European Lighthouse to Manifest Trustworthy and Green AI

Data Management Plan

Deliverable D1.2

Dissemination Level: PUBLIC

Topic: HORIZON-CL4-2022-HUMAN-02-02

Project Period: 01.09.2023 to 31.08.2026

Lead Participant: NTNU

Due Date of Deliverable: 29/02/2024 **Actual Date of Submission:** 29/02/2024

Deliverable Description

Deliverable Name	Data Management Plan
Deliverable Number	D1.2
Work Package	WP 1
Associated Task	T 1.4
Covered Period	M1-M36
Due Date	M6
Completion Date	M6
Submission Date	29/02/2024
Deliverable Lead Partner	NTNU
Deliverable Author(s)	Pankaj Pandey
Version	1.0

Dissemination Level		
PU	Public, fully open, e.g., web (Deliverables flagged as public will be automatically published in CORDIS project's page)	**X**
SEN	Sensitive, limited under the conditions of the Grant Agreement	
EU-R	EU RESTRICTED under the Commission Decision No2015/444	
EU-C	EU CONFIDENTIAL under the Commission Decision No2015/444	
EU-S	EU SECRET under the Commission Decision No2015/444	

Document History

Version	Date	Change History	Author(s)	Organisation
0.1	29/02/2024	Initial Version	Pankaj Pandey	NTNU
0.2	29/02/2024	Review Comments	Antti Tahvanainen, Ioana Bara-Busila, Jean-David Bodenan	ESF
1.0	29/02/2024	Final Version	Pankaj Pandey	NTNU

Distribution List

Date	Issue	Group
29/02/2024	Submission	European Commission

This document is released as part of the ENFIELD project, which receives funding from the European Commission through the HORIZON-CL4-2022-HUMAN-02 call and is governed by Grant Agreement No. 101120657.

List of Acronyms

Acronym	Definition
AAM	Advanced Air Mobility
ADAPT	Aviation Data Analytics Platform
AI	Artificial Intelligence
ANN	Artificial Neural Network
ASSET	Adaptive Security for Smart Internet of Things in eHealth
CCIS	Center for Cyber and Information Security
DAS	Data Acquisition Station
DMO	Data Management Officer
DMP	Data Management Plan
DPIA	Data Protection Impact Assessment
DPO	Data Protection Officer
ECoE	Eratosthenes Centre Of Excellence, Cyprus
EDP CNET	CNET Centre For New Energy Technologies Sa, Portugal
ENFIELD	European Lighthouse to Manifest Trustworthy and Green AI
EO	Earth Observation
ERA	European Research Area
ESF	Fondation Europeenne De La Science, France
EV	Electric Vehicle
FAIR	Findable, Accessible, Interoperable, Reusable
GDPR	General Data Protection Regulation
GNSS	Global Navigation Satellite Systems
GPS	Global Positioning System
GPU	Graphics Processing Unit
INESC TEC	INESC TEC - Instituto De Engenhariade Sistemas E Computadores, Tecnologia E Ciencia, Portugal
IoT	Internet of Things
ISK	Iskraemeco, Merjenje In Upravljanjeenergije, D.D., Slovenia
KNOW	KNOW-Center Gmbh Research Center For Data-Driven Business & Big Data Analytics, Austria
MAS	Multi Agent System
ML	Machine Learning
NORCICS	Norwegian Centre for Cybersecurity in Critical Sectors
NRS	Norsk Regnesentral, Norway
NTNU	Norges Teknisk-Naturvitenskapelige Universitet, Norway
OpenAIRE	Open Access Infrastructure for Research in Europe
RCN	Research Council of Norway
SAR	Synthetic Aperture Radar
SINTEF	SINTEF AS, Norway
SSH	Social Sciences and Humanities
TELENOR	TELENOR ASA, Norway
TES	Third-party Exchange Scheme
TIS	Third-party Innovation Scheme
TU/e	Technische Universiteit Eindhoven, Netherlands
UPB	Universitatea Nationala Pentru Stiinta Si Tehnologie Politehnica Bucuresti, Romania

Executive Summary

The ENFIELD project, with its ambitious objectives and collaborative framework, places a strong emphasis on responsible, ethical, and effective data management. The Data Management Plan (DMP) aligns with the FAIR principles, ensuring that data is Findable, Accessible, Interoperable, and Reusable. This comprehensive plan delineates strategies, policies, and practices to maintain the integrity, accessibility, and usability of the generated data.

Key Highlights:

- **FAIR Principles Integration:** The DMP integrates FAIR principles into every facet of data management, from collection and storage to sharing and preservation. This ensures that ENFIELD's data is aligned with global standards and maximally beneficial to the broader scientific community.

- **Data Quality Assurance:** ENFIELD prioritises data quality to guarantee the reliability and validity of insights. Rigorous validation processes, meticulous documentation standards, and adherence to best data collection and processing practices are implemented to uphold high data quality standards.

- **Publication and Open Access:** ENFIELD commits to using the Open Research Europe scholarly open-access publishing service for Horizon Europe. This commitment expedites publication, supporting research integrity, reproducibility, and transparency. Scientific publications will be made available through open-access options, with selected publications accessible through Golden Open Access. The remaining publications will be accessible on the project website and the OpenAIRE repository.

- **Dissemination Efforts:** Project partners will make significant efforts to share project results through institutional open-access platforms. NTNU's 'NTNU Open' is an example of such a platform, contributing to the broader dissemination of ENFIELD outcomes.

In conclusion, the ENFIELD Data Management Plan is a foundational guide for responsible and impactful data management throughout the project's lifecycle. By embracing FAIR principles, ensuring data quality, and leveraging collaborative platforms, ENFIELD aims to contribute valuable insights to the scientific community while fostering transparency, accountability, and long-term impact.

