RESEARCH DATA MANAGEMENT POLICY THE DEPARTMENT OF CIVIL ENGINEERING & MANAGEMENT

Version: 1.1

17-06-2024

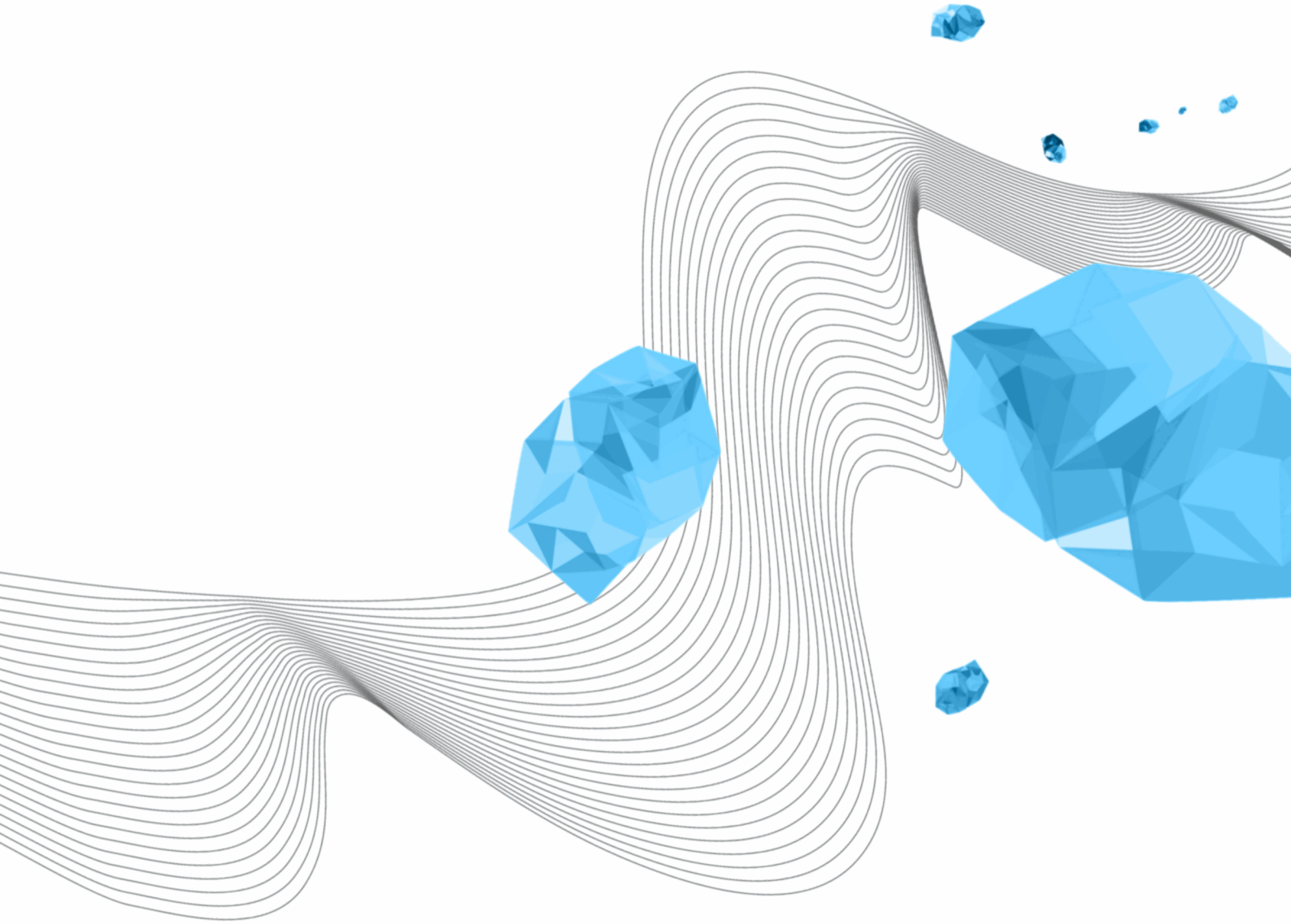

TABLE OF CONTENTS

1 Introduction 3
2 Department-specific roles and responsibilities 3
3 Data management plan (DMP) 4
4 Privacy regulations 4
5 Data storage and transfer 4
6 Data documentation 5
7 Data sharing 6
8 Data archiving 6
9 Data registration 7
Appendix 8

