

Data Management Plan

TRUST: Norwegian Centre for Trustworthy AI

Reference number	
Date	10 th October 2025
Version	0.1 (initial draft-living document)

Revision history		
Edited by	Date	Description
Morten Dæhlen	6 th October 2025	First draft created
Jenny Graver	9 th October 2025	Reviewed and additions
Malcolm Langford	10th October 2025	Reviewed and additions
Morten Dæhlen	10th October 2025	Reviewed with minor adjustments and accepted

Table of contents

1. Introduction	2
2. Project Information	2
3. Data Description.....	3
4. Standards and Metadata.....	4
5. Data Storage and Security	4
6. Data Sharing, Access and Preservation	5
7. Ethical and Legal Compliance	6
8. Responsibility, Resources and Review/Updates	6

1. Introduction

The aim of this document is to provide an overarching Data Management Plan for TRUST – The Norwegian Centre for Trustworthy AI. It is intended as a living document that will be updated as TRUST's research practices evolve.

TRUST is firmly committed to the principles of Open Science. For research data, this means that TRUST will handle research data in a transparent, ethical, and secure manner. Research data management corresponding to the principles of Open Data helps ensure that the research process is open and reproducible, thus increasing both the quality and trustworthiness of the research, its outputs, and resulting innovations.

TRUST is the recipient of a significant amount of public funds through the Research Council's AI centre call and public sector partners; with significant additional funds from private sector partners. TRUST has a responsibility to manage this funding well. Since research data is a valuable commodity, making data as open as possible and as closed as necessary will help maximise the value of these financial contributions while also respecting ethical principles and legal requirements.

This document details the practical principles of research data management that TRUST's research will adhere to in the context of current legal requirements and provisions on intellectual property and confidentiality in the Consortium Agreement. For information about the ethical underpinnings of the centre's activities, we refer to our Responsible Research and Innovation (RRI) policy which is annexed to the Consortium Agreement.

2. Project Information

Project title:

TRUST – The Norwegian Centre for Trustworthy AI

Principal investigators / team members: The centre involves researchers from the fields of data and computation sciences, social science, law, ethics, as well as researchers and staff from user partners; see project description.

Project duration:

2025–2030

Funding source(s):

The Research Council of Norway and partner institutions, see project description.

Project overview:

TRUST is a national centre comprising multiple interconnected research projects exploring the development and application of *trustworthy artificial intelligence (AI)*. Each project may generate or process different types of data, ranging from user studies to experimental data.

This overarching **Data Management Plan (DMP)** applies to all subprojects. However, as data handling occurs locally at each partner institution, TRUST also builds upon and aligns with the **data management policies of those institutions** (e.g., the University of Oslo,

SINTEF, NR, etc.). Each subproject will follow the procedures of the institution at which the project is located, including writing its own specific data management plan as needed.

3. Data Description

Data to be collected/created:

TRUST will collect and generate a variety of data types, including but not limited to:

- **Survey data:** e.g., responses from user studies on AI explainability.
- **Experimental data:** e.g., logs from machine learning model evaluations or AI-driven decision-support systems.
- **Time series:** e.g., system performance metrics, user interaction traces, or environmental sensor data.
- **Images:** e.g., annotated datasets for computer vision model training.
- **Text:** e.g., documents, transcripts, or model-generated content for natural language processing tasks.
- **Video:** e.g., large sets of information resulting from two spatial dimensions and high resolution, typically requiring 60 frames per second for accurate representation.

In principle, all data types required for developing, training, and testing AI models may be used.

Data formats:

Most data will be stored in standard, interoperable formats, such as:

- CSV, JSON, XML for tabular or structured data
- PNG, JPEG for image data
- TXT, PDF, or DOCX for textual content
- HDF5 or TFRecord for machine learning datasets

However, some raw or intermediate data (e.g., binary model checkpoints or simulation outputs) may appear and must be handled by this DMP. These proprietary representations of data will depend on the specific research tools that are used. Note that use of such data otherwise confirms to the provisions of the Consortium Agreement concerning any data that is proprietary or confidential.

Estimated volume of data:

Data volume will vary by project. For example, an AI model trained on image data may require terabytes of storage, while survey data may be in megabytes. Large datasets will be managed using **national or international data storage infrastructures** such as **NIRD (Norwegian Infrastructure for Research Data)**.

Purpose of the data:

The data supports investigations across TRUST's research areas—such as responsible uncertainty quantification, Secure AI and digital twins and hybrid modelling.

Examples include:

- Developing algorithms that can explain their reasoning.

- Studying public attitudes towards ethical and legitimate automated decision-making.
- Developing AI decision support tools and digital twins from large volumes of legal judgments and administrative decisions
- Collaborating with industry to test the trustworthiness of AI solutions in real-world environments.

Where appropriate, data may also be reused by **user partners** for innovation purposes.