Table of Contents

1. Introduction

The ENFIELD project stands at the forefront of advancing Artificial Intelligence (AI) research, aiming to establish a distinctive European Center of Excellence. This collaborative initiative unites 30 consortium members from 18 countries, representing leading educational and research institutions, industries, and SMEs. The project is strategically focused on the pillars of Adaptive, Green, Human-Centric, and Trustworthy AI, which are integral for the successful development, deployment, and acceptance of AI in Europe. ENFIELD's overarching goal is to strengthen the EU's competitive position in AI, bringing together top talents, technologies, and resources to address industry challenges in healthcare, energy, manufacturing, and space. The initiative aspires to deliver impactful research outcomes and significant socio-economic benefits for European citizens and businesses.

The project's objectives include developing and delivering over 75 unique AI solutions, spanning algorithms, methods, simulations, services, datasets, and prototypes. Complementary to these solutions, ENFIELD aims to contribute to academic discourse with 180 scientifically influential publications and 200 peer-reviewed presentations. The creation of strategic documents like the Common Research Roadmap and Vision, Safety and Security Risk Assessment Framework, White Paper, and Gender and Ethics Framework will further guide the project's trajectory.

ENFIELD's strength lies in its collaborative nature, fostering a dynamic European AI network. Consortium members bring diverse expertise to tackle critical issues in AI research and innovation collectively. The collaboration spans borders and sectors, creating a synergistic environment for advancing the field on a continental scale.

With a commitment to inclusivity, ENFIELD plans to support more than 76 individual researchers and 18 small-scale projects through Open Calls. This initiative will facilitate knowledge exchange, foster innovation, and contribute to a robust and diverse AI research community.

ENFIELD recognises the importance of education and outreach. Beyond research activities, the project will conduct summer schools and hackathons, providing skills development and knowledge exchange platforms. Strategic outreach methods will enhance community engagement, ensuring the project's expansion and long-term sustainability.

Purpose of the Data Management Plan

Aligned with the ambitious goals and collaborative spirit of the ENFIELD project, the Data Management Plan (DMP) is formulated to provide guidance for the responsible, ethical, and efficient handling of data. Embracing the FAIR principles—guaranteeing that data is Findable, Accessible, Interoperable, and Reusable—the DMP delineates strategies, policies, and practices to maintain the integrity, accessibility, and usability of data. This plan is essential for promoting transparency, accountability, and the enduring impact of the ENFIELD initiative. The incorporation of FAIR principles will be woven into every facet of data management, spanning collection and storage to sharing and preservation, ensuring that the data produced by ENFIELD yields maximal benefits for the broader scientific community.

2. Data Governance

2.1 Data Ownership

In the context of the ENFIELD project, the issue of data ownership is central to the responsible management of the wealth of information generated and utilised. The consortium collectively owns the data generated within the project, fostering a collaborative spirit where insights and knowledge are shared among members. While the consortium maintains ownership of the collective dataset, individual consortium members retain ownership of their contributed data, ensuring a balanced and equitable approach to data ownership.

2.2 Data Stewardship

The Data Management Officer (DMO) will be the primary data steward overseeing the end-to-end data management lifecycle. The DMO will be responsible for enforcing data governance policies, ensuring compliance with legal and ethical standards, and managing data quality, privacy, and security. Regular communication and collaboration with consortium members will be maintained to address emerging data challenges and opportunities.

2.3 Data Access and Sharing

The ENFIELD project highly values data access and sharing to encourage collaboration and foster innovation. Consortium agreements will delineate the terms of data access, specifying which consortium members have access to particular datasets based on their roles and responsibilities within the project. Mechanisms for controlled and secure data sharing with external entities, such as industry partners or regulatory bodies, will be established to facilitate broader collaboration while safeguarding sensitive information.

3. Data Collection and Generation

3.1 Types of Data

ENFIELD will generate various types of data, including but not limited to:

- Research data (algorithms, methods, simulations, services, prototypes)
- Publications and presentations
- Strategic documents (Common Research Roadmap and Vision, Safety and Security Risk Assessment Framework, White Paper, Gender and Ethics Framework)
- Educational materials (summer school and hackathon content)

3.2 Data Sources

Data will be sourced from various activities, including research experiments, surveys, and collaborative efforts among consortium members. The diverse origin of data ensures a comprehensive and multi-faceted approach to addressing the challenges outlined in the project objectives.

3.3 Data Quality

ENFIELD emphasises the importance of data quality to guarantee the dependability and accuracy of the insights generated. Stringent validation procedures, documentation norms, and commitment to optimal practices in data collection and processing will be implemented to uphold elevated standards of data quality.

4. Data Storage and Security

4.1 Storage Infrastructure

Data will be stored on secure servers provided by consortium members, adhering to industry-leading security protocols. Storage infrastructure will comply with relevant data protection regulations, ensuring the confidentiality and integrity of the data.

4.2 Encryption and Access Controls

Access controls will be implemented to restrict data access to authorised personnel based on their roles within the consortium, minimising the risk of data breaches. Furthermore, to safeguard against unauthorised access, sensitive data will be encrypted during storage and transmission.

4.3 Data Backup and Recovery

Regular and systematic backups will be conducted to prevent data loss. Robust procedures for data recovery will be established to mitigate potential risks and ensure the continuity of research activities.

5. Data Sharing and Preservation

5.1 Data Sharing Policy

ENFIELD embraces a data-sharing policy that encourages openness and collaboration within the consortium. Shared datasets will be made available through secure and controlled channels, with access protocols defined in consortium agreements.

5.2 Metadata and Documentation

Comprehensive documentation and necessary metadata will be created for each dataset, providing context and details necessary for understanding, interpreting, and reproducing the results. This commitment to documentation aligns with the FAIR principles of ensuring data is Findable and Reusable.