Document management:

Version	Author(s)	Description
1.0	Markus Berger Victor Wanningen	Updated the CEM Data Management Policy (DMPo) from Nov 2022 and created alignment with the faculty Research Data Management (RDM) policy
1.1	Markus Berger Victor Wanningen	Consulted with the cluster leads and RDM coordinators

1 INTRODUCTION

This document specifies the Research Data Management Policies of the University of Twente (RDM policy UT) and the Faculty of Engineering Technology (RDM policy ET) for the Department of Civil Engineering & Management (CEM). It follows the same structure as the RDM policy ET and links to it as much as possible to avoid repetition while providing targeted guidance for CEM. In the annexes of this document, it specifies aspects of data storage, sharing, and archiving for the clusters Construction Management Engineering (CME), Marine Fluvial Systems (MFS), Multidisciplinary Water Management (MWM), Soil MicroMechanics (SMM), and Transport Systems (TS).

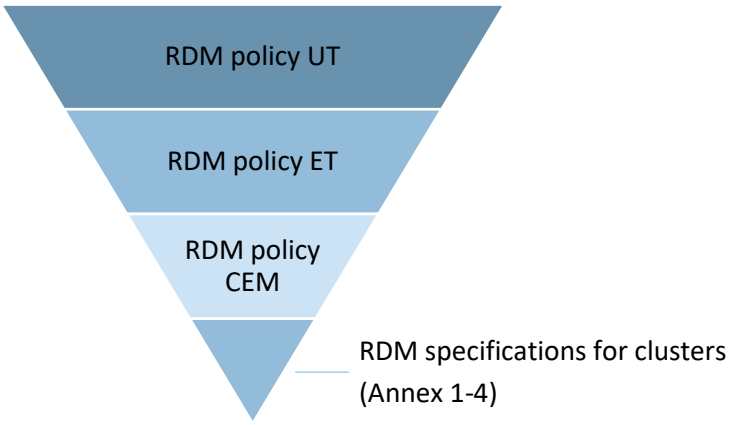

Figure 1. Hierarchy of RDM policies and specifications from UT to department level

2 DEPARTMENT-SPECIFIC ROLES AND RESPONSIBILITIES

In addition to the roles and responsibilities described in the RDM policy ET (faculty board, head of research chair, researcher), two additional roles are defined:

The Department Board

- Responsible for this policy and its implementation
- In particular, the portfolio holder research ensures that the policy is reviewed annually in consistency with the RDM policy ET

Coordinator data storage

- Within each cluster (CME, MFS, MWM, SMM, TS) a coordinator is assigned who is responsible for managing the storage and backup of research data in the group, e.g. setting writing/reading rights of project folders:
  - CME: Qinshuo Shen
  - MFS: Emre Ozturk
  - MWM: Han Su
  - SMM: Hongyang Cheng
  - TS: Baran Ulak

3 DATA MANAGEMENT PLAN (DMP)

In addition to the DMP requirements described in the RDM policy ET (chapter 3), this section specifies the data collection at CEM.

The type of primary and secondary data which is used and collected in CEM projects differs substantially, and also differs per cluster in the department. The size of the data also depends on the type of data (small data vs big data). Note that, raw data does not necessarily have to be big data (GBs), but processed data can easily become big (e.g. origin-destination matrices at PC6 level) (TBs). We can distinguish between different types of primary data and secondary data, depending on privacy issues, data ownership, UT project or joint projects, data collection in the Netherlands or abroad, open or closed data. In general, four types of data can be distinguished.

1. Privacy sensitive primary data, e.g. surveys, experimental data and smartphone data.
2. Non privacy sensitive primary data, e.g. air pollution measurements.
3. Secondary data without user restrictions (open data), e.g. GIS data (e.g. Land cover data), LCA datasets, socio-economic data (population data, employment data etc.) at PC4 level/grid level, public/open data: infrastructure network data (public transport, bicycle, car networks etc.), etc.
4. Secondary data with user restrictions (contract with data supplier), such as micro data from CBS, transport data from transport operators, or GSM data from mobile phone companies.

