization or pseudonymization is needed.

DSP7. What will be the estimated total costs (€) involved in the preparation of data for preservation?

€ 0,00

Only costs for preparing data for preservation. Costs of preservation of the data is part of the next section, Making data (publicly) available. Use cost estimations which possibly are available in the research project budget. For more information, see this guide for estimating research data management costs. For costs of UT RDM services see this overview.

Making data (publicly) available

This section is about depositing and making data (publicly) available in trusted repositories such as 4TU.ResearchData and DANS.

Use the table below to fill in the answers to the following questions for each type of data you collected or generated. If you want to add or change an item, please return to the table Data Collection. In case of a new item, don't forget to add information, also in the table of Data Storage and the Data Preservation section.

1. **Which of the data will be selected to be deposited in a trusted repository?**
2. **Which open file formats will be used for making data publicly available?**
3. **Which trusted repository will be selected for depositing the selected data?**
4. **Which of the selected data will be made publicly available?**
5. **Which license of use will be assigned to the public data?**

The UT research data management policy, faculty data policies and many funders strongly recommend making data publicly available in a trusted repository (having for instance a CoreTrustSeal / Data Seal of Approval). Choose 4TU.ResearchData for beta-sciences or technical-sciences data, DANS for humanities, health sciences, social and behavioral sciences, oral history and spatial sciences. Find other trusted data repositories on the CoreTrustSeal website. Be sure that public data sets will get a persistent identifier, such as DOI (Digital Object Identifier) for easy retrieval and sustainable linking and citation.
When making data publicly available, use non-proprietary (open) persistent formats, such as pdf, txt, rif, dot, tif, flac, csv or xml. More information about preferred formats you can find at DANS or 4TU.ResearchData.

In general, public access to data is preferred, and many funders demand this. Sometimes embargo is needed. In other cases, e.g. for reasons of confidentiality, data cannot be made public. About assigning a use license to public data, you can find more information on the 4TU.RD website.

Type of data	DA1. To be published	DA2. Open file format	DA3. Repository	DA4. Public
nft_prices_raw_2023-06_v1.1_ [secondary]]	◉ Yes ○ No	.csv	○ 4TU.RD ○ DANS ○ Other	○ Yes ○ Embargo ◉ No

Type of data	DA1. To be published
p2p_loans_raw_2020_v1.1_ [secondary]	○ Yes ◉ No

Type of data	DA1. To be published	DA2. Open file format	DA3. Repository	DA4. Public
Bubble_detection_code_v1.1	(•) Yes ( ) No	.qmd	(•) 4TU.RD ( ) DANS ( ) Other	( ) Yes ( ) Embargo (•) No
Bubble_detection_code_v2.1	(•) Yes ( ) No	.qmd	(•) 4TU.RD ( ) DANS ( ) Other	( ) Yes ( ) Embargo (•) No
P2P_loan_analysis_code_v1.1	(•) Yes ( ) No	.pynb	(•) 4TU.RD ( ) DANS ( ) Other	( ) Yes ( ) Embargo (•) No
P2P_loan_analysis_code_v2.2	(•) Yes ( ) No	.qmd	(•) 4TU.RD ( ) DANS ( ) Other	( ) Yes ( ) Embargo (•) No
P2P_loan_analysis_code_v3.3	(•) Yes ( ) No	.pynb	(•) 4TU.RD ( ) DANS ( ) Other	( ) Yes ( ) Embargo (•) No
P2P_loan_analysis_output_tables	( ) Yes ( ) No			
P2P_loan_analysis_plots	( ) Yes ( ) No			
P2P_loan_analysis_model_summaries	( ) Yes ( ) No			
Bubble_detection_output_tables	( ) Yes ( ) No			
Bubble_detection_plots	( ) Yes ( ) No			
Bubble_detection_model_summaries	( ) Yes ( ) No			

For those data that are not made publicly available, what are the reasons for this? Will availability of the data be restricted to a certain group or community, and if so, which, why and how?

Due to licensing constraints and terms of use from the original data sources (e.g. CoinMarketCap, OpenSea, Bondora), the data cannot be made fully publicly available in an open repository. However, a curated and well-documented version of the data will be deposited in 4TU.ResearchData with access restrictions. Interested researchers can submit a formal request for access, which will be evaluated on a case-by-case basis by the author. This approach ensures compliance with source platform policies while also supporting academic transparency and reproducibility.

Both the UT research data management policy and funder policies demand data to be as open as possible. Possible reasons for not making data publicly available are legal restrictions on personal data, intellectual property, political sensitivity, commercial interests (also patenting), and public security.

DA6. When necessary, how will specific software for viewing, analysing or reusing the data be made available?

Although the data will not be made publicly available, all software used for analysis in this research is based on open-source programming languages and tools, specifically R and Python. These environments are freely accessible and well-documented, which ensures that, if access to the data is granted under controlled conditions (e.g., within the research group or upon request), others will be able to view, analyze, and potentially reproduce results without restrictions related to proprietary software. Relevant packages and scripts are also written in standard, human-readable formats (e.g., .R, .py, .ipynb) to ensure future interpretability and reusability.

In some cases specific software is needed for viewing or analysing data. It is important that this software is available for verification and/or reuse of the data.

DA7. Will exclusive or restricted use of the data be claimed by third parties?

○ Yes ◉ No

DA8. What are the estimated total costs (€) for depositing the data?

€ 0,00

Check any cost arrangement between your faculty with a particular data repository. If no specific arrangement apply, check depositing costs at the website of the repository. Use cost estimations which possibly are available in the research project budget. For more information, see this guide for estimating research data management costs. For costs of UT RDM services see this overview.