5.3 Data Preservation

ENFIELD is committed to preserving critical datasets for at least five years after the project's conclusion. Preservation efforts will include storing datasets in curated repositories and adhering to the FAIR principles of ensuring data is Accessible and Interoperable, thus maximising the potential for long-term impact and usability.

6. Ethical and Legal Considerations

6.1 Informed Consent

In alignment with ethical standards, ENFIELD will ensure that data collection processes adhere to informed consent principles. Participants will be fully informed about the nature of data collection activities and will have the opportunity to provide explicit consent.

6.2 Data Protection and Privacy

Data processing within the ENFIELD project will strictly adhere to relevant data protection regulations, safeguarding the privacy of individuals and entities. The project will implement privacy-preserving measures to uphold the rights of data subjects.

6.3 Intellectual Property

Intellectual property rights will be managed in accordance with consortium agreements, ensuring a fair and transparent allocation of rights among consortium members. Clear guidelines on intellectual property ownership and usage will be established to avoid disputes and foster collaboration.

7. Monitoring and Review

The Data Management Officer will play a proactive role in monitoring and reviewing data management practices regularly. This ongoing oversight ensures that the Data Management Plan remains adaptive to emerging standards, technologies, and project requirements. The FAIR principles will be integral to these monitoring and review processes, providing a framework to assess the findability, accessibility, interoperability, and reusability of project data. Adjustments to the plan will be made as needed to enhance its effectiveness and align with best practices in the

field of data management. This iterative approach ensures ENFIELD's data management practices remain robust, ethical, and aligned with the project's goals.

8. Integration of Open Science Practices

8.1 Orientation Phase

ENFIELD embraces open science practices throughout its orientation phase, seeking to integrate diverse knowledge production strategies within the Network of Excellence. The project strongly emphasises research perspectives, including Social Sciences and Humanities (SSH) dimensions, with knowledge, data, and tools sharing serving as core principles. This orientation aims to create a collaborative and inclusive environment where stakeholders contribute to shaping the research agenda and activities.

8.2 Implementation Methodology

The implementation methodology of ENFIELD is rooted in open collaboration with all relevant actors, stakeholders, and knowledge entities. The project seeks to benefit society by employing an open and participatory approach to research. ENFIELD interacts with multiple stakeholders within the four pillars and verticals, adopting a bidirectional process to incorporate diverse experiences and perspectives into operations and decision-making processes.

8.3 Co-Creation Approach

ENFIELD follows a co-creation approach, involving stakeholders actively in the early stages of the research process. This approach ensures that advanced solutions align with the needs and requirements of end-users. Through workshops, qualitative research methods, webinars, and information sessions, the consortium engages with stakeholders in a bidirectional manner to extract real-world needs and better understand their requirements and limitations.

8.4 Pseudo-Anonymisation for GDPR Compliance

All data gathered, mainly through qualitative methods such as surveys and workshops, will undergo pseudo-anonymisation before processing. This ensures compliance with the General Data Protection Regulation (GDPR) and safeguards the privacy of participants. Pseudonyms will be consistent within the research team and throughout the project, maintaining transparency and ethical standards.

9. Open Access to Research Outputs

9.1 Scholarly Open-Access Publishing

ENFIELD is committed to using the Open Research Europe scholarly fully open-access publishing service for Horizon Europe. This commitment facilitates rapid publication, ensuring outputs that support research integrity, reproducibility, and transparency. Scientific publications from the project will be made available through open-access options, with selected publications accessible through Golden Open Access. The project deliverables classified 'public' will be made available on the project website. As much as possible, project publications will also be made available on the OpenAIRE platform. Furthermore, project collaborators will exert every effort to disseminate project outcomes via their respective institutional open-access platforms, such as NTNU's https://ntnuopen.ntnu.no/

9.2 Publicly Available e-Trainings/Courses/Webinars

ENFIELD ensures the open sharing of project results through its website, institutional repositories, and other channels. Digital training, courses, and webinars produced by ENFIELD will be made publicly available through the project's website, YouTube, social media platforms, and other public services. This commitment extends to capacity-building activities, enhancing accessibility to knowledge generated by the project.

10. Research Data Management and Outputs

10.1 Types of Data/Research Outputs

ENFIELD will generate diverse research data and outputs, including experimental data from industry partners, aggregated datasets, logs, source code, technical reports, manuals, and qualitative data from surveys, workshops, webinars, and interviews. The project is committed to pre-processing data at the time of collection to remove personally identifiable information, ensuring anonymity and GDPR compliance.

10.2 Handling of Third-Party Data

For data obtained from third parties, primarily through Open Calls, a meticulous procedure will be defined for data access, processing, and further use. Confidentiality agreements among ENFIELD partners, Open Calls participants, and evaluators will be established to ensure the non-disclosure of sensitive data. The handling of third-party data will adhere to the FAIR principles, emphasising Findability, Accessibility, Interoperability, and Reusability.

10.3 FAIR Principles Implementation

ENFIELD commits to implementing the FAIR principles in data management:

- Findability: Using European Open Science Cloud and Open Research Europe for rapid publication and data access.

- Accessibility: Continuous integration of new databases and services, providing commercial data for non-commercial purposes within the project.
- Interoperability: Adherence to standards, formats, and vocabularies to ensure seamless interoperability.
- Reusability: Promoting open-source code, using Creative Commons licenses for research data and academic publishing, and applying self-archiving services for research communities.

11. Gender Integration and Bias Reduction in AI

11.1 Integration of Sex and Gender Dimensions

ENFIELD is committed to integrating sex and gender dimensions into foundational research and innovation developments across all four building blocks of the project. Special attention will be given to the Human-centric AI pillar and the four verticals, recognising that incorporating the gender dimension is crucial for achieving better research outcomes and increasing the acceptability of these outcomes. The consortium adopts an intersectional approach, understanding gender as non-binary and intersecting with other social, cultural, and economic dimensions.