Each project DMP describes what type of data are used in the project, what agreements with data suppliers are made, how the different types of data are stored during the project, if and how different data types are shared during the project, and how the different types of data are preserved and made accessible after the end of the project. In the case of secondary data for which contracts with data suppliers have been set up, these contracts often do not allow data sharing for third parties or the use of the data beyond a single project. This is also to be described in the DMP.

4 PRIVACY REGULATIONS

See RDM policy ET, chapter 4.

5 DATA STORAGE AND TRANSFER

To specify the ET faculty's working rules on data storage (RDM policy ET, chapter 5), the storage requirements for user data, general group data, project data, and education data are specified for the five research clusters (CME, MFS, MWM, AMM, TS) in annexes 1-4. SMM has recently developed as an independent cluster, having been part of CME until March 2024. In the current version of the document, the storage requirements for SMM are still associated with those of CME.

LISA offers different types and qualities of UT network storage; different in terms of functionality, redundancy, and pricing. The project DMP should describe the location of project data (during the project) and data which is shared and/or archived (after the project ended).

If (a working copy of) the data needs to be (temporarily) stored in the cloud, on one of the non-UT network storage options given in the UT Storage-Decision-Tree tool, the folder on the cloud server needs to be shared with the PI. All data that is stored on a personal cloud (e.g. SurfDrive or OneDrive) needs to be moved to the P-drive or MS Teams/SharePoint by completion of the project.

Personal work files can be kept on the M-drive; M-drive should not be used for research data backup or storage as it is a personal drive, which will be deleted after leaving UT.

To safely and securely transfer research data to other researchers (also non-UT researchers), SURFfilesender should be used.

6 DATA DOCUMENTATION

To fulfil the data documentation requirements of RDM policy ET (chapter 6), all data generated in a project (e.g., PhD, Postdoc, third-party, etc.) within the CEM department needs to have a metadata and a README file. The metadata file summarizes the details of data being used/generated in the research. Table 1 can be used for this purpose. The PI of each project is responsible for producing and updating the metadata file.

Table 1. Template for projects' metadata

General Information		
Name of the project		
Start date		
End data		
Involved researchers		
Partners		
Funding source		
Data		
1	**Type** (e.g., survey, field data)	
	Format (e.g., .docx, .pdf, .mat)	
	Required software (e.g., Matlab, MS Word)	
	Source	Primary (i.e., collected) / Secondary (i.e., accessed)
	Level of development	Raw / Processed
	Statues	Final / Ongoing
	Date of collection/access	
	Date of last update	
	Owner	
	Permission for reuse	Yes / No
	Reuse condition (e.g., anonymization, license)	
	Contact person for reuse authorization	
	Primary storage location	
	Back-up Storage location	
	Comments	

To support the understanding and use of the data, each dataset shall contain a README file. It is recommended to use the README file template provided for data documentation on Areda. Associated guidance can be found here.

7 DATA SHARING

The ET faculty's data sharing requirements can be found in the RDM policy ET (chapter 6). For running research projects in CEM, the use of UniShare, MS SharePoint/Teams, or P-drive is recommended to share data with other researchers. If data are shared with other UT staff or students, these persons can be given access to a shared folder on a network drive.

8 DATA ARCHIVING

To specify the ET faculty's working rules on data archiving (RDM policy ET, chapter 8), the data archiving requirements for closed research projects in the five CEM clusters (CME, MFS, MWM, SMM, TS) are described in annexes 1-4.

The PI or responsible CEM professor is responsible for archiving of project data and data related material. The general recommendations for data preservation are:

1. Upon completion of the project, the principal investigator or responsible researcher(s) should store a backup of relevant project data on either P-drive or Areda-Pure (as described by research clusters in annexes 1-4).
2. If required by funders or journals, and/or being possible and desired, data (especially underpinning scientific publications) is also published at a national Data Repository (DANS or 4TU.ResearchData), or an alternative trusted repository (e.g. if agreed upon in international collaboration projects), using table 2 as guidance for handling different types of data. Every researcher can upload up to 1TB of data per year to the 4TU.ResearchData repository free of charge. By default, this data will be stored for a minimum of 15 years. The data should be openly made available, if possible. Only if there are specific restrictions (e.g. personal data, collaborations with others that do not allow to make the data openly available), an embargo can be used, or the data can be stored in the DANS repository with restricted access.