4. Standards and Metadata

Documentation:

Metadata are essential for ensuring that research data align with the FAIR principles, that is, that the data are Findable, Accessible, Interoperable and Reusable. All datasets will be accompanied by **metadata** stored in a centralized **TRUST data dictionary**, with references to more detailed metadata systems at each partner institution.

For example, a dataset on AI ethics perception might include metadata such as:

- Data creator(s) and date of collection
- Variable descriptions (e.g., “trust_score,” “participant_age”)
- Collection methodology and consent details

Metadata standards:

TRUST will adopt existing metadata representation schemes such as **Dublin Core**, **DataCite**, or **DCAT**, and will extend or customize these if necessary to meet project-specific needs (e.g., AI model metadata standards such as **ML Schema**).

Version control practices:

All datasets, code, and documentation will be versioned using institutional or collaborative version control systems (e.g., **GitLab**, **GitHub Enterprise**, or **local ICT-managed repositories**) to ensure traceability and reproducibility.

5. Data Storage and Security

Storage solutions:

Data will be stored using approved institutional and commercial cloud services, such as:

- **EduCloud** (University of Oslo’s academic cloud)
- data.sintef.no (SINTEFs research data cloud)
- **TSD (Services for Sensitive Data)** for projects involving personal or sensitive information.

Data that is highly sensitive and/or requires strong data minimisation treatment under the GDPR will be stored in a highly secure environment like TSD. For example, a medical AI project will store patient-related data only in TSD, while non-sensitive model data can reside in EduCloud. This is the current approach for the S2-Data research infrastructure that will be

used for many TRUST projects. Currently, S2-Data includes decisions on asylum seekers in TSD, but judgments in international investment law in EduCloud.

Backup procedures:

Data will be backed up automatically in line with institutional IT policies. For instance, universities typically ensure daily backups with 30-day recovery options.

Access control:

Data access will be regulated according to sensitivity level:

- **Open access** for public, non-sensitive datasets.
 - **Restricted access** using two-factor authentication for sensitive or proprietary datasets.
 - **Encryption** (both in transit and at rest) for all personal or confidential data.
-

6. Data Sharing, Access and Preservation

Plans for sharing:

Data and publications will be shared through institutional repositories (e.g., UiO's **Dataverse**) or European infrastructures (e.g., **Zenodo**, **OpenAIRE**).

Example: A de-identified dataset on human–AI collaboration may be made publicly available after publication, subject to an embargo period if required.

Ethical and legal considerations:

All sharing will follow the **TRUST Policy for Responsible Research and Innovation (RRI)**, including compliance with:

- Informed consent from participants.
- Proper anonymization of personal data.
- Respect for copyright and data ownership.

Licenses applied:

TRUST will primarily use **Creative Commons (CC-BY or CC0)** licenses for open data. However, for commercial or confidential collaborations, **proprietary licensing** (e.g., under NDA) may be necessary.

Expected audience for reuse:

Researchers, educators, policymakers, and private/public sector actors interested in AI ethics, data governance, and responsible innovation.

Archiving strategy and duration:

Data will be archived indefinitely when of long-term value (e.g., benchmark datasets). Personal data will be anonymised before archiving. Data containing personal data will be stored and deleted according to data protection requirements for each subproject. To meet sustainability goals, the DMP of each subproject will also include **data disposal and deletion procedures** for outdated or redundant data.

7. Ethical and Legal Compliance

Consent and anonymization:

If human participants are involved, TRUST will use robust **informed consent procedures**. Examples include:

- Written consent forms explaining data usage and storage duration.
- Anonymization, Pseudonymization (replacing names with IDs) or **synthetic data generation** to safeguard privacy.

Compliance:

All activities will adhere to **ethical standards, institutional policies, and legal frameworks**. TRUST will ensure participants' rights, intellectual property protection, and secure data processing throughout the research lifecycle.

Data protection:

All handling of personal or sensitive data will comply with **GDPR** and relevant Norwegian and EU regulations. Subprojects processing personal data will be notified to the Norwegian Agency for Shared Services in Education and Research (Sikt) for review of GDPR compliance (UiO) or follow equivalent procedures where required at other institutions.

8. Responsibility, Resources and Review/Updates

Responsibility:

All researchers and support staff are responsible for following this DMP. One member of the leadership team will serve as the **Data Management Officer (DMO)**, overseeing adherence across all subprojects and coordinate with the structures set in place for the RRI.

Resources:

Each subproject will allocate **0-3% of its budget** (there will be project that is not using data) for data management tasks—covering documentation, metadata creation, secure storage, and compliance reporting.

Review/Updates:

Data management practices will be reviewed **annually**. Each subproject of TRUST must report its data management activities, and the overall DMP will be updated to reflect new technologies, ethical standards, or legal requirements.