11.2 Transversal Priority and Inclusive Framework

Integrating the gender dimension is considered a transversal priority in ENFIELD, permeating the entire project. An inclusive Gender and Ethics framework will be collaboratively implemented for the project's activities, involving Work Package and task leaders in the co-creation process. This framework will ensure that gender considerations are embedded in all aspects of the project, contributing to a more inclusive and ethically sound research environment.

11.3 Gender-Inclusive Outcomes

ENFIELD is dedicated to delivering gender-inclusive outcomes. The consortium commits to incorporating sex and gender dimensions in developing the EU-wide roadmap for research and innovation of Safe and Secure AI, as well as the White Paper on Safe and Secure AI made in EU. These efforts align with the findings and recommendations of the Gendered Innovations/Innovation through Gender Expert Group and are in accordance with the New European Research Area (ERA) for Research & Innovation objectives, the EU Gender Equality Strategy (2020-2025), EU LGBTIQ Equality Strategy 2020-2025, and the Ljubljana Declaration on Gender Equality in Research & Innovation.

11.4 Addressing Data Bias in ML Algorithms

Recognising the importance of data-centric machine learning (ML), ENFIELD focuses on researching and addressing data bias in ML algorithms. The project will investigate various types

of data biases, including sample bias, exclusion bias, observer bias, gender and racial bias, and association bias. Methods to reduce or eliminate these biases will be developed, and results concerning sex and gender aspects will be disseminated along with solutions for eliminating data bias in ML.

11.5 Non-Discrimination in Data Collection

ENFIELD will ensure sex and gender-based non-discrimination in data collection by utilising diverse datasets for ML algorithms. These datasets will be large and varied enough to represent different populations and contexts. Pre-processing techniques will be employed to reduce bias in the original data before training the model. Metadata related to gender, sex, ethnicity, age, and other factors will be included in the data preparation process.

11.6 In-Processing Methods for Bias Reduction

ENFIELD will use in-processing methods to address bias during model training, such as Classification with Fairness Constraints or Adversarial de-biasing. The model will undergo verification for fairness, ensuring the inclusion of different social groups, non-discrimination, and fair representation of sex and gender.

11.7 Research, Education, and Dissemination

Sex and gender aspects will be integrated as research subjects in all three verticals where appropriate. The research results related to gender dimension approaches will be included in published papers. Educational activities, including summer schools, hackathons, and dissemination activities like debates and webinars, will discuss sex and gender unbiased aspects. Gender considerations will also be incorporated into ENFIELD's Open Calls for TES and TIS, ensuring a holistic and inclusive approach across all project activities.

12. Green AI Pillar

This section outlines the strategies for managing, documenting, and sharing data generated during the research and development of the Green AI Pillar in ENFIELD. The DMP aligns with the FAIR principles (Findable, Accessible, Interoperable, and Reusable) to ensure transparency, accessibility, and reproducibility.

12.1 Data Collection and Generation

12.1.1 Data Sources

Data will be collected from various sources, including:

- Experimental trials and simulations on Green AI approaches.

- Performance metrics related to energy consumption, carbon emissions, and ML model efficiency.
- Edge device deployment and optimisation strategies.
- Symbolic AI and data-driven models for Hybrid AI in the edge-to-cloud continuum.

12.1.2 Data Types

The collected data will include:

- Performance metrics (energy consumption, carbon emissions, ML metrics).
- Experiment results and simulations.
- Symbolic constraints and optimisation parameters.

12.2 Data Organization and Documentation

12.2.1 Naming Conventions

Files and datasets will follow standardised naming conventions, including experiment details, date, and purpose, to facilitate easy identification and retrieval.

12.2.2 Metadata

Comprehensive metadata will be created for each dataset, providing details on the experiment, variables, algorithms used, and any specific characteristics relevant to the data.

12.2.3 Data Structure

Datasets will be organised into a clear folder structure, reflecting different aspects of Green AI research, such as edge deployment, metrics monitoring, and edge-to-cloud continuum orchestration.

12.3 Data Storage and Backup

Data will be securely stored in dedicated repositories with access controls. Regular automated backups will be implemented to prevent data loss. Multiple copies of critical datasets will be maintained.

12.4 Data Access and Sharing

12.4.1 Access Control

Access to data will be restricted based on project roles and responsibilities. Only authorised personnel will have access to specific datasets.

12.4.2 Data Sharing

Data will be shared among project members through secure communication channels. A data-sharing schedule will be established to ensure timely access to relevant datasets.

12.5 Data Preservation and Archiving

Data will be preserved for at least five years beyond the project's completion. Archiving will use standard formats and maintain metadata for future understanding and utilisation.

12.6 Expected Outcomes and Dissemination

12.6.1 Publications

Results and findings will be disseminated through at least 30 scientific journals and conferences. Publications will include details about developed Green AI tools, orchestration approaches, and metrics standardisation.

12.6.2 Open-Source Contributions

ENFIELD will contribute to open-source libraries and tools relevant to Green AI on the edge/edge-to-cloud continuum, promoting collaboration and community engagement.

12.7 Resources

The following resources will be utilised for data management:

- SINTEF's IoT lab for experimental edge devices
- Access to Amazon's data science servers with Nvidia A30 GPUs
- KNOW's big data technologies (Apache Hadoop, Apache Spark, Apache Storm, CUDA server)

12.8 Compliance

This Data Management Plan adheres to the FAIR principles, ensuring that data generated in the Green AI Pillar of ENFIELD is transparent, accessible, and reusable. Any deviations from this plan will be documented and communicated within the project team.