Table 2 describes the CEM data preservation plan, depending on the different types of data (raw/cleaned/processed/models/model outputs/publication data; primary/secondary; privacy sensitive or not).

Table 2 Data Preservation Plan

Data Type	Primary/ Secondary/ model	Privacy sensitive	Archived	Location	Duration
Raw data	Primary	Yes/no	Not Public	UT Network Storage if size in GBs, else alternative LISA archive service (e.g. Areda)	10 years
	Secondary	no	Not Public	UT Network Storage if size in GBs, else alternative LISA archive service (e.g. Areda).	10 years
	Secondary	Yes	Not archived	Data sources for which an NDA has been signed data need to be deleted after the project closure	-
Cleaned/ processed data (input to models)	Primary	Yes	Not public	UT Network Storage	10 years
	Secondary	No	Public	Data Repository (DANS or 4TU.RD), recommended, see point 2	15 years

Aggregated Anonymized Data	Primary	No	Public	Data Repository (DANS or 4TU.RD)	15 years
Model scripts and open source codes	Model	No	Preferably public, else or by request	Data Repository (DANS or 4TU.RD) and other relevant, e.g. GITHub	15 years
Model software (external)	Model	No	Not archived		
Modelled Results/Output	Model	No	Not public, only by request	UT Network Storage, if size in MBs, else alternative LISA archive (e.g. Areda) . Data should potentially be reproduced with the model scripts, software and raw/cleaned data.	10 years
Data in Publication (spreadsheets, figures, raster datasets, etc.)	Primary/ Secondary/ Model	No	Public	Data Repository (DANS, 4TU.RD, Zenodo)	15 years

9 DATA REGISTRATION

See RDM policy ET, chapter 9.

APPENDIX

Annex 1: CME and SMM: data documentation, storage and archiving locations

CME uses the data storage and backup facilities provided by the LISA (i.e., UT Group/Project drive, aka P-drive) as the primary space for data management and storage. Based on the current data and project data need for the next 5 years, CME would require 400 GB of space for data storage and backup. The following would apply to all members of the department:

- Staff and researchers are not entirely free to use any form of storage strategy for personal work files. Due to technical and security concerns, the staff members are suggested to consult with this decision tool to explore viable/recommended options for the storage of data. In principle, local hard drives, and GDPR-compliant online services such as OneDrive, and m-drive can be used for the storage of personal data.
- Important to note is that M drive will be deleted upon the termination of the employment contract. Therefore, one must make sure that it is only used for the storage of personal work files only. Research data should be stored in the corresponding Project Drive folder.
- For the running projects, the researcher should maintain a current folder on the P-drive (P:\ET\CME\Current). Other than the cluster group members, Qinshuo Shen, i.e., the IT admin of the CME group has access to these folders;
- For completed projects, all the data need to be transferred to an archive folder in the P-drive (P:\ET\CME\Archive);
- Upon completion of the project, data should also be published in the 4TU Data Repository or Areda-Pure;
- The education data should be also archived, next to Canvas, in MS Teams/SharePoint. This can be placed in the Teams channel of SME staff.

Table 4 represents CME data storage configuration. The main working folder for the members of the CME department is P:\ET\CME\Current. While the project or research is ongoing, all related data should be stored under this folder. When the project is finished, the principal investigator or responsible researcher(s) should transfer project data to P:\ET\CME\Archive . All the working data of the department under locations P:\ET\CME\Current , and P:\ET\CME\Archive automatically backup by LISA to corresponding locations: P:\ET\CME\CurrentBackup and P:\ET\CME\ArchiveBackup .

Table 3. **CME** data storage configuration

	Location	**Amount (Gb)**	**LISA Quality standard**	**Backup**	**Cost (€) per TB**	**Total cost per year**
Working	P:\ET\CME\Current	200	Normal Quality Storage	Duplicated and 28 day back-up	75	15
	P:\ET\CME\Archive	200	Normal Quality Storage	Single storage and 28 day back-up	75	15
Education	MS SharePoint – Teams Channel of CME staff			Recovery of 30 days		

Annex 2: MFS: data documentation, storage and archiving locations

At the moment, data management is divided per project.