13. Adaptive AI Pillar

This section outlines the strategies and procedures for managing, documenting, and sharing data generated during the research and development of Adaptive AI in ENFIELD. The DMP adheres

to the FAIR principles (Findable, Accessible, Interoperable, and Reusable) to ensure data quality, transparency, and collaboration.

13.1 Data Collection and Generation

13.1.1 Data Sources

Data will be collected from various sources, including but not limited to:

- Experimental trials and simulations
- Brain-inspired algorithms development
- Edge device deployment challenges
- Incremental learning robustness and trustworthiness studies

13.1.2 Data Types

The collected data will include:

- Experimental results
- Brain-inspired algorithm models
- Edge device optimisation strategies
- Metrics related to Adaptive AI characteristics

13.2. Data Organization and Documentation

13.2.1 Naming Conventions

Consistent and standardised naming conventions will be established for files and datasets. Descriptive names will include information about the experiment, date, and purpose.

13.2.2 Metadata

Comprehensive metadata will be created for each dataset, including details on the experiment, variables, algorithms used, and any specific characteristics relevant to the data.

13.2.3 Data Structure

Datasets will be organised in a clear folder structure, reflecting the different stages of the Adaptive AI development process. This structure will facilitate easy navigation and retrieval of specific datasets.

13.3 Data Storage and Backup

Data will be stored securely in dedicated repositories with access control mechanisms. Regular automated backups will be implemented to prevent data loss. Multiple copies of critical datasets will be maintained.

13.4 Data Access and Sharing

13.4.1 Access Control

Access to data will be restricted based on project roles and responsibilities. Only authorised personnel will have access to specific datasets.

13.4.2 Data Sharing

Data will be shared among project members through secure communication channels. A data-sharing schedule will be established to ensure timely access to relevant datasets.

13.5 Data Preservation and Archiving

Data will be preserved for at least five years beyond the project's completion. Archiving will be done using standard formats, and metadata will be maintained for future understanding and utilisation.

13.6 Expected Outcomes and Dissemination

13.6.1 Publications

Results and findings will be disseminated through at least 30 scientific journals and conferences. Publications will include details about the developed algorithms, metrics, and outcomes.

13.6.2 Open-Source Contributions

ENFIELD will contribute to open-source libraries and tools relevant to Adaptive AI on the edge, promoting collaboration and community engagement.

13.7 Resources

The following resources will be utilised for data management:

- Open ML: Platform for sharing data, code, models, and experiments.
- Snellius: Dutch National Supercomputer for capability computing.
- HTI Labs: Laboratory infrastructure for human-technology interaction experiments.
- Teralab: AI-DIH for advanced computing infrastructure.

13.8 Compliance

This Data Management Plan complies with the FAIR principles and ensures transparency, accessibility, and reproducibility of the data generated in the ENFIELD project. Any deviations from this plan will be documented and communicated within the project team.

14. Human-Centric AI Pillar

This section outlines the strategies for managing, documenting, and sharing data generated during the research and development of the Human-Centric AI Pillar in ENFIELD. The goal is to ensure transparency, accessibility, and reproducibility of the research outcomes.

14.1 Data Collection and Generation

14.1.1 Data Sources

Data will be collected from various sources, including:

- Experimental trials and simulations on AI-based decision systems.
- Metrics related to transparency, accountability, and trust in AIS.
- User interactions and experiences in the loop of automated decisions.
- Field testing capturing user experiences.

14.1.2 Data Types

The collected data will include:

- Performance metrics of AIS systems.
- Explanation models and methodologies.
- User feedback and experiences.
- Metrics assessing interpretability and quality of explanations.

14.2 Data Organization and Documentation

14.2.1 Naming Conventions

Files and datasets will adhere to standardised naming conventions, including details such as experiment information, date, and purpose.

14.2.2 Metadata

Comprehensive metadata will be created for each dataset, providing details on experiments, variables, algorithms used, and specific characteristics relevant to the data.

14.2.3 Data Structure

Datasets will be organised into a clear folder structure reflecting different aspects of Human-Centric AI research, such as explanation methodologies, user interaction, and performance metrics.

14.3 Data Storage and Backup

Data will be securely stored in dedicated repositories with access controls. Regular automated backups will be implemented to prevent data loss. Multiple copies of critical datasets will be maintained.

14.4 Data Access and Sharing

14.4.1 Access Control

Access to data will be restricted based on project roles and responsibilities. Only authorised personnel will have access to specific datasets.

14.4.2 Data Sharing

Data will be shared among project members through secure communication channels. A data-sharing schedule will be established to ensure timely access to relevant datasets.

14.5 Data Preservation and Archiving

Data will be preserved for at least five years beyond the project's completion. Archiving will use standard formats, and metadata will be maintained for future understanding and utilisation.

14.6 Expected Outcomes and Dissemination

14.6.1 Publications

Results and findings will be disseminated through at least 40 scientific journals and conferences. Publications will include details about AI-based decision support systems, explanation methodologies, and human-AI shared decision-making.

14.6.2 Toolkits and Guidelines

A toolkit covering the AI model lifecycle and guidelines for ethical AI will be provided to developers, policymakers, and consumers. These resources will promote controllable and accountable AI and integrate bias mitigation.

14.7 Resources

The following resources will be utilised for data management:

- Access to SurfSara supercomputers (TU/e)
- Human-Technology Interaction Labs for empirical studies (TU/e)
- Data and analytics from telco network operations (TELENOR)
- Access to PRECIS Data Center – including DGX 4 A100 cluster (UPB)
- HPC infrastructure for testing developed methods and tools (NRS)

14.8 Compliance

This Data Management Plan adheres to best practices, ensuring that data generated in the Human-Centric AI Pillar of ENFIELD is transparent, accessible, and reusable. Any deviations from this plan will be documented and communicated within the project team.

15. Trustworthy AI Pillar

This section outlines the strategies for managing, documenting, and sharing data generated during the research and development of the Trustworthy AI Pillar in ENFIELD. The focus is on ensuring transparency, accountability, and the responsible use of AI technologies.

15.1 Data Collection and Generation

15.1.1 Data Sources

Data will be collected from various sources, including:

- Experiments assessing the robustness and safety of AI tools.
- Privacy-preserving variants of AI tasks and techniques.
- Cybersecurity risk management and threat landscape analysis.

15.1.2 Data Types

The collected data will include:

- Formal approaches to abstraction and modelling of uncertainty in AI tools
- Privacy-preserving AI solutions and their privacy guarantees
- Cybersecurity risk management framework and threat landscape data

15.2 Data Organization and Documentation

15.2.1 Naming Conventions

Standardised naming conventions will be adopted for files and datasets, incorporating details such as experiment information, date, and purpose.

15.2.2 Metadata

Comprehensive metadata will be created for each dataset, providing details on experiments, variables, algorithms used, and specific characteristics relevant to the data.

15.2.3 Data Structure

Datasets will be organised into a clear folder structure reflecting different aspects of Trustworthy AI research, such as risk-aware AI tools, benchmarks, privacy-preserving solutions, and cybersecurity analyses.

15.3 Data Storage and Backup

Data will be securely stored in dedicated repositories with access controls. Regular automated backups will be implemented to prevent data loss. Multiple copies of critical datasets will be maintained.

15.4 Data Access and Sharing

15.4.1 Access Control

Access to data will be restricted based on project roles and responsibilities. Only authorised personnel will have access to specific datasets.

15.4.2 Data Sharing

Data will be shared among project members through secure communication channels. A data-sharing schedule will be established to ensure timely access to relevant datasets.

15.5 Data Preservation and Archiving

Data will be preserved for at least five years beyond the project's completion. Archiving will use standard formats, and metadata will be maintained for future understanding and utilisation.

15.6 Expected Outcomes and Dissemination

15.6.1 Publications

Results and findings will be disseminated through at least 30 scientific journals and conferences. Publications will include details about risk-aware AI tools, benchmarks, privacy-preserving solutions, and cybersecurity analyses.

15.6.2 Community Building

Efforts will be made to build a thriving community on trustworthy AI, privacy-preserving AI, and cybersecurity of AI systems. Workshops and events will be organised within established international conferences and competitions.

15.7 Resources

The following resources will be utilised for data management:

- Access to AI tools, datasets, models, and other research resources from consortium partners
- Lab facilities of the Norwegian Centre for Cybersecurity of Critical Sectors (NORCICS) and the Center for Cyber and Information Security (CCIS)
- Data from Telenor operations for testing developed methods and tools

15.8 Compliance

This Data Management Plan adheres to best practices, ensuring that data generated in the Trustworthy AI Pillar of ENFIELD is transparent, accessible, and reusable. Any deviations from this plan will be documented and communicated within the project team.

16. Energy Vertical

This section outlines the strategies for managing, documenting, and sharing data generated during the research and development focusing on the Energy Vertical. The aim is to ensure transparency, accountability, and responsible use of AI technologies in addressing challenges and implementing novel solutions in the energy sector.

16.1 Data Collection and Generation

16.1.1 Data Sources

Data will be collected from various sources, including:

- Experiments assessing the application of AI in the energy sector, incorporating techniques such as artificial neural networks (ANN), rule-based systems, expert systems, and multi-agent systems (MAS).
- Data from the renewable energy and energy storage industry, including use cases like short-term forecasting, predictive maintenance, and energy battery dispatch.

16.1.2 Data Types

The collected data will include:

- Physics-based models of electrical grids and energy storage assets.
- AI-assisted decision support models for human operators under forecast uncertainty.
- Privacy-preserving machine learning techniques for smart meter data.

16.2 Data Organization and Documentation

16.2.1 Naming Conventions

Standardized naming conventions will be adopted, including details such as experiment information, date, and purpose.

16.2.2 Metadata

Comprehensive metadata will be created for each dataset, providing details on experiments, variables, algorithms used, and specific characteristics relevant to the data.

16.2.3 Data Structure

Datasets will be organised into a clear folder structure reflecting different aspects of the research, such as physics-based modelling, AI-assisted decision support, and privacy-preserving ML.

16.3 Data Storage and Backup

Data will be securely stored in dedicated repositories with access controls. Regular automated backups will be implemented to prevent data loss. Multiple copies of critical datasets will be maintained.

16.4 Data Access and Sharing

16.4.1 Access Control

Access to data will be restricted based on project roles and responsibilities. Only authorised personnel will have access to specific datasets.

16.4.2 Data Sharing

Data will be shared among project members through secure communication channels. A data-sharing schedule will be established to ensure timely access to relevant datasets.

16.5 Data Preservation and Archiving

Data will be preserved for at least five years beyond the project's completion. Archiving will use standard formats, and metadata will be maintained for future understanding and utilisation.

16.6 Expected Outcomes and Dissemination

16.6.1 Publications

Results and findings will be disseminated through at least 15 scientific journals and conferences. Publications will include details about AI-based decision prototypes, computational methods for optimisation, federated learning algorithms, and more.

16.6.2 Community Building

Efforts will be made to build a thriving community in the energy sector, organising workshops and events within established international conferences and competitions.

16.7 Resources

- Utilization of resources from INESC TEC Laboratory of Smart Grids and Electric Vehicles.
- EDP CNET's Laboratory of EV Charging Stations, and energy datasets from ISK.

16.8 Compliance

This Data Management Plan adheres to best practices, ensuring that data generated in the Energy Vertical of ENFIELD is transparent, accessible, and reusable. Any deviations from this plan will be documented and communicated within the project team.