Members of the group are free to use a personal data storage strategy such as:

- Raw data and processed data is individually stored by every researcher in their own PC with copy on external hard disks (for temporary storage) or SURFDrive. All external hard disks should be encrypted by using Bitlocker.
- Some projects like RiverCare and Sandbox make use of external servers like Deltares and the version control system of Github for scripts and fieldwork collected data.
- Data supporting journal publications is available as supplementary material and in some cases as separate dataset in 4TU.RD.

According to the policy of the group, we plan to improve the data management strategy as follows:

- Raw data and processed data should be stored in a central folder located on the P:\ET\CEM\WS\Projects and P:\ET\CEM\CSNBE\Projects. The folder should be accessible remotely via UTwente VPN connection to prevent the use of multiple working directories.
- All MFS members can access the Projects folder if there are no user restrictions or privacy issues.
- If there are user restrictions or privacy issues, the PI sets the reading/writing rights of the project folder.
- Each Projects folder will be divided per researcher following a similar data structure that can be easily archived and understood after the project is finished.
- Data for each researcher will follow the data type structure:
    - raw data,
    - scripts,
    - processed data,
    - outputs.
- Selected processed data and model output are made available as supplementary information or supporting dataset to published articles. Data repository depends on the review process and access type:
    - Zenodo is used for restricted access and
    - 4TU.RD or a journal repository is used for open access.
- For completed projects, all the data folders should be documented following Table 1 template and uploaded to Areda-Pure. Additionally the project folder should be moved to P:\ET\CEM\WS\ClosedProjects and P:\ET\CEM\CSNBE\ClosedProjects.
- Personal work files can be kept on the M-drive.
- All course material on Canvas and other drives should be stored and aggregated on the P-drive (P:\ET\CEM\WS\Education or P:\ET\CEM\CSNBE\Education), as a backup and to enable a smoother transition of the course material to another staff member if needed. For files published on Canvas in .pdf format, this includes versions of these files in the original file format (e.g. .docx or .pptx).

Table 4 represents MFS data storage configuration. All the working data of the department under locations:

- **P:\ET\CEM\WS\Projects**
- **P:\ET\CEM\CSNBE\Projects**

When the project is finished, the principal investigator or responsible researcher(s) should upload the project data to Areda-Pure and migrate the folder contents to the following locations:

- **P:\ET\CEM\WS\ClosedProjects**
- **P:\ET\CEM\CSNBE\ClosedProjects**

Table 4. **MFS** data storage configuration

	Location	**Amount**	**Data service type**	**Backup**	**Cost (€) per TB**
	P:\ET\CEM\WS\Projects P:\ET\CEM\CSNBE\Projects	~1TB	UT-drive	Duplicated and 31 day back-up	75
	P:\ET\CEM\WS\ClosedProjects P:\ET\CEM\CSNBE\ClosedProjects	~20TB	UT-drive	Duplicated and 31 day back-up	75

Annex 3: MWM: data documentation, storage and archiving locations

MWM uses the data storage and backup facilities provided by UT Group/Project drive (i.e. the P-drive) and Microsoft (i.e. OneDrive and SharePoint) as the primary media for data management and storage. The following applies to all members of the group:

- Members of the group shall use abovementioned back-up facilities to maintain a folder for all their running projects, which allows for sharing documents, parallel working on documents, and synchronizing with local storage on notebooks. Further, at least one additional group member shall have access to the project folder to make sure it is accessible in an emergency situation. By default, this is the secretary of the group (Dorette Olthof).
- For project where the involved data is larger than 200 GB, members can use external hard drive to maintain the project data and use another external hard drive to back up the data. In this case, members need to discuss the external hard drive management with their supervisors, e.g. where the hard drive will be kept and whether or not the external hard drive will be moved often. Further, hard drives need to be encrypted and should only be used short term before the data is offloaded to the P-drive.
- Relevant education data (ppt files, readers, exams, etc.) shall be stored in the MWM-staff teams channel (i.e. in MS SharePoint) – in this way it is backed up and accessible to all staff members. After a course has been given, the teaching material should be updated annually.
- For each publication, completed projects, internal master thesis where the total size of data is not large (e.g. < 200 GB), all the data need to be transferred to the archive folder in the P-drive using predefined folder template structure (P:\ET\MWM\Archive). Completed projects, internal PhD and master thesis need to link to the relevant publications; Large data (> 200 GB) can be archived in P-drive or AREDA.
- Processed data and model output selected to be reported in academic publications are made available as supplementary information (SI) to these publications, or stored in trusted open-access repositories referred to the publication; preferably, publication and data both have DOI's and are mutually referenced.
- Personal work files of group members can be kept on the M-drive; M-drive should not be used for research data backup or storage as it is a personal drive, which will be deleted after leaving UT;

Table 5 represents the MWM data storage configuration. All the data of the group under locations P:\ET\MWM\ \Archive are automatically backed up by UT.

Table 5. **MWM** data storage configuration

	Location	**Amount (Tb)**	**LISA Quality standard**	**Backup**	**Cost (€) per Tb**	**Cost (€) per year**
Education data	MS SharePoint – Teams channel MWM-staff			Recovery for 30 days		
Running projects	P-drive P:\ET\MWM\Projects	0.5		Hourly backup for recent 2 d & daily backup for recent 30 d	75	38
	MS OneDrive/SharePoint			Recovery for 30 days		
Closed projects/ theses	P-drive P:\ET\MWM\Archive	7	Normal Quality Storage	Hourly backup for recent 2 d & daily backup for recent 30 d	75	525
	Total					563

Annex 4: Transport Systems (TS) Cluster group: data documentation, storage and archiving locations

Project related data is stored in project folders at the standard Group folder (P-drive) \\ad.utwente.nl\org\ET\TEM\Projects. All cluster members can access project folders if there are no user restrictions or privacy issues. If there are user restrictions/privacy issues, only project researchers, PI, responsible TS professor and the IT admin of the TS group (Dr. Baran Ulak) has access to these folders. The TS IT Admin sets the reading/writing rights of the project folder.

Raw and processed secondary data which can be shared among members of the cluster (e.g. CBS data, OViN Travel Survey Data, GIS data) are placed in the "RESEARCHDATA" project folder at P:\ET\TEM\ Projects\RESEARCHDATA). Subfolders are organized by data source (e.g. OVIN, Fietstelweek, etc.). Every folder on P:\ET\TEM\Projects has the DMP description and a text file with metadata (in English).

Education: course materials, data and documents related to BSc and MSc courses is stored in the folder //Org/ET/TEM/EDU/ Each CEM course run by TS members has a separate folder, including thesis projects for BSc (//Org/ET/TEM/EDU/MOD12/BSc thesis projects and MSc thesis (Org/ET/TEM/EDU/Mthesis)

Microsoft Teams Channels / Microsoft SharePoint are used for group documents (e.g. meeting minutes) and joint working on reports, papers and course materials. Any project related documents stored on Teams Channels/Sharepoint will be moved to the project folder on P drive for archiving, after the end of the project. At the minimum two cluster members (at the minimum the cluster admin and PI) are owners. Google Drive folders (using cluster members' UT accounts), OneDrive or Sharepoint can be used for project proposals writing with non-UT partners. UniShare is used to store and use data in projects together with international partners or with national partners. At least two cluster members and project PIs/Co-PIs are owners and admins of each project UniShare storage.

Table 6: **TS** data storage and archive locations

	location	volume	quality	Backup
	Running project folders: P:\ET\TEM\Projects.	7TB;	LISA standard (high)	duplicated and 28 day back-up
	Education: P:\ET\TEM\EDU		LISA standard (high)	duplicated and 28 day back-up
	P:\ET\TEM\Users\ (data storage, non-project related)		LISA standard (high)	duplicated and 28 day back-up
	Closed project folders P:\ET\TEM\Archive\ClosedProjects		LISA standard (high)	duplicated and 28 day back-up
	Unishare storage	1TB;	LISA standard (high)	duplicated and 28 day back-up
	MS Teams channel; sharing documents for joint writing			
	Sharepoint/Onedrive/Google drive – project proposals with non-UT partners			