17. Healthcare Vertical

This section delineates the strategies for managing, documenting, and sharing data generated during the research and development focusing on the Healthcare Vertical. The primary goal is to ensure transparency, accountability, and the responsible use of AI technologies in addressing challenges and implementing novel solutions in the healthcare sector.

17.1 Data Collection and Generation

17.1.1 Data Sources

Data will be collected from various sources, including:

- Health digitalisation records, sensory data, and simulated systems biology data.
- Clinical Big Data from various sources, including publicly available datasets like www.kaggle.com, www.sleepdata.org/, European Health Data Space, and potential access to health data sources from Norway (https://bbmri.no/health-data, https://helsedata.no/en/data-sources/).

- Data shared by envisaged partners and resources like TU/e Health Data Portal, ALAMEDA, and the Digital Twin Lab based on the Research Council of Norway's ASSET project.

17.1.2 Data Types

The collected data will include:

- Complex, heterogeneous, and noisy datasets relevant to healthcare.
- Adaptive AI approaches to limit the need for extensive data.
- Data related to Human-Centric AI, Trustworthy AI, and Responsible AI.

17.2 Data Organization and Documentation

17.2.1 Naming Conventions

Standardized naming conventions will be adopted for files and datasets, incorporating details such as experiment information, date, and purpose.

17.2.2 Metadata

Comprehensive metadata will be created for each dataset, providing details on experiments, variables, algorithms used, and specific characteristics relevant to the healthcare data.

17.2.3 Data Structure

Datasets will be organised into a clear folder structure reflecting different aspects of healthcare research, such as Responsible AI, Trustworthy AI, and Human-Centric AI.

17.3 Data Storage and Backup

Data will be securely stored in dedicated repositories with access controls. Regular automated backups will be implemented to prevent data loss. Multiple copies of critical datasets will be maintained.

17.4 Data Access and Sharing

17.4.1 Access Control

Access to data will be restricted based on project roles and responsibilities. Only authorised personnel will have access to specific datasets.

17.4.2 Data Sharing

Data will be shared among project members through secure communication channels. Open-source developments will be made public to fuel community collaboration.

17.5 Data Preservation and Archiving

Data will be preserved for at least five years beyond the project's completion. Archiving will use standard formats, and metadata will be maintained for future understanding and utilisation.

17.6 Expected Outcomes and Dissemination

17.6.1 Publications

Results and findings will be disseminated through at least 15 scientific journals and conferences. Publications will include details about the Responsible AI framework, adaptive AI approaches, and industrial use cases in healthcare.

17.6.2 Community Building

Efforts will be made to build a thriving community in the healthcare sector, organising workshops and events within established international conferences and competitions.

17.7 Resources

- Utilization of publicly available data from sources like www.kaggle.com, www.sleepdata.org/, European Health Data Space, and potential access to health data sources from Norway.
- Collaboration with partners for shared data and resources.
- Utilisation of specific platforms like TU/e's Health Data Portal, ALAMEDA, and the Digital Twin Lab based on the Research Council of Norway's ASSET project.

17.8 Compliance

This Data Management Plan adheres to best practices, ensuring that data generated in the Healthcare Vertical of ENFIELD is transparent, accessible, and reusable. Any deviations from this plan will be documented and communicated within the project team.

18. Manufacturing Vertical

This section outlines the strategies for managing, documenting, and sharing data generated during the research and development focusing on the Manufacturing Vertical. The goal is to ensure transparency, accountability, and the responsible use of AI technologies in addressing challenges and implementing novel solutions in the manufacturing sector.

18.1 Data Collection and Generation

18.1.1 Data Sources

Data will be collected from various sources, including:

- Industry 4.0 data from Cyber-Physical Production Systems, including data on Smart Autonomous, Collaborative Product Service, and Hyperconnected Factories.
- AI-driven Digital Transformation data, including machine vision systems, embedded robotic systems, real-time edge AI systems, knowledge discovery, decision support, and virtual assistants.
- Data related to the recent shift to Industry 5.0, focusing on Green, Adaptive, Trustworthy, and Human-centric AI.

18.1.2 Data Types

The collected data will include:

- Data capturing the AI edge-to-cloud continuum architectures in manufacturing.
- Data on the adaptivity and explainability of AI systems embedded in products.
- Human-centric AI data for both blue-collar and white-collar workers.

18.2 Data Organization and Documentation

18.2.1 Naming Conventions

Standardised naming conventions will be adopted for files and datasets, incorporating details such as experiment information, date, and purpose.

18.2.2 Metadata

Comprehensive metadata will be created for each dataset, providing details on experiments, variables, algorithms used, and specific characteristics relevant to manufacturing data.

18.2.3 Data Structure

Datasets will be organised into a clear folder structure reflecting different aspects of manufacturing research, such as Smart Autonomous Factory, Product-Service Factory, and Hyperconnected Factory.

18.3 Data Storage and Backup

Data will be securely stored in dedicated repositories with access controls. Regular automated backups will be implemented to prevent data loss. Multiple copies of critical datasets will be maintained.

18.4 Data Access and Sharing

18.4.1 Access Control

Access to data will be restricted based on project roles and responsibilities. Only authorised personnel will have access to specific datasets.

18.4.2 Data Sharing

Data will be shared among project members through secure communication channels. Open-source developments will be made public to foster community collaboration.

18.5 Data Preservation and Archiving

Data will be preserved for at least five years beyond the project's completion. Archiving will use standard formats, and metadata will be maintained for future understanding and utilisation.

18.6 Expected Outcomes and Dissemination

18.6.1 Publications

Results and findings will be disseminated through at least 10 scientific journals and conferences. Publications will include details about conceptual models for Industry 5.0 and AI, AI Maturity Models, contributions to AI4Europe catalogue, and experimental facility results.

18.6.2 Community Building

Efforts will be made to build a thriving community in the manufacturing sector, organising workshops and events within established international conferences and competitions.

18.7 Resources

- Utilization of data from industrial members and European Industrial Clusters in manufacturing domains.
- Collaboration with partners for shared data and resources.

18.8 Compliance

This Data Management Plan adheres to best practices, ensuring that data generated in the Manufacturing Vertical of ENFIELD is transparent, accessible, and reusable. Any deviations from this plan will be documented and communicated within the project team.

19. Space Vertical

This section outlines the strategies for managing, documenting, and sharing data generated during the research and development focusing on the Space Vertical. The objective is to ensure transparency, accountability, and the responsible use of AI technologies in addressing challenges and implementing innovative solutions in the space sector.

19.1 Data Collection and Generation

19.1.1 Data Sources

Data will be collected from various sources, including:

- Global Navigation Satellite Systems (GNSS) data (e.g., GPS, Galileo).
- Satellite Communication Systems data.
- Satellite-based Earth Observation (EO) data, including Big Earth Data from passive and active satellite sensors (e.g., satellite image-based earth observation for meteorology, disaster risk reduction, marine surveillance).
- Advanced Air Mobility (AAM) data, especially in terms of communication and navigation.

19.1.2 Data Types

The collected data will include:

- GNSS and satellite communication optimization data.
- Big Earth Data from satellite sensors.
- AAM sector data, focusing on communication and navigation.
- Data related to certification, regulation, energy consumption, and cybersecurity challenges.

19.2 Data Organization and Documentation

19.2.1 Naming Conventions

Standardized naming conventions will be adopted for files and datasets, incorporating details such as experiment information, date, and purpose.

19.2.2 Metadata

Comprehensive metadata will be created for each dataset, providing details on experiments, variables, algorithms used, and specific characteristics relevant to space data.

19.2.3 Data Structure

Datasets will be organized into a clear folder structure reflecting different aspects of space research, such as EO-based applications, satellite communication, and AAM scenarios.

19.3 Data Storage and Backup

Data will be securely stored in dedicated repositories with access controls. Regular automated backups will be implemented to prevent data loss. Multiple copies of critical datasets will be maintained.

19.4 Data Access and Sharing

19.4.1 Access Control

Access to data will be restricted based on project roles and responsibilities. Only authorized personnel will have access to specific datasets.

19.4.2 Data Sharing

Data will be shared among project members through secure communication channels. Guidelines for Space, EO, and AAM AI-based Solution Development will be made publicly available.

19.5 Data Preservation and Archiving

Data will be preserved for at least five years beyond the project's completion. Archiving will use standard formats, and metadata will be maintained for future understanding and utilization.

19.6 Expected Outcomes and Dissemination

19.6.1 Publications

Results and findings will be disseminated through at least 5 scientific journals and conferences. Publications will include details about the guidelines for AI-based solutions, AI/ML algorithms for predicting loss of satellite communications, and AI algorithms for EO-based applications.

19.6.2 Community Building

Efforts will be made to build a thriving community in the space sector, organising workshops and events within established international conferences and competitions.

19.7 Resources

- Specific datasets will be created based on L-band satellite propagation models for initial algorithm training.

- Boeing's ground and flight test campaigns results will be evaluated.
- Boeing's Big Data Cluster tailored to Air Traffic Analytics (ADAPT) will be considered.
- ECoE's high-spec server and infrastructure of the EO Satellite Data Acquisition Station (DAS) will be utilized for processing Big Earth Data from optical and SAR sensors.

19.8 Compliance

This Data Management Plan adheres to best practices, ensuring that data generated in the Space Vertical of the Trustworthy AI Pillar of ENFIELD is transparent, accessible, and reusable. Any deviations from this plan will be documented and communicated within the project team.

20. Local Data Management Plan and Data Protection Officer

All ENFIELD consortium partners are dedicated to adhering to relevant sections of the Consortium Agreement, Articles in the Grant Agreement, bilateral agreements for data management, relevant national and European Union's rules and regulations. This Data Management Plan prioritises compliance with both the General Data Protection Regulation (GDPR) and the FAIR principles. Under the GDPR, each legal entity holding personal data is responsible for its management and must appoint a Data Protection Officer (DPO). Consequently, every partner organisation within ENFIELD is mandated to establish a data protection/data management plan for personal data management, including procedures for regular reviews, updates, and the appointment of a DPO. Partners will conduct Data Protection Impact Assessments (DPIAs) as required by the GDPR, under the supervision of their respective organization's DPO. Furthermore, the responsibility of Data Protection Officer and Data Management Officer could be a shared responsibility or two different individuals in the positions of Data Protection Officer and Data Management Officer.

ENFIELD partners possessing data ownership must ensure the existence of respective data management plan that adheres to the FAIR principles. All partners are expected to fully comply with GDPR and FAIR principles, including the establishment of data management plans covering both sensitive and non-sensitive personal data, along with procedures for DPIAs. When necessary, partners should collaborate with their organisation's DPO to conduct DPIAs. Compliance with FAIR principles may be outlined within the partner level data management plan or another organisational policy document.

ESF is tasked with overseeing the implementation of the Gender and Ethical framework within the project, ensuring partners' awareness and compliance.

References

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

European Commission, Ethics and data protection, 14 Nov 2018, https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/ethics/h2020_hi_ethics-data-protection_en.pdf

European Commission, EU data protection rules, https://commission.europa.eu/law/law-topic/data-protection/eu-data-protection-rules_en

European Commission, European AI Alliance, https://futurium.ec.europa.eu/en/european-ai-alliance/pages/official-documents-and-reports

European Commission, A European approach to artificial intelligence, https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence