



AI-Ops Framework for Automated, Intelligent and Reliable Data/AI Pipelines Lifecycle with Humans-in-the-Loop and Coupling of Hybrid Science-Guided and AI Models



WP7: Project Coordination, Scientific, Technical and Innovation Management

D7.2: Data Management Plan

Deliverable Leader: ATHENA

Due Date: Jun 30, 2024 (M6)

Dissemination Level: Public

Version: v2.0










This deliverable presents the preliminary Data Management Plan for the AI-DAPT project. It defines the strategy to manage all research data used or generated by the project under EC guidelines. Research data that is expected to be used or generated by the project are identified and categorized, while guidelines are provided for data handling and publication methodology. The project's Ethics-and-Fairness-by-design approach towards the construction of the AI-DAPT Trustworthy framework is also explained in this document.

Further Information: <http://www.ai-dapt.eu/>

Disclaimer. The views represented in this document only reflect the views of the authors and not the views of the European Union. The European Union is not liable for any use that may be made of the information contained in this document. Furthermore, the information is provided "as is" and no guarantee or warranty is given that the information is fit for any particular purpose. The user of the information uses it at its sole risk and liability.



Partners

	ATHINA- Erevnitiko Kentro Kainotomias stis Technologies tis Pliroforias, ton Epikoinonion kai tis Gnosis (ATHENA)	Greece
	Fraunhofer Gesellschaft zur Forderung der Angewandten Forschung ev (FRAUNHOFER)	Germany
	South East Technological University (SETU)	Ireland
	Uninova-Instituto de Desenvolvimento de Novas Tecnologias (UNINOVA)	Portugal
	Universitat Politecnica de Catalunya (UPC)	Spain
	Suite5 Data Intelligence Solutions Limited (Suite5)	Cyprus
	MCS Datalabs (MCS)	Germany
	Eyfyia gia Epicheiriseis Etaireia Periorismenis Evthinis - Intelligence For Business Ltd (WITSITE)	Greece
	University of Cyprus (UCY)	Cyprus
	S&D Consulting Europe SRL (S&D)	Italy
	Gioumpitek Meleti Schediasmos Ylopoiisi kai Polisi Ergon Pliroforikis Etaireia Periorismenis Efthynis (UBITECH)	Greece
	Charite - Universitaetsmedizin Berlin (CHARITE)	Germany
	BIBA - Bremer Institut fuer Produktion und Logistik GmbH (BIBA)	Germany
	OHS Engineering GmbH (OHS)	Germany
	Etaireia Promitheias Aeriou Thessalonikis - Thessalias Monoprosopi Anonymos Etaireia (ZENITH)	Greece
	DOMX Idiotiki Kefalaiouschiki Etaireia (DOMX)	Greece
	Made Scarl (MADE)	Italy
	Consorzio Intellimech (IMECH)	Italy



Document Log

Contributors	Eleni Lavasa (ATHENA) Marina Cugurra (S&D) Theodore Dalamagas (ATHENA)
Internal Reviewer 1	Dimitris Skoutas (ATHENA)
Internal Reviewer 2	Carlos Agostinho (UNINOVA)
Type	Report
Delivery Date	M06

History

Versions	Description
v0.1	Full draft version including contributions from both T7.3 (ATHENA) and T7.4 (S&D)
v1.0	Revision of internal reviewers 1 and 2
v2.0	Final



Executive Summary

This deliverable outlines the initial Data Management Plan (DMP) for the AI-DAPT project, focusing on strategies for handling research data in compliance with EU FAIR principles. It details the types of data to be generated, collection and storage processes, relevant standards, and privacy and ethical guidelines. The DMP is a dynamic document that will be updated as the project progresses. Key impacts include promoting a data-centric approach and ensuring compliance across the consortium, which includes various demonstrator partners. The DMP is created using the Argos platform, facilitating collaborative data management and promoting data findability, accessibility, interoperability, and reusability.

The document also addresses the relevant ethical and legal frameworks, outlines the roles and responsibilities for ethics management, and provides a preliminary analysis of ethical considerations for each demonstrator. Data privacy and security are emphasized, adhering to GDPR and other regulations. The project aims to make research outputs widely accessible while protecting sensitive data and includes measures for intellectual property management and open science practices. The DMP will continue to evolve with detailed dataset descriptions and ongoing ethical assessments as the project advances.



Table of Contents

Executive Summary.....	iv
1 Introduction.....	1
1.1 AI-DAPT Project Overview	1
1.2 Deliverable Purpose and Scope	1
1.3 Impact and Target Audiences.....	2
1.4 Deliverable Methodology	2
1.5 Dependencies in AI-DAPT and Supporting Documents	3
1.6 Document Structure.....	3
2 AI-DAPT Management of Research Data.....	4
2.1 Data Summary.....	4
2.1.1 Demonstrator data.....	5
2.1.2 Project data.....	5
2.2 Findability	6
2.3 Accessibility.....	6
2.4 Interoperability.....	7
2.5 Reusability	7
2.6 Privacy & Security considerations	8
2.7 IPR management.....	8
2.8 Allocation of resources	9
3 Ethics	10
3.1 Assigned roles and responsibilities	12
3.2 Ethical considerations in the AI-DAPT demonstrators	13
4 Data management in AI-DAPT demonstrators	14
5 Conclusions and Next Steps.....	15
References	16
List of Acronyms/Abbreviations.....	17
Annexes	18
Annex I.....	18



List of Figures

FIGURE 1. OVERVIEW OF THE RESEARCH DATA MANAGEMENT STRATEGY IN AI-DAPT..... 4

FIGURE 2. THE AI-DAPT H2020 COMMUNITY WAS ESTABLISHED IN ZENODO..... 6

FIGURE 3. TRANSVERSAL NATURE OF THE LEGAL AND ETHICAL ASPECTS IN AI-DAPT 10

FIGURE 4. AI-DAPT DMP IN ARGOS 14

FIGURE 5. DATASET DESCRIPTION IN ARGOS..... 18



List of Tables

TABLE 1. ETHICS-RELATED ROLES AND THEIR RESPONSIBILITIES. 12



1 Introduction

1.1 AI-DAPT Project Overview

Today, Artificial Intelligence (AI) has paved a long way since its inception and has started experiencing exponential growth across various industries and shaping our world in ways that were once thought impossible. As AI transitions from research to deployment, leveraging the appropriate data to develop and evaluate AI models has evolved into one of its greatest challenges. Data are in fact the raw material and the most indispensable asset fuelling much of today's progress in AI, generating previously unattainable insights, assisting more evidence-based decision-making, and bringing tangible business/economic benefits and innovation to all involved stakeholders. However, despite their instrumental role in determining performance, fairness, and robustness of AI systems, data are paradoxically characterised as the most under-valued and de-glamorised aspect of AI while a data-centric focus is typically lacking in the current AI research.

AI-DAPT aims to deliver an innovative and impactful research agenda that will provide tangible benefits to a variety of stakeholders that struggle with making AI services. Seeking to reinstate the pure data-related work in its rightful place, and reinforcing the generalizability, reliability, trustworthiness, and fairness of AI solutions, AI-DAPT vision relies on the implementation of an AIOps framework to support and automate AI pipelines that continuously learn and adapt based on their context. It enables proper purposing, collection, documentation, (bias) valuation, annotation, curation and synthetic generation of data, while keeping humans-in-the-loop across five axes: (i) Data Design for AI, (ii) Data Nurturing for AI, (iii) Data Generation for AI, (iv) Model Delivery for AI, (v) Data-Model Optimization for AI.

AI-DAPT brings forward a two-fold data-centric mentality in AI:

- **Data:** AI-driven automation for data pipelines based on Explainable AI (XAI) techniques as well as synthetic data generation and observability.
- **Model:** Automation on AI model building and hybrid science-AI solutions, bringing together data-driven AI models and science-based (first-principles) models that build on high-quality data.

Bridging the gap between data-centric and model-centric AI, AI-DAPT will turn over a new leaf in trustworthy AI and will nurture an ecosystem involving all AI and data value-chain stakeholders. The aim is to enhance their prosperous collaboration in order to deliver and apply innovative AI-driven methods that rely on smart and dynamic end-to-end automation of data, AI training/inference pipelines in the cloud-edge computing continuum.

To demonstrate the actual innovation and added value that can be derived through the AI-DAPT scientific advancements, the AI-DAPT results will be validated in two ways:

- By applying them to tackle real-world challenges in four (4) key industries: Health, Robotics, Energy, and Manufacturing.
- By integrating them into various AI solutions, whether open source or commercial, already present in the market.

1.2 Deliverable Purpose and Scope

This deliverable presents the first version of the project's Data Management Plan (DMP), as the outcome of tasks T7.3 – "Research Data Management" and T7.4 "Legal and Ethical Management Activities", in M6 of the project, within activities of WP7 – Coordination". The DMP establishes the strategy for the management of research data and outputs within the AI-DAPT project. This strategy



is in line with EU guidelines on FAIR (Findable, Accessible, Interoperable and Reusable) data management [<https://www.go-fair.org/fair-principles/>]. Within the scope of the DMP is to:

- document information on the types, formats and other properties of the data expected to be generated or handled by the project,
- define reliable and secure processes for collecting/storing/sharing/accessing the data,
- identify relevant standards/vocabularies for data/metadata,
- identify data and other research outputs that can be made publicly available,
- explain the rules and methodologies that partners will have to follow to ensure compliance with privacy and ethical regulations,
- provide publication guidelines and support in IPR management.

This is a preliminary documentation of the DMP, since not all relevant information is available at this early stage of the project (M6). The DMP will be updated with new information in the course of the project, to reflect progress in research and development activities and monitor adherence to the established data management rules and practices.

1.3 Impact and Target Audiences

The DMP lays out the rules and methodologies for managing all research data and outputs within the AI-DAPT framework. These should be followed by all project participants, therefore the DMP is addressed to the entire AI-DAPT consortium, including demonstrator, technical and administrative partners alike. AI-DAPT introduces a data-centric mentality and intends to develop novel methods for AI-based automation in data processing pipelines. A well-structured, regularly updated DMP, as well as compliance to it by all involved parties, are anticipated to have major impact in achieving these goals.

1.4 Deliverable Methodology

The main points to be taken into account in the design of the DMP, highlighted in the project's Description of Action (DoA), were considered while creating this deliverable. Using this preliminary description as a roadmap, the project-level strategy towards data management was constructed and documented. In this process, the ethical and legal framework established in D6.1 has been considered. Then, the DMP was created in the Argos¹ platform, developed by the OpenAIRE EU project. This publicly available community tool supports the creation, collaborative processing and maintenance of DMPs for EU-funded projects. Apart from the project-level inputs (e.g. name, number of GA, short description), the created plan supports the generation of individual descriptions for each dataset (in general, digital asset) expected to be handled or generated in the project. The description templates provided by the tool address every relevant aspect of data management under the FAIR principles, including privacy, security and ethical considerations. Once finalized, individual descriptions as well as the entire plan can be exported in various formats (including docx, pdf, xml format). The tool also supports versioning, to keep track of the evolution of the DMP along the progress of the project.

An initial registry of demonstrator data sources was retrieved from the results of questionnaires circulated by UCY partners in the context of WP1 activities. These results, along with rich context information on the demonstrator use cases and requirements are presented in deliverable (D1.1, 2024) "AI-DAPT Automated AI Pipeline End User Needs and Scientific & Technology Radar". For each of the datasets listed therein by each demonstrator, a corresponding description was created within the project's DMP in Argos, and the available information on the dataset was filled. Then, demonstrator partners were asked to fill in the descriptions in more detail, following the structure of the template. They were provided with the option to do so either directly in the Argos platform, or in

¹ <https://argos.openaire.eu/>



the corresponding exported documents. Under this approach, demonstrator-specific inputs to the DMP were acquired. In this process, the preliminary analysis of demonstrator-specific ethical considerations presented in D1.1 was taken into account.

1.5 Dependencies in AI-DAPT and Supporting Documents

The first version of the DMP for the AI-DAPT project, presented in this deliverable D7.2, has been implemented in direct consideration of:

- The preliminary version of the DMP outlined in the DoA,
- The ethical and legal framework established in deliverable (D6.1, 2024) - “Dissemination, Communication, Engagement & Innovation Plan”, in section 2.4.5.2 “Regulatory and Ethical Compliance”,
- Information on demonstrator data sources, privacy and ethical constraints presented in deliverable (D1.1, 2024) - “AI-DAPT Automated AI Pipeline End User Needs and Scientific & Technology Radar”.

Subsequent versions of the DMP will be created and maintained in the Argos platform, where they can be publicly available.

1.6 Document Structure

The deliverable, after this introductory part in Section 1, is structured as follows:

- Section 2 elaborates on the adopted strategy for the management of research data in AI-DAPT. An overview and categorization of the anticipated assets is provided, while FAIR practices in their management are explained. The project’s approach to ensure the privacy and security of these assets, as well as IPR management, is also presented in this section.
- Section 3 discusses ethical and legal aspects of data management in AI-DAPT and introduces the Ethics-and-Fairness-by-design approach towards designing and implementing the AI-DAPT Trustworthy Framework. The allocation of relevant roles and responsibilities to support this approach, as well as an initial assessment of legal and ethical considerations in the demonstrators, are included in this section.
- The construction of the DMP per demonstrator in the Argos platform is presented in Section 4, while the actual output of the tool is included in Annex I of the document.
- Finally, concluding remarks are drawn and next steps are discussed in Section 5.



2 AI-DAPT Management of Research Data

The AI-DAPT strategy towards research data management is constructed under the guidelines² provided by the European Commission on FAIR (Findable, Accessible, Interoperable and Reusable) Data Management in Horizon 2020. The FAIR principles (Wilkinson, et al., 2016) were established in 2016 by FORCE 11 group³, to determine the best practices in research data preparation, in a human- and machine-readable format. AI-DAPT is a data-intensive project, in the sense that it will generate, process and handle data from various sources, in order to develop and validate novel AI-based automation methods in data and model pipelines. FAIR data management is therefore a key factor in achieving the desired high level of automation, as well as in assisting human-machine collaboration. Moreover, AI-DAPT supports open access policy for the project's research outputs, in accordance with the H2020 guidelines⁴ regarding Open Access to Scientific Publications and Research Data. At the same time, special attention is paid to the protection of personal and corporate data in compliance with GDPR and relevant acts, and secure data exchange within AI-DAPT is guaranteed. In the rest of this section, a brief summary of the data sources identified at this early stage is provided, and FAIR practices in the AI-DAPT management of this data are specified.

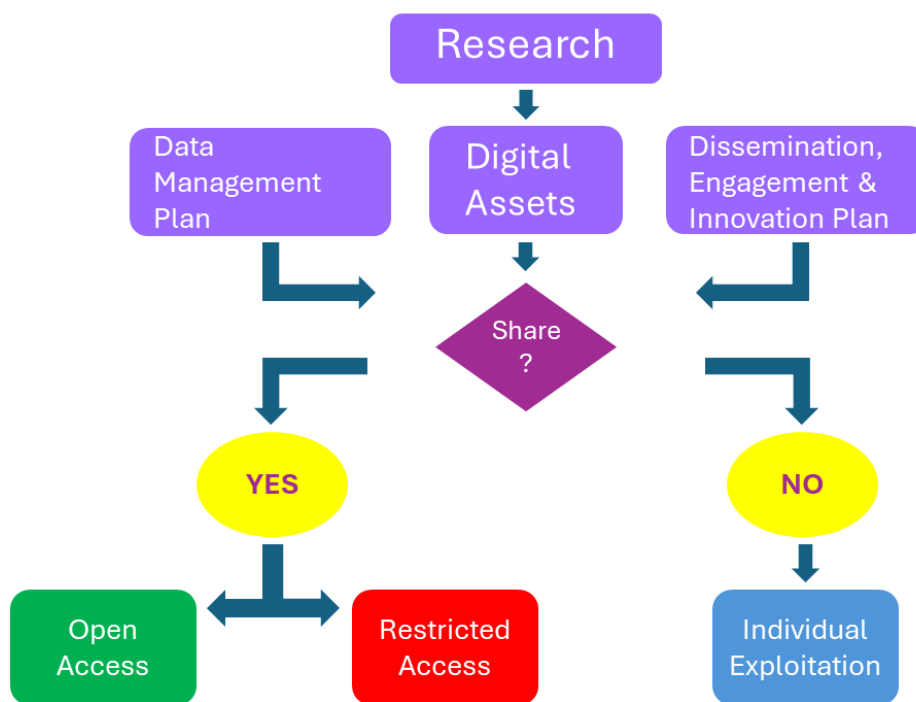


Figure 1. Overview of the Research Data Management strategy in AI-DAPT.

2.1 Data Summary

Data sources in AI-DAPT include existing datasets from demonstrator partners, data to be collected by demo partners for the pilot's needs, derivative and synthetic data to be generated during the

² European Commission, Directorate-General for Research & Innovation (2016). Guidelines on FAIR Data Management in Horizon 2020

³ <https://force11.org/>

⁴ European Commission. Horizon 2020 Programme. Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020.



project, and open-source datasets. Additional research outputs such as AI models and software code are anticipated as well.

2.1.1 Demonstrator data

Detailed descriptions of demonstrator use cases and overall data sources have been presented in (D1.1, 2024) and are briefly summarized here for completeness, to provide the reader with the necessary context for the DMP and especially the privacy and ethical considerations involved. Then, the specific description, based on the Argos H2020 template, of each dataset expected by the demonstrators, can be found in the Annex.

Demonstrator 1: Health - Personalized Medicine Based on Non-invasive Glucose Monitoring

The first demonstrator, involving MCS and CHARITE, aims to use Photoplethysmography (PPG), which detects blood volume changes using optical technology, combined with Machine Learning, to non-invasively monitor blood glucose levels in diabetic patients.

Their existing "MCS - Glucose Dataset" contains PPG metrics for 10 subjects in time-series format, with glucose levels and blood pressure recorded every ten minutes. A new dataset, with PPG metrics, motion data, and demographic information from 600 tests on 200 patients, including both diabetic and non-diabetic individuals, is planned to be generated by CHARITE as the outcome of a new clinical study, with support and equipment by MCS partners. An additional open-source dataset has been depicted in D1.1, that could be potentially combined with the above.

Demonstrator 2: Robotics & Cognitive Ergonomics - Human-Centered Automation

The second demonstrator, involving IMECH and MADE, focuses on robotics and cognitive ergonomics, with the aim to optimize working conditions through effective human-machine collaboration. To achieve this, they plan to monitor the working environment and the employees' emotions and mental state using wearables, and then analyse the collected bio-signals and insights to detect stress in operators collaborating with robots in the shop floor.

Since there is no pre-existing dataset for this demo, both IMECH and MADE are in the process of generating one (each). Currently, the necessary equipment (i.e. wearable devices) is being purchased and tested. Several open-source datasets for the detection of stress from wearable data have been identified and the possibility to use them alongside the anticipated IMECH and MADE datasets is being investigated.

Demonstrator 3: Energy - Cross-vector Residential Demand-Response (DR) Through Smart Heating

The third demonstrator, involving ZENITH and DOMX, is focused on improving personalized load and price forecasts in energy consumption, as well as enhancing the accuracy of demand response predictions, ultimately reducing peak load.

Several datasets are already available by both partners, including time-series (measured and synthetic) energy consumption datasets, static datasets with characteristics of the corresponding population, individual and large-scale aggregated gas/power consumption, as well as openly available energy and gas trading data. Some relevant open datasets have been retrieved, that could be used along with the existing ones.

Demonstrator 4: Manufacturing - Predictive Maintenance of Production Assets

The fourth demonstrator, involving OHS and BIBA, aims to improve manufacturing maintenance quality and efficiency, detect events to reduce costs, and enhance predictive maintenance services.

Both static and dynamic datasets on maintenance processes and the involved equipment and dedicated efforts are available by OHS, while relevant open-source datasets have been identified and have been catalogued in D1.1.

2.1.2 Project data



In the context of research and development activities, AI-DAPT will generate new data and other digital assets, including derivative/compiled data, synthetic data and hybrid science-guided and AI models. Derivative/compiled data may involve joining selected, non-sensitive features from demo data with open-source or synthetic data. Synthetic data generation is an integral part of AI-DAPT research and development activities, so multiple synthetic datasets are expected to be handled by the project. Additional digital assets are anticipated such as hybrid science-guided and AI models addressing the demonstrator scenarios, as well as XAI methods for data & AI pipeline automation, in the form of software code. Software documentation and the project's public deliverables and scientific publications are also amongst AI-DAPT's expected research data.

2.2 Findability

The data collected and generated in AI-DAPT will be identifiable and locatable through unique identification mechanisms, such as Unique Identifiers (UIDs), Uniform/International Resource Identifiers (URIs/IRIs) and Digital Object Identifiers (DOIs). Files will follow standardized naming conventions and versioning. All data will be inventoried and annotated with appropriate metadata according to AI-DAPT standards and those promoted by the European Data Portal⁵, using the DCAT Application Profile for metadata. Metadata will be published alongside the data in a machine-readable format, using standard terms.

AI-DAPT will use the Zenodo repository maintained by CERN⁶ to ensure the project's research data is findable. A community has been established for AI-DAPT on Zenodo (<https://zenodo.org/communities/ai-adapt/>), where all public datasets and scientific publications generated by the project will be uploaded (Figure 1). These materials will also be registered with the European Commission Funded Research (OpenAIRE) community, as well as uploaded in the project's web site (<https://www.ai-dapt.eu>) to maximize findability. Research outputs uploaded in Zenodo will be assigned DOIs and enriched with metadata, including Grant Number and Project Acronym. Additionally, the repository will provide version control.

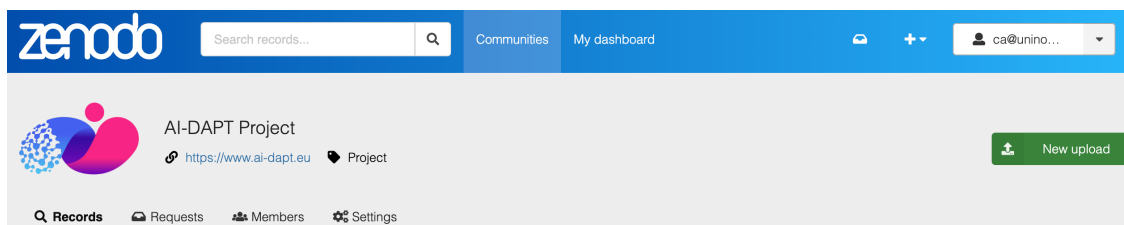


Figure 2. The AI-DAPT H2020 community was established in Zenodo.

2.3 Accessibility

During the project's lifetime, information on the following aspects will be detailed for each dataset before determining if it will be openly accessible: a) the nature and scale of the data, b) its potential usefulness to others, c) the existence of similar data and potential synergies, and d) the possibility for integration and reuse by external users. Overall, AI-DAPT's research data will be made available, when possible, without compromising privacy, ethical considerations, or commercial sustainability, to parties with a legitimate research interest. The project will ensure immediate open and free access to peer-reviewed scientific publications. Machine-readable copies of published or final peer-reviewed manuscripts will be made openly accessible either by the scientific publisher (under a "gold" access model supported by most academic journals) or via trusted repositories such as the EC's Open

⁵ <https://data.europa.eu/en>

⁶ Conseil Européen pour la Recherche Nucléaire, or European Council for Nuclear Research



Research Europe portal and Zenodo, as well as repositories operated by research partners (ATHENA, FHG, SETU, UCY, UPC). Additionally, the AI-DAPT website (<https://www.ai-dapt.eu>) will provide access to public deliverables and other project materials after EC approval, while research assets such as AI models will be accessible at the AI -on-Demand platform⁷. Open-source code will be made available on GitHub.

Certain pilot data, especially from demo sites, will remain confidential due to privacy concerns. This is due to the extreme added value of these data for the given company, to avoid compromising the commercial interests of the business partner in the highly competitive contemporary market. This data will only be shared with trusted parties such as the technical support partners of the demonstrators. The AI-DAPT platform will provide a secure and trustworthy environment to do so.

2.4 Interoperability

AI-DAPT will place particular effort in making data to be used in the project interoperable, since this is a necessary requirement to achieve automated and AI-based mining, harvesting, selection, cleaning, annotation, and/or enrichment/augmentation of data for AI. A preliminary identification of the structure/model of the data that are to be handled within the AI-DAPT framework, is done with use of the inputs by demo partners in the Dataset descriptions generated in Argos. Any domain-specific, open standards/ontologies/data models that demo partners already apply to their (meta)data are queried and will be adopted/expanded within the AI-DAPT framework. In parallel, within the project's standardization activities, relevant open standards/ontologies/data models to the pilot domains are investigated. Combining the results of both tasks, AI-DAPT will proceed to align with several widely adopted open standards/ontologies/data models and build on those to establish the AI-DAPT metadata standard. In this process, demonstrator data that did not already follow a standard schema, will be mapped to open and widely used ontologies.

2.5 Reusability

The outcomes generated by the project will be made available with Creative Commons Licenses (CC BY, CC0 or equivalent), and data re-usability will be facilitated for both confidential and public datasets. Once a data asset is marked as public, and, therefore, made available in the aforementioned repositories (in Sections 2.2 & 2.3) it will be fully reusable. Further promoting the reusability and aiming to improve the reproducibility of results, open science practices will be incorporated either as part of demonstrator activities (WP5 - Demonstration, Benchmarking, Business Validation and Impact Assessment) or as dedicated tasks within the AI-DAPT communication and dissemination strategy (WP6 - Dissemination, Communication, Stakeholder Engagement, Exploitation).

More specifically, an early and open sharing strategy of research outputs will be pursued. Early information about the design specifications and architecture of the AI-DAPT platform and components for the AI models will be shared as preprints in repositories such as Zenodo, Preprints, and arXiv before being published in open-access journals. Although most journals allow preprint sharing prior to publication, the consortium will ensure that the policies of the targeted journals permit this practice without jeopardizing future publication by using the policy check services of Sherpa Romeo⁸.

For activities focusing on applied research, such as the design of hybrid physical/ML models, federated data sharing mechanisms, advanced ML-driven data curation methods, and data sovereignty mechanisms, the consortium will promote the pre-registration of relevant approaches in targeted repositories like OSF and AsPredicted to facilitate early validation of research hypotheses. In parallel selected AI models will be openly published (e.g. as Jupyter Notebooks) in the AI-on-Demand platform

⁷ <https://www.ai4europe.eu>

⁸ <https://www.sherpa.ac.uk/romeo/>



and in GitHub, being therefore available to external researchers and data scientists for further elaboration and improvement.

Adding to that, the consortium will establish mechanisms to engage external stakeholders in the ongoing co-creation and co-validation of project results. These activities will allow external stakeholders to access and experiment with services for free and provide feedback. This open access will be available during the project's implementation. After the project concludes, open access will be maintained for Open-Source components under a straightforward licensing scheme (e.g., Apache License 2.0). Participation in open peer-review processes will be encouraged through the publication of research results on the Open Research Europe portal, which promotes open peer-review, a process with which AI-DAPT partners are already familiar. The consortium will adopt an inclusive approach of involving demonstrators as data providers and their end-users/clients in the co-creation and co-validation of the project's results.

To further ensure reproducibility and reinforce Open Science practices, AI-DAPT will implement several measures. It will avoid unnecessary duplication by utilizing existing state-of-the-art technologies and extensively validate project results in demonstrators, reporting findings, lessons learned, and highlighting any weaknesses or inaccuracies. Quality will be maintained through a well-established assurance process supervised by the Technical Coordinator. Additionally, AI-DAPT will establish synergies with other projects to validate data exchanges between platforms and develop a concrete replication plan for its components based on validation findings.

2.6 Privacy & Security considerations

The collection, processing, and sharing of personal data will be conducted in accordance with GDPR, Data Governance Act and AI Act towards fortifying security and privacy of personal and sensitive data and offering data providers with the necessary tools to have full control of their data and AI pipelines.

AI-DAPT will focus on creating secure and privacy-preserving solutions by design, ensuring that data subjects can confidently share their data. Processes and mechanisms to meet security and privacy requirements in data collection activities within AI-DAPT include the anonymization of personal/sensitive data, and data storage/sharing in secure and trustworthy repositories such as the AI-DAPT repository hosted by the project coordinator at internal servers of ATHENA RC. State-of-the-art firewalls, network security, encryption and authentication are used to protect data stored/shared in this repository. Upon development of the AI-DAPT platform and the use of additional servers provided by the institutions of project partners, or cloud services provided by AWS (Amazon Web Services), secure data storage and sharing therein will be ensured with use of similar advanced mechanisms.

2.7 IPR management

All consortium members are committed to contributing their background knowledge for the project's success. To ensure mutual respect for each other's intellectual property (IP) rights, partners with patents, copyrights, or other protected IP necessary for the project will provide them on fair and reasonable terms. Partners will only use another partner's protected items (products, information, source code, etc.) when the licensing conditions have been clearly communicated.

For intellectual property resulting from the project (foreground IP), partners agree to co-own based on their contributions to its design and development. Appropriate IP measures will be selected by the co-owning parties. If necessary, the consortium will use services like the Horizon IP Scan from the EC IP Helpdesk to develop a joint IP management strategy and prevent conflicts, enhancing the exploitation plan for project results. Partners will respect each other's rights during the dissemination of project results, informing the consortium and Management Team about planned publications or



activities that might reveal sensitive information, allowing time for protection measures to be established.

Two key issues are considered: a) "freedom to operate," ensuring AI-DAPT and its outputs can be exploited without infringing existing patents, and b) the ability to patent the project's innovations. A preliminary patent search indicates that AI-DAPT's key features are not currently patented or legally protected, providing the consortium with the freedom to protect itself and deter competition. Careful drafting, with the help of a patent attorney, will be necessary to navigate around clauses of relevant patents, but no major obstacles have been identified so far.

2.8 Allocation of resources

Project's outputs will be curated initially by their main creator, while on the second stage this responsibility will be of the leader of the Data Management task (T7.3), who will be responsible for the quality assurance and the proper storage of the outputs. Free storage facilities provided by partners, such as the servers provided by ATHENA RC to host the project's internal repository, will be used to the extent possible. In case costs incur, these will be handled by the project coordinator.



3 Ethics

The legal and ethical activities and issues have a transversal nature across the AI-DAPT workplan. In fact, the work and outcomes of T7.4 “Legal and Ethical Management Activities”, which intends to support the AI-DAPT partners in performing their research activities and experiments in an ethical way and in compliance with the current applicable regulatory systems, following responsible research ethics guidelines, are strictly interrelated with both the technical work packages and the demonstration and testing activities, as shown by the following figure.

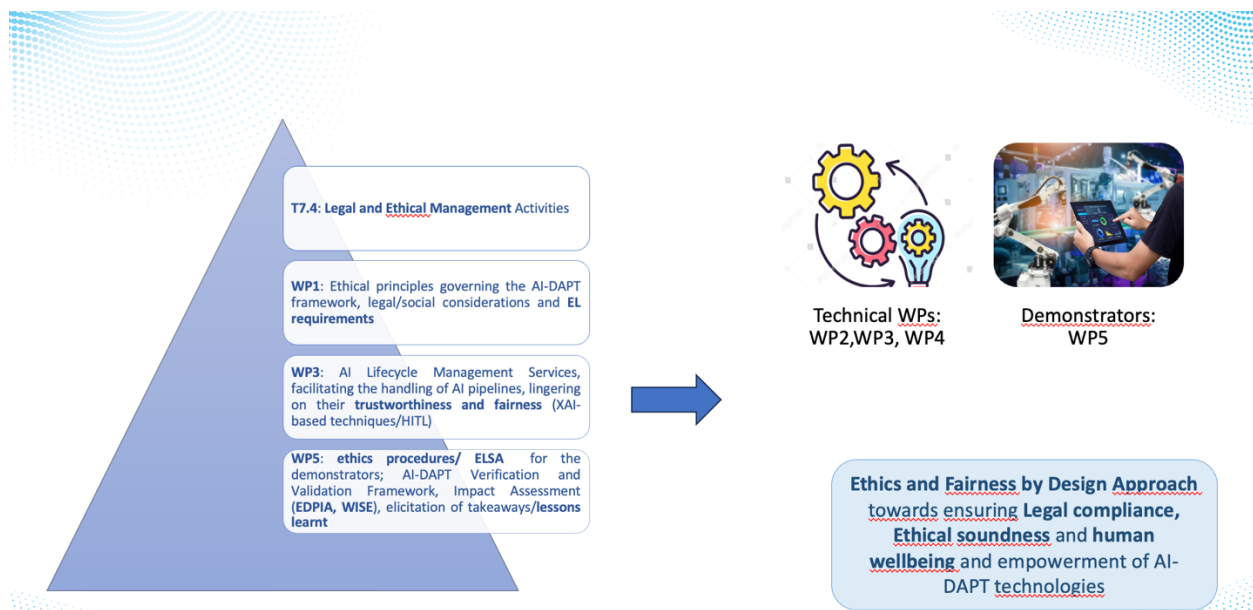


Figure 3. Transversal nature of the legal and ethical aspects in AI-DAPT

The **AI-DAPT Trustworthy Framework** will be elaborated through the legal and ethical review and requirement elicitation, following the Ethics-and-Fairness-by-design approach, towards ensuring legal compliance, ethical soundness and human well-being in AI-DAPT technological developments and validation.

In this direction, the initial set of European legal, regulatory and ethical sources relevant for AI-DAPT system and technological assets have been identified as part of WP6 activities and presented in D6.1. They have been classified in two main categories, respectively regarding Artificial Intelligence and Data. In particular, the initial list of legal and regulatory sources includes:

- Artificial Intelligence Domain:
 - AI Act (AIA) COM(2021) 206 final, Proposal for a Regulation of the European Parliament and of the Council laying down harmonized rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts;
 - AI Liability Directive (AILD) Proposal, COM (2022) 496 final “Proposal for a Directive of the European Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence;



- Revised Product Liability Directive (RPLD) Proposal, COM (2022) 495 final, “Proposal for a Directive of the European Parliament and of the Council on liability for defective product”;
- The AI innovation package to support Artificial Intelligence startups and SMEs⁹, adopted in January 2024;
- Ethics Guidelines for Trustworthy AI and ALTAI Assessment List;
- Data:
 - Data Governance Act (Regulation (EU) 2022/868);
 - Data Act (Regulation EU 2023/2854)
 - GDPR, Regulation (EU) 2016/679
 - Regulation on the free flow of non-personal data Regulation 2018/1807
 - e-Privacy Directive (Directive 2002/58/EC on privacy and electronic communications, replacing the Directive 97/66/EC and partially amended by Directive 2009/136/EC) and e-Privacy Regulation Proposal COM(2017) 10 final 2017/0003
 - Open Data Directive (including High Value Datasets). Directive (EU) 2019/1024
 - NIS 2 Directive EU 2022/2555)
 - Cybersecurity Act 2019/881/EU
 - Digital Services Act (DSA) (Regulation (EU) 2022/2065)
 - European Data Strategy (COM 2020 66 final
 - IDSA Rulebook 2023, concerning the IDS Data Sovereignty paradigm

In the next months this initial reference framework will be further enriched, monitored and deeply analysed in order to elicit legal and ethical requirements. This initial list of EU-wide regulatory sources might be enriched at a later stage, also taking into account the future developments under the AI, Data, and Robotics partnership. Furthermore, the Guidelines on the responsible use of generative AI in research¹⁰, recently published by the European Commission (20 March 2024) with the aim of supporting the European research community in the responsible use of generative artificial intelligence, will be deepened to assess their potential implications and impact on AI-DAPT technological developments. In addition to this, the AI-DAPT Regulatory Framework will be integrated with demonstrator-specific relevant regulatory and ethical sources, which will be considered for eliciting the legal and ethical requirements specific to them, if opportune.

Besides the legal survey, the context and initial analysis of user needs in the demonstrators and initial ethical considerations have been identified and analysed in the framework of WP1 “Automated AI Pipeline Lifecycle Management Framework”, in particular T1.1 “Automated AI Pipelines End-User Needs, Ethics Analysis, Constraints and Considerations” and reported in D1.1.

More specifically, a survey was performed on AI-DAPT demonstrators in Health, Robotics, Energy, and Manufacturing, directed to investigate the viewpoints, expectations and concerns regarding the ethical dimensions relevant for their pilot cases, their familiarity with ethical considerations, and how these considerations apply to their role as pilots: in other words, the consultation was functional to identify and address the key ethical constraints, as well as the pilots’ needs and desiderata, in the future technical development of the project, whilst revolutionizing AI with a data-centric approach seamlessly integrated into a model-centric, science-guided strategy across the AI-Ops lifecycle, enhancing generalizability, reliability, trustworthiness, and fairness in AI solutions.

As regards the ethics activities related to the demonstrators, another important step is the Ethics and Data Protection Impact Assessment Methodology within WP5, which will be elaborated in the next months within T5.1 “Verification and Validation Framework Definition and Baseline Impact

⁹ COM(2024) 28 final, Communication on boosting startups and innovation in trustworthy artificial intelligence

¹⁰ ERA Forum Stakeholders, “Living guidelines on the responsible use of generative AI in research – first version”, March 2024



Assessment” and described in D5.1 “Demonstrators Evaluation Framework and Use Case Plan”. It will be defined and implemented as part of the overall AI-DAPT Verification and Validation Framework.

Certain roles and responsibilities have been nominated for the implementation of the chosen ethical approach, and a preliminary analysis of ethical aspects has been performed, as explained in the following paragraphs.

3.1 Assigned roles and responsibilities

Ethics-related roles in AI-DAPT, as well as their responsibilities, are presented in Table 1. Roles include the Ethics Mentor (EM), Ethics Advisory Board (EAB) & Ethics Demonstrator Managers (EDM). The specific partners undertaking these roles are also presented in this table.

Table 1. Ethics-related roles and their responsibilities.

Role	Responsibilities	Name	Organization
Ethics Mentor	<ul style="list-style-type: none"> Internal organization and guidance on ethics management and assessment Interaction with the EAB and orchestration of its work Ensure that the WP leaders include ethics considerations in their workflows (ethics-by-design) 	Marina Cugurra	S&D
Ethics Advisory Board	<ul style="list-style-type: none"> Composition: intra-Consortium Experts Review and analysis of the AI-DAPT technology and use cases to identify all potential legal and socio-technical challenges/obstacles Oversight and Opinions on the AI-DAPT research activities, outputs and experiments to ensure legal compliance and ethical soundness 	Marina Cugurra	S&D
		Paulo Figueiras	UNINOVA
		Dimitris Christou	ZENITH
		Dimitris Panopoulos	SUITE5
Ethics Demonstrator Manager	<ul style="list-style-type: none"> Contribute to the fine-tuning of the ethics protocols and their implementation in the context of each experiment case Key role in timely identifying any ethics issues that might occur during experiments' operations 	Demo 1- Health Andreas F.H. Pfeiffer	CHARITE
		Demo 2- Robotics Francesco Dellino	MADE
		Demo 3- Energy Akis Gousios	ZENITH
		Demo 4 - Manufacturing Karl Hribernik	BIBA



The Kick-off-Meeting of the AI-DAPT Ethics Advisory Board & Ethics Demonstrator Managers was organized on 19 June 2024 in order to discuss the respective roles and responsibilities, to provide an overview of the main ethics issues and activities of AI-DAPT, as well as to delve into each of the demonstrators' planned operations to explore the legal and ethical dimensions at stake.

3.2 Ethical considerations in the AI-DAPT demonstrators

AI-DAPT will develop a Data and AI Risk Assessment framework, based on methodologies like the ALTAI-driven Human Rights Impact Assessment. This framework will: a) analyse the data needs and AI features of various components, b) identify potential risks and ethical issues according to current and proposed regulations, c) assess the severity and likelihood of these risks, and d) provide guidance for technical decisions to ensure compliance with AI-DAPT solutions.

In this context, a preliminary analysis of user needs and identification of ethical constraints and considerations in the AI-DAPT demonstrators has been conducted by S&D and UCY partners, and results are reported in detail in D1.1.

In summary, the specific needs of the demonstrators identified so far, with respect to data protection and ethical constraints, are listed below:

- Demonstrator 1 - Health: The data collection process will be done through a clinical trial study led by CHARITE, with the use of both bio-signals from wearable devices and blood test results. This involves both personal and sensitive data of human participants. An ethics approval has been submitted from CHARITE partners to the relevant authorities in their hospital to enable the initiation of the study and data collection. Informed consent must be acquired by all participants in the study.
- Demonstrator 2 - Robotics: Data for this demonstrator will also be collected from wearable devices, while the corresponding levels of stress will be annotated by participants themselves. This involves both personal and sensitive data of human participants. An ethics approval has not been submitted yet by either IMECH or MADE partners, as they have no relevant experience. They will be guided to do so with the assistance of the EM and demo 1 partners who are familiar with this process. Informed consent from operators that participate in the study should be explicitly stated.
- Demonstrator 3 - Energy: The available data include personal information of the company's clients, also confidential corporate data. Partners already anonymize their clients' personal data, however, advanced data protection and security mechanisms must be in place to handle this data.
- Demonstrator 4 - Manufacturing: Data sources in this case include confidential data of the company's clients, so that the data must be securely shared with AI-DAPT consortium for the needs of the demonstrator.

The activities towards the implementation of AI-DAPT Trustworthy Framework are therefore ongoing, with the mentioned Kick-off Meeting (KoM) of EAB and EDMs and the scheduling of dedicated sessions per demonstrator underway, as well as with the analysis and enrichment of the Legal, Regulatory and Ethical Framework and resulting Legal and Ethical Requirements identification, both at project-level and at demonstrator-level, as well as with the Ethics and Data Protection Impact Assessment Methodology to be adopted in WP5.



4 Data management in AI-DAPT demonstrators

The project's DMP has been created in Argos community tool and filled with the individual dataset descriptions of each demonstrator (Figure 4). Argos provides description templates to ensure all aspects related to privacy, security, ethics and FAIR data management are addressed for each asset.

The information documented therein is more complete for already existing datasets as in the case of Demonstrators 3 – Energy and 4- Manufacturing, as well as for the existing MCS glucose dataset in Demonstrator 1- Health. Some of the datasets have not been collected yet, so their descriptions will certainly be updated once the data is actually in place. More specific information on the exact mechanisms for data secure exchange and storage will be made available as soon as the corresponding design and development of the AI-DAPT platform has progressed.

A total of sixteen (16) datasets have been registered so far for the demonstrators, distributed as following:

- Demo 1 – Health: 3
- Demo 2 – Robotics: 2
- Demo 3 – Energy: 6
- Demo 4 – Manufacturing: 5

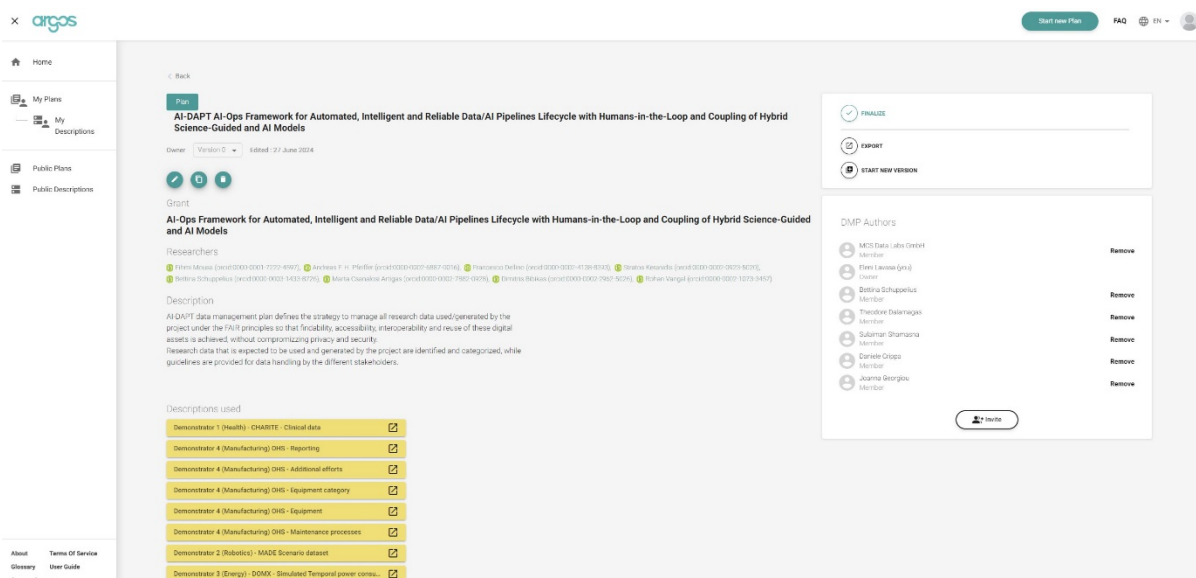


Figure 4. AI-DAPT DMP in Argos

The first version of the demonstrator-specific DMP, as exported by Argos, is provided in Annex I of this deliverable.



5 Conclusions and Next Steps

The AI-DAPT project underscores the crucial role of data in developing reliable and fair AI solutions. By creating an AIOps framework to support the entire AI pipeline, AI-DAPT aims to ensure that AI models are generalizable, trustworthy, and fair. The project highlights the importance of automating data management processes, incorporating explainable AI techniques, and combining data-driven AI models with science-based models to enhance the overall AI ecosystem. The project's Data Management Plan (DMP) is key to managing research data in alignment with FAIR principles, ensuring legal and ethical standards, as well as open access to research outputs without compromising privacy and sustainability. AI-DAPT's impact will be demonstrated through real-world applications in health, robotics, energy, and manufacturing, showcasing its practical benefits and integrating its solutions into existing AI tools.

The next steps involve continuously updating the DMP, implementing the AIOps framework to automate AI pipelines, and validating AI-DAPT solutions in real-world contexts. Ensuring ethical and legal compliance, fostering stakeholder engagement and collaboration, and promoting open science practices are essential. Efficient documentation and management of research data and outputs are necessary to support these initiatives, with all project participants adhering to the established guidelines. By following these steps, AI-DAPT aims to create a robust framework for managing AI development and deployment, driving innovation and collaboration across multiple industries while ensuring data-driven AI models are effective and ethically sound.



References

D1.1, 2024. *"AI-DAPT Automated AI Pipeline End User Needs and Scientific & Technology Radar"*, s.l.: s.n.

D6.1, 2024. *"Dissemination, Communication, Engagement & Innovation Plan"*, s.l.: s.n.

Wilkinson, M., Dumontier, M., Aalbersberg, I. & al., e., 2016. The FAIR Guiding Principles for scientific data management and stewardship.. *Sci Data*, Volume 3, p. 160018.



List of Acronyms/Abbreviations

Acronym/ Abbreviation	Description
AI	Artificial Intelligence
AIA	Artificial Intelligence Act
AILD	AI Liability Directive
DMP	Data Management Plan
DoA	Description of Action
DOI	Digital Object Identifier
EAB	Ethics Advisory Board
EC	European Commission
EDM	Ethics Demonstrator Manager
EM	Ethics Mentor
EU	European Union
FAIR	Findable, Accessible, Interoperable, Reusable
GA	Grant Agreement
GDPR	General Data Protection Regulation
H2020	Horizon 2020
IP(R)	Intellectual Property Rights
IRI	International Resource Identifier
KoM	Kick-off Meeting
UID	Unique Identifiers
URI	Uniform Resource Identifier
WP	Work Package



Annexes

Annex I

The output DMP from the Argos platform is presented here, including all demonstrator dataset descriptions. An indicative screenshot from editing a particular description in Argos is presented in Figure 5 below.

Figure 5. Dataset description in Argos

AI-DAPT AI-Ops Framework for Automated, Intelligent and Reliable Data/AI Pipelines Lifecycle with Humans-in-the-Loop and Coupling of Hybrid Science-Guided and AI Models

Version 0

Description

AI-DAPT data management plan defines the strategy to manage all research data used/generated by the project under the FAIR principles so that findability, accessibility, interoperability and reuse of these digital assets is achieved, without compromising privacy and security.

Research data that is expected to be used and generated by the project are identified and categorized, while guidelines are provided for data handling by the different stakeholders.

Funder

European Commission || EC

Grant

AI-Ops Framework for Automated, Intelligent and Reliable Data/AI Pipelines Lifecycle with Humans-in-the-Loop and Coupling of Hybrid Science-Guided and AI Models/ No 101135826

Researchers

Dimitris Bibikas (orcid:0000-0002-2962-5026), Rohan Vangal (orcid:0000-0002-1073-3457), Francesco Dellino (orcid:0000-0002-4138-8393), Bettina Schuppelius (orcid:0000-0003-1433-8726), Fihmi Mousa (orcid:0000-0001-7222-4597), Marta Csanalosi Artigas (orcid:0000-0002-7882-0928), Stratos Keranidis (orcid:0000-0002-0923-5020), Andreas F. H. Pfeiffer (orcid:0000-0002-6887-0016), Stefan Kabisch (orcid:0000-0003-1792-1757), Daniele Crippa, Sulaiman Shamasna, Carl Hans

Organizations

MADE SCARL, University of Cyprus, OHS ENGINEERING GMBH, Athena Research and Innovation Center in Information and Communication Technologies, DOMX IDIOTIKI KEFALAIOUCHIKI ETAIREIA, CHARITE - UNIVERSITAETSMEDIZIN BERLIN, CONSORZIO INTELLIMECH, MCS DATALABS, ETAIREIA PROMITHEIAS AERIOU THESSALONIKIS - THESSALIAS MONOPROSOPI ANONYMOS ETAIREIA

1. Main Info

Title of DMP: AI-DAPT AI-Ops Framework for Automated, Intelligent and Reliable Data/AI Pipelines Lifecycle with Humans-in-the-Loop and Coupling of Hybrid Science-Guided and AI Models

Description:

AI-DAPT data management plan defines the strategy to manage all research data used/generated by the project under the FAIR principles so that findability, accessibility, interoperability and reuse of these digital assets is achieved, without compromising privacy and security.

Research data that is expected to be used and generated by the project are identified and categorized, while guidelines are provided for data handling by the different stakeholders.

Researchers:

Dimitris Bibikas ([orcid:0000-0002-2962-5026](https://orcid.org/0000-0002-2962-5026))

Rohan Vangal ([orcid:0000-0002-1073-3457](https://orcid.org/0000-0002-1073-3457))

Francesco Dellino ([orcid:0000-0002-4138-8393](https://orcid.org/0000-0002-4138-8393))

Bettina Schuppelius ([orcid:0000-0003-1433-8726](https://orcid.org/0000-0003-1433-8726))

Fihmi Mousa ([orcid:0000-0001-7222-4597](https://orcid.org/0000-0001-7222-4597))

Marta Csanalosi Artigas ([orcid:0000-0002-7882-0928](https://orcid.org/0000-0002-7882-0928))

Stratos Keranidis ([orcid:0000-0002-0923-5020](https://orcid.org/0000-0002-0923-5020))

Andreas F. H. Pfeiffer ([orcid:0000-0002-6887-0016](https://orcid.org/0000-0002-6887-0016))

Stefan Kabisch ([orcid:0000-0003-1792-1757](https://orcid.org/0000-0003-1792-1757))

Daniele Crippa

Sulaiman Shamasna

Carl Hans

Organizations:

MADE SCARL

University of Cyprus

OHS ENGINEERING GMBH

Athena Research and Innovation Center in Information and Communication Technologies

DOMX IDIOTIKI KEFALAIOUCHIKI ETAIREIA

CHARITE - UNIVERSITAETSMEDIZIN BERLIN

CONSORZIO INTELLIMECH

MCS DATALABS

ETAIREIA PROMITHEIAS AERIOU THESSALONIKIS - THESSALIAS MONOPROSOPI ANONYMOS ETAIREIA

Contact: [Eleni Lavasa](#)

2. Funding

Funding organizations: [European Commission](#) | [EC](#)

Grants: [AI-Ops Framework for Automated, Intelligent and Reliable Data/AI Pipelines Lifecycle with Humans-in-the-Loop and Coupling of Hybrid Science-Guided and AI Models](#)

Project: [AI-DAPT](#)

3. License

License: [CC-BY-4.0](#)

Access Rights: [Restricted](#)

Part of

AI-DAPT AI-Ops Framework for Automated, Intelligent and Reliable Data/AI Pipelines Lifecycle with Humans-in-the-Loop and Coupling of Hybrid Science-Guided and AI Models

Description

Demonstrator 1 (Health) - MCS Glucose Dataset

Description

Timeseries dataset, with PPG metrics (photoplethysmogram; can be used to detect blood volume changes over time) of 10 subjects measured in 3 channels at 50 Hz sampling rate. Annotated with glucose levels and blood pressure every ten minutes.

A PPG channel refers to a specific set of sensors and components within a wearable multi-channel PPG system designed for photoplethysmography (PPG) measurements. Since PPG is a technique that uses light (green, red, and infrared) to monitor blood volume changes in tissues, each channel contributes unique insights into the interaction between light and tissues

Researchers

Fihmi Mousa (orcid:0000-0001-7222-4597), Sulaiman Shamasna

Description

1.1 Brief description of the described research output

1.1.1 What kind of research output are you describing?

Research Data

1.1.2 Is it physical or digital?

Digital

1.1.3 Are you generating or re-using it?

Re-used

This dataset has been used by MCS for Non-invasive glucose monitoring and will be re-used in the scope of the demonstrator.

1.1.4 What is the type of the described dataset?

Observational

A timeseries dataset with PPG metrics of 10 subjects measured in 3 channels at 50 Hz frequency. Annotated with glucose levels and blood pressure every ten minutes.

1.1.5 What is its format?

CSV format

1.1.6 What is its expected size?

unknown

1.1.7 Why are you collecting/generating or re-using it?

- To share information
- To improve a product
- To combine with other data
- Other

This information is shared by MCS with the AI-DAPT consortium, to be used for experimentation in the development and evaluation of the project's AI-Ops framework, also in the development & validation of specific solutions for the demonstrator. Since the data providers (MCS) already use this

information for predictive analytics, its re-use aims to the improvement of accuracy in the prediction of glucose levels.

1.1.8 What is its origin / provenance?

Data is collected and owned by MCS.

1.1.9 To whom might it be useful ('data utility')?

- Researchers
- Research communities
- Industry

This dataset may be useful to researchers and practitioners in the Health domain, for the non-invasive study/prediction of blood glucose levels in patients suffering from diabetes

2.1 Publications

2.1.1 Does the described output support any scientific publication?

No

2.1.2 Is there a data availability statement provided along with the publication?

No

2.3 Software

2.3.1 Does the described output use or support any software?

No

3.1.1 Making data findable, including provisions for metadata

3.1.1.1 What type(s) of persistent identifier(s) are used for the described dataset / output?

Data Identifiers

3.1.1.2 Will you provide metadata for the described dataset / output?

Yes

3.1.1.3 What type(s) of metadata?

Descriptive

3.1.1.4 Do the metadata use standardised vocabularies?

No

3.1.1.6 Are the metadata searchable?

No

3.1.1.8 Are keywords provided in the metadata?

No

3.1.1.9 Are metadata harvestable?

No

3.2.1 Repository

3.2.1.1 In which repository will the dataset / output be deposited?

AI-DAPT repository/database

3.2.1.2 Is the selected repository a trusted source?

Yes

- Has certification
- Supports authentication and authorization of users
- Has data security mechanisms in place

3.2.1.5 Does the repository(ies) assign datasets / outputs with persistent identifiers?

No

3.2.2 Data

3.2.2.1 What is the described dataset / output title?

MCS Glucose dataset

3.2.2.2 How is the dataset / output shared?

Shared

The dataset is expected to be shared under restricted access policies. Access will be granted to technical support partners of the demonstrator.

3.2.2.3 What is the reason of limiting access to the dataset / output?

Access to this dataset is limited due to its contents including personal and sensitive information of the people participating in the study.

3.2.2.5 Are there any methods or tools required to access the dataset / output?

No

3.2.2.8 Is the described dataset / output supported by a data access committee?

Yes

3.2.2.9 Please specify how the dataset / output will be accessed during and after the project ends

Restricted access to authorized users, sharing contracts

3.2.3 Metadata

3.2.3.1 Will you provide metadata even if the described dataset / output can not be openly shared?

Yes

3.2.3.3 Do metadata provide information about how to access the described dataset / output?

No

3.2.3.4 Will metadata remain available after the dataset / output is no longer available?

No

3.3 Making data and other outputs interoperable

3.3.1 Does your (meta)data use a controlled vocabulary?

No

3.3.3 Have you applied a standard schema for your (meta)data?

No

3.3.4 Will you provide a mapping to more commonly used ontologies?

Yes

TBD

3.3.5 What is the methodology followed?

TBD

3.3.6 What community-endorsed interoperability best practices are followed?

TBD

3.3.7 Does the described dataset / output provide qualified references with other outputs?

No

3.4 Increasing data and other outputs reuse

3.4.1 What internationally recognised licence will you use for your dataset / output?

TBD

3.4.2 What reusability and / or reproducibility methods are followed?

- Readme files
- Variable definitions
- Units of measurement

3.4.3 Will you provide the described dataset / output in the public domain?

No

3.4.4 Do you intend to ensure (re)use by third parties after your project finishes?

No

3.4.5 Is provenance well documented?

No

3.4.6 What documented procedures for quality assurance do you have in place?

Set up of scientific and technical committee

4.1 Allocation of resources

4.1.1 What will be the cost of making the described output FAIR?

The cost is covered by the EC funding of the project

4.1.2 How will this cost be covered?

Infrastructure Grant

The cost is covered by the EC funding of the project

4.1.3 Identify the people who will be responsible and their role(s) in the management of the described output

a. Fihmi Mousa (orcid:0000-0001-7222-4597)

Project Management

b. Sulaiman Shamasna

Data Scientist

5.1 Data Security

5.1.1 What security measures are followed?

Passwords

To be checked

5.1.2 What conditions do the security measures meet?

- Data access
- Data storage
- Data sharing

5.1.3 How will you preserve the described dataset / output in the long term?

Data will be stored and maintained in MCS data infrastructures

6.1 Ethical aspects

6.1.1 Are there any ethical or legal issues that can have an impact on sharing the described dataset / output?

yes

Personal & sensitive (medical) data. The dataset is proprietary and sharing contracts need to be in place for usage (within &) out of the scope of this project.

The clinical trial study protocol for Demonstrator #1 will be drafted and delivered to the relevant ethics acceptance committees (EU and CHARITE), at the beginning of T5.2 Demonstrators Use Cases Detailing, Coordination and Execution Planning (M7).

6.1.2 Does the described dataset / output contain sensitive information?

Yes

6.1.3 Does the described dataset / output contain personal data?

Yes

6.1.4 What are the methods used for processing and accessing sensitive/personal information?

- Anonymising data where necessary

- Privacy constraints and applicable ethical norms
- Data accompanied by informed consent statements

7.1 Other

7.1.1 Do you make use of other procedures for data management?

No

Powered by



Part of

AI-DAPT AI-Ops Framework for Automated, Intelligent and Reliable Data/AI Pipelines Lifecycle with Humans-in-the-Loop and Coupling of Hybrid Science-Guided and AI Models

Description

Demonstrator 1 (Health) - CHARITE & MCS GlucoNIV

Description

Timeseries dataset with primary signal being PPG. Secondary data such as motion data could be used to clean the data and detect artifacts. The dataset is planned to include at least 34 patients at risk of developing diabetes or with manifest diabetes.

PPG: 3-6 channels at 100-200 Hz, Accelerometer: 3 channels: 25-200 Hz, Gyroscope: 25-200 Hz

A PPG channel refers to a specific set of sensors and components within a wearable multi-channel PPG system designed for photoplethysmography (PPG) measurements. Since PPG is a non-invasive technique that uses light (green, red, and infrared) to monitor blood volume changes in tissues, each channel contributes unique insights into the interaction between light and tissues.

PPG signals will be collected with two devices each time, one wearable at the wrist and one finger clip.

Researchers

Bettina Schuppelius (orcid:0000-0003-1433-8726), Andreas F. H. Pfeiffer (orcid:0000-0002-6887-0016), Marta Csanalosi Artigas (orcid:0000-0002-7882-0928), Stefan Kabisch (orcid:0000-0003-1792-1757), Fihmi Mousa (orcid:0000-0001-7222-4597), Sulaiman Shamasna

Description

1.1 Brief description of the described research output

1.1.1 What kind of research output are you describing?

Research Data

1.1.2 Is it physical or digital?

Digital

1.1.3 Are you generating or re-using it?

New

Data to be collected by CHARITE and MCS in the context of AI-DAPT Demonstrator 1 - Health

1.1.4 What is the type of the described dataset?

Sample or specimen data

1.1.5 What is its format?

Probably CSV

1.1.6 What is its expected size?

unknown (approx. 600 observations)

1.1.7 Why are you collecting/generating or re-using it?

- To obtain information
- To share information

- To develop a product
- To improve a product
- To combine with other data
- Other

1.1.8 What is its origin / provenance?

Data will be owned by CHARITE and shared with MCS upon the consent of subjects participating in the study.

1.1.9 To whom might it be useful ('data utility')?

- Researchers
- Research communities
- Industry

This dataset may be useful to researchers and practitioners in the Health domain, for the development of a non-invasive method to study and/or predict blood glucose levels in healthy patients as a screening tool, in patients at risk of developing diabetes and patients with overt diabetes.

2.1 Publications

2.1.1 Does the described output support any scientific publication?

No

2.1.2 Is there a data availability statement provided along with the publication?

No

2.3 Software

2.3.1 Does the described output use or support any software?

No

3.1.1 Making data findable, including provisions for metadata

3.1.1.1 What type(s) of persistent identifier(s) are used for the described dataset / output?

- Projects identifiers
- Other

Other

NCT Number (ID for clinical studies registered on ClinicalTrials.gov)

3.1.1.2 Will you provide metadata for the described dataset / output?

Yes

3.1.1.3 What type(s) of metadata?

Descriptive

3.1.1.4 Do the metadata use standardised vocabularies?

No

3.1.1.6 Are the metadata searchable?

No

3.1.1.8 Are keywords provided in the metadata?

No

3.1.1.9 Are metadata harvestable?

No

To be determined

3.2.1 Repository

3.2.1.1 In which repository will the dataset / output be deposited?

AI-DAPT repository/database

3.2.1.2 Is the selected repository a trusted source?

Yes

- Has certification
- Supports authentication and authorization of users
- Has data security mechanisms in place

3.2.1.5 Does the repository(ies) assign datasets / outputs with persistent identifiers?

No

3.2.2 Data

3.2.2.1 What is the described dataset / output title?

CHARITE & MCS GlucoNIV

3.2.2.2 How is the dataset / output shared?

Shared

The dataset is expected to be shared under restricted access policies. Access will be granted to technical support partners of the demonstrator.

3.2.2.3 What is the reason of limiting access to the dataset / output?

Access to this dataset is limited due to its contents including personal and sensitive information of the people participating in the study.

3.2.2.5 Are there any methods or tools required to access the dataset / output?

No

3.2.2.8 Is the described dataset / output supported by a data access committee?

Yes

Through the Charité Clinical Trial Office, Team Data protection

3.2.2.9 Please specify how the dataset / output will be accessed during and after the project ends

Restricted access to authorized users, sharing contracts

3.2.3 Metadata

3.2.3.1 Will you provide metadata even if the described dataset / output can not be openly shared?

No

To be decided

3.2.3.3 Do metadata provide information about how to access the described dataset / output?

No

3.2.3.4 Will metadata remain available after the dataset / output is no longer available?

No

3.3 Making data and other outputs interoperable

3.3.1 Does your (meta)data use a controlled vocabulary?

No

To be decided

3.3.2 If you created the vocabulary, where can it be found?

To be decided

3.3.3 Have you applied a standard schema for your (meta)data?

No

3.3.4 Will you provide a mapping to more commonly used ontologies?

No

To be decided

3.3.5 What is the methodology followed?

To be determined

3.3.6 What community-endorsed interoperability best practices are followed?

To be determined

3.3.7 Does the described dataset / output provide qualified references with other outputs?

Yes

This dataset contains the actual PPG metrics for all subjects as time-series data. This is linked to a dataset, describing the characteristics of the population participating in the clinical trial (e.g. age, sex, blood pressure etc.) and their blood glucose with the clinical standard methods.

3.4 Increasing data and other outputs reuse

3.4.1 What internationally recognised licence will you use for your dataset / output?

To be determined

3.4.2 What reusability and / or reproducibility methods are followed?

- Readme files
- Variable definitions
- Units of measurement

3.4.3 Will you provide the described dataset / output in the public domain?

No

3.4.4 Do you intend to ensure (re)use by third parties after your project finishes?

No

To be decided

3.4.5 Is provenance well documented?

Yes

3.4.6 What documented procedures for quality assurance do you have in place?

Set up of scientific and technical committee

4.1 Allocation of resources

4.1.1 What will be the cost of making the described output FAIR?

The cost is covered by the EC funding of the project

4.1.2 How will this cost be covered?

Infrastructure Grant

The cost is covered by the EC funding of the project

4.1.3 Identify the people who will be responsible and their role(s) in the management of the described output

a. Marta Csanalosi Artigas (orcid:0000-0002-7882-0928)

Research assistant

b. Bettina Schuppelius (orcid:0000-0003-1433-8726)

Research assistant

c. Andreas F. H. Pfeiffer (orcid:0000-0002-6887-0016)

Senior Professor

d. Stefan Kabisch (orcid:0000-0003-1792-1757)

Study physician, clinical researcher

e. Fihmi Mousa (orcid:0000-0001-7222-4597)

Project Management

f. Sulaiman Shamasna

Data Scientist

5.1 Data Security

5.1.1 What security measures are followed?

- Passwords
- Physical access control

5.1.2 What conditions do the security measures meet?

- Data access
- Data storage
- Data sharing

5.1.3 How will you preserve the described dataset / output in the long term?

Data will be stored and maintained in CHARITE data infrastructures for a minimum of 10 years after study completion.

6.1 Ethical aspects

6.1.1 Are there any ethical or legal issues that can have an impact on sharing the described dataset / output?

yes

Personal & sensitive (medical) data. The dataset is proprietary and sharing contracts need to be in place for usage (within &) out of the scope of this project.

The clinical trial study protocol for Demonstrator #1 will be drafted and delivered to the relevant ethics acceptance committees (EU and CHARITE), at the beginning of T5.2 Demonstrators Use Cases Detailing, Coordination and Execution Planning (M7).

6.1.2 Does the described dataset / output contain sensitive information?

Yes

6.1.3 Does the described dataset / output contain personal data?

Yes

6.1.4 What are the methods used for processing and accessing sensitive/personal information?

- Anonymising data where necessary

- Privacy constraints and applicable ethical norms
- Data accompanied by informed consent statements
- National laws

7.1 Other

7.1.1 Do you make use of other procedures for data management?

No

Powered by



Part of

AI-DAPT AI-Ops Framework for Automated, Intelligent and Reliable Data/AI Pipelines Lifecycle with Humans-in-the-Loop and Coupling of Hybrid Science-Guided and AI Models

Description

Demonstrator 1 (Health) - CHARITE - Clinical data

Description

Clinical data on the patients participating in the clinical trial for the CHARITE & MCS GlucoNIV dataset, e.g. age, sex, blood pressure etc. and their blood glucose determined at different times by the clinical standard methods.

Researchers

Andreas F. H. Pfeiffer (orcid:0000-0002-6887-0016), Marta Csanalosi Artigas (orcid:0000-0002-7882-0928), Bettina Schuppelius (orcid:0000-0003-1433-8726), Stefan Kabisch (orcid:0000-0003-1792-1757)

Description

1.1 Brief description of the described research output

1.1.1 What kind of research output are you describing?

Research Data

1.1.2 Is it physical or digital?

Digital

1.1.3 Are you generating or re-using it?

New

Data to be collected by CHARITE in the context of AI-DAPT Demonstrator 1 - Health

1.1.4 What is the type of the described dataset?

Sample or specimen data

1.1.5 What is its format?

Probably CSV

1.1.6 What is its expected size?

unknown (approx. 600 observations)

1.1.7 Why are you collecting/generating or re-using it?

- To obtain information
- To improve a product
- To combine with other data

1.1.8 What is its origin / provenance?

Data will be owned by Charité and shared with MCS datalabs upon the consent of subjects participating in the study.

1.1.9 To whom might it be useful ('data utility')?

- Researchers

- Research communities
- Industry

This dataset may be useful to researchers and practitioners in the Health domain, for the development of a non-invasive method to study and/or predict blood glucose levels in healthy patients as a screening tool, in patients at risk of developing diabetes and patients with overt diabetes.

2.1 Publications

2.1.1 Does the described output support any scientific publication?

No

2.1.2 Is there a data availability statement provided along with the publication?

No

2.3 Software

2.3.1 Does the described output use or support any software?

No

3.1.1 Making data findable, including provisions for metadata

3.1.1.1 What type(s) of persistent identifier(s) are used for the described dataset / output?

- Data identifiers
- Other

DOI

In addition the trial will be registered at ClinicalTrials.gov and receive a NCT number (ID for clinical studies registered on ClinicalTrials.gov).

3.1.1.2 Will you provide metadata for the described dataset / output?

No

TBD

3.2.1 Repository

3.2.1.1 In which repository will the dataset / output be deposited?

AI-Dapt repository/database, Zenodo

<https://zenodo.org/>

ZENODO builds and operates a simple and innovative service that enables researchers, scientists, EU projects and institutions to share and showcase multidisciplinary research results (data and publications) that are not part of the existing institutional or subject-based repositories of the research communities. ZENODO enables researchers, scientists, EU projects and institutions to: easily share the long tail of small research results in a wide variety of formats including text, spreadsheets, audio, video, and images across all fields of science. display their research results and get credited by making the research results citable and integrate them into existing reporting lines to funding agencies like the European Commission. easily access and reuse shared research results.

3.2.1.2 Is the selected repository a trusted source?

Yes

- Has certification
- Supports authentication and authorization of users
- Has data security mechanisms in place

3.2.1.5 Does the repository(ies) assign datasets / outputs with persistent identifiers?

Yes

3.2.1.7 Does the repository support versioning?

Unknown

3.2.2 Data

3.2.2.1 What is the described dataset / output title?

CHARITE Clinical data

3.2.2.2 How is the dataset / output shared?

Shared

The dataset is expected to be shared under restricted access policies. Access will be granted to technical support partners of the demonstrator.

3.2.2.3 What is the reason of limiting access to the dataset / output?

Access to this dataset is limited due to its contents including personal and sensitive information of the people participating in the study.

3.2.2.5 Are there any methods or tools required to access the dataset / output?

No

3.2.2.8 Is the described dataset / output supported by a data access committee?

Yes

Through the Charité Clinical Trial Office, Team Data protection

3.2.2.9 Please specify how the dataset / output will be accessed during and after the project ends

Restricted access to authorized users, sharing contracts

3.2.2.10 Please specify how long after the project has ended the dataset / output will be made accessible for

TBD

3.2.3 Metadata

3.2.3.1 Will you provide metadata even if the described dataset / output can not be openly shared?

No

TBD

3.2.3.3 Do metadata provide information about how to access the described dataset / output?

No

3.2.3.4 Will metadata remain available after the dataset / output is no longer available?

No

3.3 Making data and other outputs interoperable

3.3.1 Does your (meta)data use a controlled vocabulary?

Yes

3.3.3 Have you applied a standard schema for your (meta)data?

No

3.3.4 Will you provide a mapping to more commonly used ontologies?

No

3.3.5 What is the methodology followed?

TBD

3.3.6 What community-endorsed interoperability best practices are followed?

TBD

3.3.7 Does the described dataset / output provide qualified references with other outputs?

Yes

This dataset contains the characteristics of the population participating in the clinical trial (e.g. age, sex, BMI, blood pressure etc.) and their blood glucose measurements with clinical standard methods. This data will be incorporated with the CHARITE & MCS GlucoNIV dataset, which contains the PPG metrics for all subjects as time-series data.

3.4 Increasing data and other outputs reuse

3.4.1 What internationally recognised licence will you use for your dataset / output?

Creative Commons Attribution 4.0

3.4.2 What reusability and / or reproducibility methods are followed?

- Variable definitions
- Units of measurement

3.4.3 Will you provide the described dataset / output in the public domain?

No

3.4.4 Do you intend to ensure (re)use by third parties after your project finishes?

Yes

TBD

3.4.5 Is provenance well documented?

No

3.4.6 What documented procedures for quality assurance do you have in place?

Set up of scientific and technical committee

4.1 Allocation of resources

4.1.1 What will be the cost of making the described output FAIR?

The cost is covered by the EC funding of the project

4.1.2 How will this cost be covered?

Infrastructure Grant

The cost is covered by the EC funding of the project

4.1.3 Identify the people who will be responsible and their role(s) in the management of the described output

a. Marta Csanalosi Artigas (orcid:0000-0002-7882-0928)

Research assistant

b. Bettina Schuppelius (orcid:0000-0003-1433-8726)

Research assistant

c. Andreas F. H. Pfeiffer (orcid:0000-0002-6887-0016)

Senior Professor

d. Stefan Kabisch (orcid:0000-0003-1792-1757)

Study physician, clinical researcher

5.1 Data Security

5.1.1 What security measures are followed?

- Firewall
- Passwords
- Physical access control

5.1.2 What conditions do the security measures meet?

- Data access
- Data storage
- Data sharing

5.1.3 How will you preserve the described dataset / output in the long term?

Data will be stored and maintained in CHARITE data infrastructures for a minimum of 10 years after study completion.

6.1 Ethical aspects

6.1.1 Are there any ethical or legal issues that can have an impact on sharing the described dataset / output?

yes

Personal & sensitive (medical) data. The dataset is proprietary and sharing contracts need to be in place for usage (within &) out of the scope of this project.

The clinical trial study protocol for Demonstrator #1 will be drafted and delivered to the relevant ethics acceptance committees (EU and CHARITE), at the beginning of T5.2 Demonstrators Use Cases Detailing, Coordination and Execution Planning (M7).

6.1.2 Does the described dataset / output contain sensitive information?

Yes

6.1.3 Does the described dataset / output contain personal data?

Yes

6.1.4 What are the methods used for processing and accessing sensitive/personal information?

- Anonymising data where necessary
- Privacy constraints and applicable ethical norms
- Data accompanied by informed consent statements
- National laws

7.1 Other

7.1.1 Do you make use of other procedures for data management?

No

Powered by



Part of

AI-DAPT AI-Ops Framework for Automated, Intelligent and Reliable Data/AI Pipelines Lifecycle with Humans-in-the-Loop and Coupling of Hybrid Science-Guided and AI Models

Description

Demonstrator 2 (Robotics) - IMECH biosignals dataset

Description

Dataset from partner companies, which will consist of sensor (timeseries) data: physiological real-time signals from wearable devices. The data will be manually annotated based on operator feedback (directly or through interviews).

Data will probably be collected in 1 hour batches with a sampling frequency around 100Hz.

Researchers

Rohan Vangal (orcid:0000-0002-1073-3457), Daniele Crippa

Description

1.1 Brief description of the described research output

1.1.1 What kind of research output are you describing?

Research Data

1.1.2 Is it physical or digital?

Digital

1.1.3 Are you generating or re-using it?

New

1.1.4 What is the type of the described dataset?

Observational

Data will be collected through wearable devices and integrated with labels that indicates the task and the target feature level (ex. stress). Labels could be defined through a final survey for the test subject or with input of the operator by means of the device itself.

1.1.5 What is its format?

TBD - most likely CSV format, but could also be parquet format, to increase efficiency in storage and transfer, to the slight detriment of integration

1.1.6 What is its expected size?

Approximately 50 MB

1.1.7 Why are you collecting/generating or re-using it?

- To obtain information
- To share information
- To develop a product

1.1.8 What is its origin / provenance?

Data will be owned by IMECH's partner companies and provided to IMECH upon their consent.

1.1.9 To whom might it be useful ('data utility')?

- Researchers

- Industry

2.1 Publications

2.1.1 Does the described output support any scientific publication?

No

2.1.2 Is there a data availability statement provided along with the publication?

No

2.3 Software

2.3.1 Does the described output use or support any software?

No

3.1.1 Making data findable, including provisions for metadata

3.1.1.1 What type(s) of persistent identifier(s) are used for the described dataset / output?

Data identifiers

TBD

3.1.1.2 Will you provide metadata for the described dataset / output?

Yes

TBD

3.1.1.3 What type(s) of metadata?

Descriptive

TBD - Probably ID that includes info on test subject and test number for the specific subject and possibly static parameters of the subjects (ex. age), appropriately anonymized

3.1.1.4 Do the metadata use standardised vocabularies?

No

3.1.1.6 Are the metadata searchable?

No

3.1.1.8 Are keywords provided in the metadata?

No

3.1.1.9 Are metadata harvestable?

No

3.2.1 Repository

3.2.1.1 In which repository will the dataset / output be deposited?

AI-DAPT repository/database

3.2.1.2 Is the selected repository a trusted source?

Yes

- Has certification
- Supports authentication and authorization of users
- Has data security mechanisms in place

3.2.1.5 Does the repository(ies) assign datasets / outputs with persistent identifiers?

No

3.2.2 Data

3.2.2.1 What is the described dataset / output title?

IMECH - Physiological signals dataset

3.2.2.2 How is the dataset / output shared?

Shared

Confidentiality on specific parameters will need to be explored, depending on the exact method for the dataset collection.

3.2.2.3 What is the reason of limiting access to the dataset / output?

To be decided whether this dataset can be publicly available

3.2.2.5 Are there any methods or tools required to access the dataset / output?

No

3.2.2.8 Is the described dataset / output supported by a data access committee?

Yes

3.2.3 Metadata

3.2.3.1 Will you provide metadata even if the described dataset / output can not be openly shared?

Yes

3.2.3.2 Under which license will metadata be provided?

Creative Commons Zero (CC0)

3.2.3.3 Do metadata provide information about how to access the described dataset / output?

No

3.2.3.4 Will metadata remain available after the dataset / output is no longer available?

No

3.3 Making data and other outputs interoperable

3.3.1 Does your (meta)data use a controlled vocabulary?

No

3.3.3 Have you applied a standard schema for your (meta)data?

No

3.3.4 Will you provide a mapping to more commonly used ontologies?

Yes

TBD

3.3.5 What is the methodology followed?

TBD

3.3.6 What community-endorsed interoperability best practices are followed?

TBD

3.3.7 Does the described dataset / output provide qualified references with other outputs?

No

TBD whether static data characterizing the operators participating in the study can be included as metadata to the IMECH biosignals dataset.

3.4 Increasing data and other outputs reuse

3.4.1 What internationally recognised licence will you use for your dataset / output?

TBD

3.4.2 What reusability and / or reproducibility methods are followed?

- Readme files
- Variable definitions
- Units of measurement

3.4.3 Will you provide the described dataset / output in the public domain?

No

3.4.4 Do you intend to ensure (re)use by third parties after your project finishes?

No

3.4.5 Is provenance well documented?

No

3.4.6 What documented procedures for quality assurance do you have in place?

Set up of scientific and technical committee

4.1 Allocation of resources

4.1.1 What will be the cost of making the described output FAIR?

The cost is covered by the EC funding of the project

4.1.2 How will this cost be covered?

Infrastructure Grant

The cost is covered by the EC funding of the project

4.1.3 Identify the people who will be responsible and their role(s) in the management of the described output

Daniele Crippa

Researcher

5.1 Data Security

5.1.1 What security measures are followed?

- Firewall
- Passwords

TBD

5.1.2 What conditions do the security measures meet?

- Data access
- Data storage
- Data sharing

5.1.3 How will you preserve the described dataset / output in the long term?

IMECH data infrastructures

6.1 Ethical aspects

6.1.1 Are there any ethical or legal issues that can have an impact on sharing the described dataset / output?

yes

Possible challenges regarding privacy to be resolved.

6.1.2 Does the described dataset / output contain sensitive information?

Yes

6.1.3 Does the described dataset / output contain personal data?

Yes

6.1.4 What are the methods used for processing and accessing sensitive/personal information?

- Anonymising data where necessary
- Privacy constraints and applicable ethical norms
- Data accompanied by informed consent statements
- Privacy policies
- National laws

7.1 Other

7.1.1 Do you make use of other procedures for data management?

No

Powered by



Part of

AI-DAPT AI-Ops Framework for Automated, Intelligent and Reliable Data/AI Pipelines Lifecycle with Humans-in-the-Loop and Coupling of Hybrid Science-Guided and AI Models

Description

Demonstrator 2 (Robotics) - MADE Scenario dataset

Description

Dataset from partner companies, which will consist of sensor (timeseries) data: physiological real-time signals from wearable devices. The data will be manually annotated based on operator feedback (directly or through interviews).

Data will probably be collected in 30 – 40 minutes batches with a sampling frequency around 1 out of 30 seconds.

Researchers

Francesco Dellino (orcid:0000-0002-4138-8393)

Description

1.1 Brief description of the described research output

1.1.1 What kind of research output are you describing?

Research Data

1.1.2 Is it physical or digital?

Digital

1.1.3 Are you generating or re-using it?

New

1.1.4 What is the type of the described dataset?

Observational

1.1.5 What is its format?

NoSQL

1.1.6 What is its expected size?

Approximately 600 kB for a single batch

1.1.7 Why are you collecting/generating or re-using it?

- To obtain information
- To share information
- To develop a product

1.1.8 What is its origin / provenance?

Data will be owned by MADE's partner companies and provided to MADE upon their consent.

Data will be collected through experimentation in MADE's facilities.

1.1.9 To whom might it be useful ('data utility')?

- Researchers
- Industry

2.1 Publications

2.1.1 Does the described output support any scientific publication?

No

2.1.2 Is there a data availability statement provided along with the publication?

No

2.3 Software

2.3.1 Does the described output use or support any software?

Yes

GARMIN connect Developer and API WAMP

<https://developer.garmin.com/gc-developer-program/overview/>

<https://wamp-proto.org/index.html>

3.1.1 Making data findable, including provisions for metadata

3.1.1.1 What type(s) of persistent identifier(s) are used for the described dataset / output?

Data identifiers

TBD

3.1.1.2 Will you provide metadata for the described dataset / output?

Yes

3.1.1.3 What type(s) of metadata?

Descriptive

3.1.1.4 Do the metadata use standardised vocabularies?

No

3.1.1.6 Are the metadata searchable?

Yes

3.1.1.7 How are searchable metadata provided?

Other

3.1.1.8 Are keywords provided in the metadata?

Yes

3.1.1.9 Are metadata harvestable?

Yes

3.2.1 Repository

3.2.1.1 In which repository will the dataset / output be deposited?

No SQL database, MongoDB in MADE local server

3.2.1.2 Is the selected repository a trusted source?

Yes

- Has certification
- Supports authentication and authorization of users
- Has data security mechanisms in place

3.2.1.5 Does the repository(ies) assign datasets / outputs with persistent identifiers?

Yes

3.2.2 Data

3.2.2.1 What is the described dataset / output title?

MADE - AI-DAPT dataset

3.2.2.2 How is the dataset / output shared?

Shared

Data will be shared in JSON format.

Confidentiality on specific parameters will need to be explored, depending on the exact method for the dataset collection.

3.2.2.3 What is the reason of limiting access to the dataset / output?

To be decided whether this dataset can be publicly available

3.2.2.5 Are there any methods or tools required to access the dataset / output?

No

3.2.2.8 Is the described dataset / output supported by a data access committee?

Yes

3.2.3 Metadata

3.2.3.1 Will you provide metadata even if the described dataset / output can not be openly shared?

Yes

3.2.3.2 Under which license will metadata be provided?

Creative Commons Zero (CC0)

3.2.3.3 Do metadata provide information about how to access the described dataset / output?

No

3.2.3.4 Will metadata remain available after the dataset / output is no longer available?

No

3.3 Making data and other outputs interoperable

3.3.1 Does your (meta)data use a controlled vocabulary?

No

To be determined

3.3.3 Have you applied a standard schema for your (meta)data?

No

3.3.4 Will you provide a mapping to more commonly used ontologies?

Yes

TBD

3.3.5 What is the methodology followed?

TBD

3.3.6 What community-endorsed interoperability best practices are followed?

TBD

3.3.7 Does the described dataset / output provide qualified references with other outputs?

No

3.4 Increasing data and other outputs reuse

3.4.1 What internationally recognised licence will you use for your dataset / output?

TBD

3.4.2 What reusability and / or reproducibility methods are followed?

- Readme files
- Variable definitions
- Units of measurement
- Other

3.4.3 Will you provide the described dataset / output in the public domain?

No

3.4.4 Do you intend to ensure (re)use by third parties after your project finishes?

No

TBD

3.4.5 Is provenance well documented?

Yes

3.4.6 What documented procedures for quality assurance do you have in place?

Set up of scientific and technical committee

4.1 Allocation of resources

4.1.1 What will be the cost of making the described output FAIR?

The cost is covered by the EC funding of the project

4.1.2 How will this cost be covered?

Infrastructure Grant

The cost is covered by the EC funding of the project

4.1.3 Identify the people who will be responsible and their role(s) in the management of the described output

Francesco Dellino (orcid:0000-0002-4138-8393)

Project Manager

5.1 Data Security

5.1.1 What security measures are followed?

- Firewall
- Passwords

5.1.2 What conditions do the security measures meet?

- Data access
- Data storage
- Data sharing

5.1.3 How will you preserve the described dataset / output in the long term?

MADE data infrastructures

6.1 Ethical aspects

6.1.1 Are there any ethical or legal issues that can have an impact on sharing the described dataset / output?

yes

Possible challenges regarding personal data to be resolved.

6.1.2 Does the described dataset / output contain sensitive information?

Yes

6.1.3 Does the described dataset / output contain personal data?

Yes

6.1.4 What are the methods used for processing and accessing sensitive/personal information?

- Anonymising data where necessary
- Privacy constraints and applicable ethical norms
- Data accompanied by informed consent statements

7.1 Other

7.1.1 Do you make use of other procedures for data management?

No

Powered by



Part of

AI-DAPT AI-Ops Framework for Automated, Intelligent and Reliable Data/AI Pipelines Lifecycle with Humans-in-the-Loop and Coupling of Hybrid Science-Guided and AI Models

Description

Demonstrator 3 (Energy) - DOMX - Temporal power consumption dataset

Description

Time-series dataset collected for 2 years from 100 residential and commercial buildings.

Attributes include:

Heating: Indoor/outdoor temperature, climate comfort, heating/hot water usage

User: Comfort limits, user schedules and preferences, app interactions

Energy: instant power, energy per usage scenario

All parameters are collected at maximum rate per minute.

Researchers

Stratos Keranidis (orcid:0000-0002-0923-5020)

Description

1.1 Brief description of the described research output

1.1.1 What kind of research output are you describing?

Research Data

1.1.2 Is it physical or digital?

Digital

1.1.3 Are you generating or re-using it?

Re-used

This data has been used so far by DOMX for predictive analysis of energy loads, and will be re-used and extended within the scope of the project.

1.1.4 What is the type of the described dataset?

Observational

Time-series dataset collected for 2 years from 100 residential and commercial buildings, including heating, user and energy information.

1.1.5 What is its format?

CSV format

1.1.6 What is its expected size?

approximately 5-10GB

1.1.7 Why are you collecting/generating or re-using it?

- To share information
- To make informed decisions
- To improve a product
- To combine with other data
- Other

This information is shared by DOMX with the AI-DAPT consortium, to be used for experimentation in the development and evaluation of the project's AI-Ops framework, also in the development &

validation of specific solutions for the demonstrator. Since the data providers (DOMX) already use this information for predictive analytics, its re-use aims to the improvement of accuracy in the prediction of energy loads so that the company is assisted to make more informed decisions.

1.1.8 What is its origin / provenance?

Data collected by DOMX and owned by DOMX clients, provided upon their consent for the research purposes of the project.

1.1.9 To whom might it be useful ('data utility')?

- Researchers
- Economy
- Industry

Researchers in academia or the energy industry may use this data for the statistical/correlation/predictive analysis of energy consumption patterns and influencing factors.

2.1 Publications

2.1.1 Does the described output support any scientific publication?

a. Yes

A novel system for providing explicit demand response from domestic natural gas boilers

Applied Energy

b. Yes

Minimization of natural gas consumption of domestic boilers with convolutional long-short term memory neural networks and genetic algorithm

Applied Energy

2.1.2 Is there a data availability statement provided along with the publication?

No

2.3 Software

2.3.1 Does the described output use or support any software?

Yes

<https://domx.io>

3.1.1 Making data findable, including provisions for metadata

3.1.1.1 What type(s) of persistent identifier(s) are used for the described dataset / output?

Data identifiers

3.1.1.2 Will you provide metadata for the described dataset / output?

Yes

Optionally

3.1.1.3 What type(s) of metadata?

Descriptive

3.1.1.4 Do the metadata use standardised vocabularies?

Yes

SAREF4ENER is used, the extension of Smart Applications REFerence (SAREF) suite of ontologies for the Energy domain.

3.1.1.5 Please provide URL/Description of used vocabularies

<https://saref.etsi.org/index.html>

3.1.1.6 Are the metadata searchable?

No

3.1.1.8 Are keywords provided in the metadata?

No

3.1.1.9 Are metadata harvestable?

No

3.2.1 Repository

3.2.1.1 In which repository will the dataset / output be deposited?

Own infrastructure

3.2.1.2 Is the selected repository a trusted source?

Yes

- Has certification

- Supports authentication and authorization of users
- Has data security mechanisms in place

3.2.1.5 Does the repository(ies) assign datasets / outputs with persistent identifiers?

No

3.2.2 Data

3.2.2.1 What is the described dataset / output title?

Time-series dataset on energy consumption - DOMX

3.2.2.2 How is the dataset / output shared?

Shared

The dataset is expected to be shared under restricted access policies. Access will be granted to technical support partners of the energy demonstrator.

3.2.2.3 What is the reason of limiting access to the dataset / output?

Access to this dataset is limited due to its contents including personal information of the company's customers.

3.2.2.5 Are there any methods or tools required to access the dataset / output?

No

3.2.2.8 Is the described dataset / output supported by a data access committee?

Yes

3.2.2.9 Please specify how the dataset / output will be accessed during and after the project ends

Restricted access to authorized users, sharing contracts

3.2.3 Metadata

3.2.3.1 Will you provide metadata even if the described dataset / output can not be openly shared?

Yes

3.2.3.3 Do metadata provide information about how to access the described dataset / output?

No

3.2.3.4 Will metadata remain available after the dataset / output is no longer available?

No

3.3 Making data and other outputs interoperable

3.3.1 Does your (meta)data use a controlled vocabulary?

Yes

SAREF4ENER is used, the extension of Smart Applications REference (SAREF) suite of ontologies for the Energy domain.

<https://saref.etsi.org/index.html>

3.3.3 Have you applied a standard schema for your (meta)data?

Yes

3.3.5 What is the methodology followed?

To be decided

3.3.6 What community-endorsed interoperability best practices are followed?

Use of SAREF4ENER

3.3.7 Does the described dataset / output provide qualified references with other outputs?

Yes

The time-series dataset contains energy consumption information from residential and commercial buildings participating in the pilot. This is linked to a static dataset which contains the characteristics of the pilot population.

3.4 Increasing data and other outputs reuse

3.4.1 What internationally recognised licence will you use for your dataset / output?

To be decided

3.4.2 What reusability and / or reproducibility methods are followed?

- Readme files
- Variable definitions
- Units of measurement

3.4.3 Will you provide the described dataset / output in the public domain?

No

3.4.4 Do you intend to ensure (re)use by third parties after your project finishes?

No

3.4.5 Is provenance well documented?

No

Not applicable

3.4.6 What documented procedures for quality assurance do you have in place?

Set up of scientific and technical committee

4.1 Allocation of resources

4.1.1 What will be the cost of making the described output FAIR?

The cost is covered by the EC funding of the project

4.1.2 How will this cost be covered?

Infrastructure Grant

The cost is covered by the EC funding of the project

4.1.3 Identify the people who will be responsible and their role(s) in the management of the described output

Stratos Keranidis (orcid:0000-0002-0923-5020)

5.1 Data Security

5.1.1 What security measures are followed?

- Encryption
- Passwords
- Other

Strong passwords, Encryption mechanisms, Pseudonymisation principles

5.1.2 What conditions do the security measures meet?

- Data access
- Data storage

- Data sharing

5.1.3 How will you preserve the described dataset / output in the long term?

Data will be stored and maintained in DOMX data infrastructures.

6.1 Ethical aspects

6.1.1 Are there any ethical or legal issues that can have an impact on sharing the described dataset / output?

yes

Dataset is proprietary and sharing contracts need to be in place for usage (within &) out of the scope of this project.

6.1.2 Does the described dataset / output contain sensitive information?

No

6.1.3 Does the described dataset / output contain personal data?

Yes

7.1 Other

7.1.1 Do you make use of other procedures for data management?

Yes

7.1.2 Documentation of other procedures

e.g. GDPR compliance

Powered by



Part of

AI-DAPT AI-Ops Framework for Automated, Intelligent and Reliable Data/AI Pipelines Lifecycle with Humans-in-the-Loop and Coupling of Hybrid Science-Guided and AI Models

Description

Demonstrator 3 (Energy) - DOMX - Simulated Temporal power consumption dataset

Description

Synthetic time-series dataset generated for 2 years for 100 residential and commercial buildings. This represents the baseline power consumption for all buildings contributing to the Temporal power consumption dataset. The baseline consumption is synthetically generated for each building using heat conduction models, under the assumption that the DOMX smart thermostat functionality is disabled.

Attributes include:

Heating: Indoor/outdoor temperature, climate comfort, heating/hot water usage

User: Comfort limits, user schedules and preferences, app interactions

Energy: instant power, energy per usage scenario

All parameters are simulated at maximum rate per minute.

Researchers

Stratos Keranidis (orcid:0000-0002-0923-5020)

Description

1.1 Brief description of the described research output

1.1.1 What kind of research output are you describing?

Research Data

1.1.2 Is it physical or digital?

Digital

1.1.3 Are you generating or re-using it?

Re-used

This synthetic dataset has been used so far by DOMX for the personalized calculation of energy saving, under comparison to the real temporal power consumption data, and will be re-used within the scope of the project.

1.1.4 What is the type of the described dataset?

Simulation

Synthetic time-series dataset generated for 2 years for 100 residential and commercial buildings, as the baseline power consumption without enabling the energy saving DOMX thermostat.

1.1.5 What is its format?

CSV format

1.1.6 What is its expected size?

approximately 5-10GB

1.1.7 Why are you collecting/generating or re-using it?

- To share information
- To make informed decisions
- To improve a product
- To combine with other data
- Other

This information is shared by DOMX with the AI-DAPT consortium, to be used for experimentation in the development and evaluation of the project's AI-Ops framework, also in the development & validation of specific solutions for the demonstrator. Since the data providers (DOMX) already use this information for predictive analytics and the calculation of energy saving with the DOMX smart thermostat, its re-use aims to the improvement of accuracy in the personalized prediction of energy loads so that the company is assisted to make more informed decisions.

1.1.8 What is its origin / provenance?

Data generated by DOMX and owned by DOMX clients, provided upon their consent for the research purposes of the project.

1.1.9 To whom might it be useful ('data utility')?

- Researchers
- Economy
- Industry

Researchers in academia or the energy industry may use this data for the statistical/correlation/predictive analysis of energy consumption patterns and influencing factors.

2.1 Publications

2.1.1 Does the described output support any scientific publication?

a. Yes

A novel system for providing explicit demand response from domestic natural gas boilers

Applied Energy

b. Yes

Minimization of natural gas consumption of domestic boilers with convolutional long-short term memory neural networks and genetic algorithm

Applied Energy

2.1.2 Is there a data availability statement provided along with the publication?

No

2.3 Software

2.3.1 Does the described output use or support any software?

Yes

<https://domx.io>

3.1.1 Making data findable, including provisions for metadata

3.1.1.1 What type(s) of persistent identifier(s) are used for the described dataset / output?

Data identifiers

3.1.1.2 Will you provide metadata for the described dataset / output?

Yes

3.1.1.3 What type(s) of metadata?

Descriptive

3.1.1.4 Do the metadata use standardised vocabularies?

Yes

SAREF4ENER is used, the extension of Smart Applications REFerence (SAREF) suite of ontologies for the Energy domain.

3.1.1.5 Please provide URL/Description of used vocabularies

<https://saref.etsi.org/index.html>

3.1.1.6 Are the metadata searchable?

No

3.1.1.8 Are keywords provided in the metadata?

No

3.1.1.9 Are metadata harvestable?

No

3.2.1 Repository

3.2.1.1 In which repository will the dataset / output be deposited?

AI-DAPT repository/database

3.2.1.2 Is the selected repository a trusted source?

Yes

- Has certification

- Supports authentication and authorization of users
- Has data security mechanisms in place

3.2.1.5 Does the repository(ies) assign datasets / outputs with persistent identifiers?

No

3.2.2 Data

3.2.2.1 What is the described dataset / output title?

Synthetic time-series dataset on baseline energy consumption - DOMX

3.2.2.2 How is the dataset / output shared?

Shared

The dataset is expected to be shared under restricted access policies. Access will be granted to technical support partners of the energy demonstrator.

3.2.2.3 What is the reason of limiting access to the dataset / output?

Access to this dataset is limited due to its contents including personal information of the company's customers.

3.2.2.5 Are there any methods or tools required to access the dataset / output?

No

3.2.2.8 Is the described dataset / output supported by a data access committee?

Yes

3.2.2.9 Please specify how the dataset / output will be accessed during and after the project ends

Restricted access to authorized users, sharing contracts

3.2.3 Metadata

3.2.3.1 Will you provide metadata even if the described dataset / output can not be openly shared?

Yes

3.2.3.3 Do metadata provide information about how to access the described dataset / output?

No

3.2.3.4 Will metadata remain available after the dataset / output is no longer available?

No

3.3 Making data and other outputs interoperable

3.3.1 Does your (meta)data use a controlled vocabulary?

Yes

SAREF4ENER is used, the extension of Smart Applications REference (SAREF) suite of ontologies for the Energy domain.

<https://saref.etsi.org/index.html>

3.3.3 Have you applied a standard schema for your (meta)data?

Yes

3.3.5 What is the methodology followed?

To be decided

3.3.6 What community-endorsed interoperability best practices are followed?

Use of SAREF4ENER

3.3.7 Does the described dataset / output provide qualified references with other outputs?

Yes

The synthetic time-series dataset contains baseline (i.e. with the DOMX smart thermostat disabled) energy consumption information for residential and commercial buildings participating in the pilot. This is linked to the DOMX temporal power consumption dataset which contains the real energy consumption for these buildings, with the smart thermostat operation enabled. .

3.4 Increasing data and other outputs reuse

3.4.1 What internationally recognised licence will you use for your dataset / output?

To be decided

3.4.2 What reusability and / or reproducibility methods are followed?

- Readme files
- Variable definitions
- Units of measurement

- Other

API description

3.4.3 Will you provide the described dataset / output in the public domain?

No

3.4.4 Do you intend to ensure (re)use by third parties after your project finishes?

No

3.4.5 Is provenance well documented?

No

Not applicable

3.4.6 What documented procedures for quality assurance do you have in place?

Set up of scientific and technical committee

4.1 Allocation of resources

4.1.1 What will be the cost of making the described output FAIR?

The cost is covered by the EC funding of the project

4.1.2 How will this cost be covered?

Infrastructure Grant

The cost is covered by the EC funding of the project

4.1.3 Identify the people who will be responsible and their role(s) in the management of the described output

Stratos Keranidis (orcid:0000-0002-0923-5020)

5.1 Data Security

5.1.1 What security measures are followed?

- Encryption
- Passwords
- Other

Strong passwords, Encryption mechanisms, Pseudonymisation principles

5.1.2 What conditions do the security measures meet?

- Data access
- Data storage
- Data sharing

5.1.3 How will you preserve the described dataset / output in the long term?

Data will be stored and maintained in DOMX data infrastructures.

6.1 Ethical aspects

6.1.1 Are there any ethical or legal issues that can have an impact on sharing the described dataset / output?

yes

Dataset is proprietary and sharing contracts need to be in place for usage (within &) out of the scope of this project.

6.1.2 Does the described dataset / output contain sensitive information?

No

6.1.3 Does the described dataset / output contain personal data?

Yes

7.1 Other

7.1.1 Do you make use of other procedures for data management?

Yes

7.1.2 Documentation of other procedures

e.g. GDPR compliance

Powered by



Part of

AI-DAPT AI-Ops Framework for Automated, Intelligent and Reliable Data/AI Pipelines Lifecycle with Humans-in-the-Loop and Coupling of Hybrid Science-Guided and AI Models

Description

Demonstrator 3 (Energy) - DOMX - Static dataset on pilot population

Description

Static dataset characterizing the pilot population (contributing to the time-series dataset's measurements).

Attributes include:

Building data: size, energy class, construction year, approximate location, heating zone

Occupants: # of occupants, age groups, income level

Contract type: fixed, dynamic, kWh price

Researchers

Stratos Keranidis (orcid:0000-0002-0923-5020)

Description

1.1 Brief description of the described research output

1.1.1 What kind of research output are you describing?

Research Data

1.1.2 Is it physical or digital?

Digital

1.1.3 Are you generating or re-using it?

Re-used

This data has been used so far by DOMX for predictive analysis of energy loads, and will be re-used within the scope of the project.

1.1.4 What is the type of the described dataset?

Observational

Data collected by DOMX through surveys from pilot users.

1.1.5 What is its format?

CSV format

1.1.6 What is its expected size?

Approximately 3 MB

1.1.7 Why are you collecting/generating or re-using it?

- To obtain information
- To share information
- To improve a product
- To combine with other data

1.1.8 What is its origin / provenance?

Data collected by DOMX through surveys from pilot users. Owned by DOMX clients, provided upon their consent for the research purposes of the project.

1.1.9 To whom might it be useful ('data utility')?

- Researchers
- Economy
- Industry

2.1 Publications

2.1.1 Does the described output support any scientific publication?

No

2.1.2 Is there a data availability statement provided along with the publication?

No

2.3 Software

2.3.1 Does the described output use or support any software?

No

3.1.1 Making data findable, including provisions for metadata

3.1.1.1 What type(s) of persistent identifier(s) are used for the described dataset / output?

Data identifiers

3.1.1.2 Will you provide metadata for the described dataset / output?

No

3.2.1 Repository

3.2.1.1 In which repository will the dataset / output be deposited?

AI-DAPT repository/database

3.2.1.2 Is the selected repository a trusted source?

Yes

- Has certification
- Supports authentication and authorization of users
- Has data security mechanisms in place

3.2.1.5 Does the repository(ies) assign datasets / outputs with persistent identifiers?

No

3.2.1.7 Does the repository support versioning?

No

3.2.2 Data

3.2.2.1 What is the described dataset / output title?

Static dataset characterizing the pilot population for residential and commercial building energy consumption

3.2.2.2 How is the dataset / output shared?

Shared

The dataset is expected to be shared under restricted access policies. Access will be granted to technical support partners of the energy demonstrator.

3.2.2.3 What is the reason of limiting access to the dataset / output?

Access to this dataset is limited due to its contents including personal information of the company's customers.

3.2.2.5 Are there any methods or tools required to access the dataset / output?

No

3.2.2.8 Is the described dataset / output supported by a data access committee?

Yes

3.2.2.9 Please specify how the dataset / output will be accessed during and after the project ends

Restricted access to authorized users, sharing contracts

3.2.3 Metadata

3.2.3.1 Will you provide metadata even if the described dataset / output can not be openly shared?

No

3.2.3.3 Do metadata provide information about how to access the described dataset / output?

No

3.2.3.4 Will metadata remain available after the dataset / output is no longer available?

No

3.3 Making data and other outputs interoperable

3.3.1 Does your (meta)data use a controlled vocabulary?

Yes

SAREF4ENER is used, the extension of Smart Applications REFerence (SAREF) suite of ontologies for the Energy domain.

<https://saref.etsi.org/index.html>

3.3.3 Have you applied a standard schema for your (meta)data?

No

3.3.4 Will you provide a mapping to more commonly used ontologies?

Yes

3.3.5 What is the methodology followed?

TBD

3.3.6 What community-endorsed interoperability best practices are followed?

Use of SAREF4ENER

3.3.7 Does the described dataset / output provide qualified references with other outputs?

Yes

This static dataset contains characteristics of the pilot population contributing to the energy consumption documented in the time-series dataset of residential and commercial buildings.

3.4 Increasing data and other outputs reuse

3.4.1 What internationally recognised licence will you use for your dataset / output?

To be decided

3.4.2 What reusability and / or reproducibility methods are followed?

- Readme files
- Variable definitions
- Units of measurement

3.4.3 Will you provide the described dataset / output in the public domain?

No

3.4.4 Do you intend to ensure (re)use by third parties after your project finishes?

No

3.4.5 Is provenance well documented?

No

Not applicable

3.4.6 What documented procedures for quality assurance do you have in place?

Set up of scientific and technical committee

4.1 Allocation of resources

4.1.1 What will be the cost of making the described output FAIR?

The cost is covered by the EC funding of the project

4.1.2 How will this cost be covered?

Infrastructure Grant

The cost is covered by the EC funding of the project

4.1.3 Identify the people who will be responsible and their role(s) in the management of the described output

Stratos Keranidis (orcid:0000-0002-0923-5020)

5.1 Data Security

5.1.1 What security measures are followed?

- Encryption
- Passwords
- Other

Strong passwords, Encryption mechanisms, Pseudonymisation principles

5.1.2 What conditions do the security measures meet?

- Data access
- Data storage

- Data sharing

5.1.3 How will you preserve the described dataset / output in the long term?

Data will be stored and maintained in DOMX data infrastructures.

6.1 Ethical aspects

6.1.1 Are there any ethical or legal issues that can have an impact on sharing the described dataset / output?

yes

Dataset is proprietary and sharing contracts need to be in place for usage out of the scope of this project.

6.1.2 Does the described dataset / output contain sensitive information?

No

6.1.3 Does the described dataset / output contain personal data?

Yes

7.1 Other

7.1.1 Do you make use of other procedures for data management?

Yes

7.1.2 Documentation of other procedures

e.g. GDPR compliance

Powered by



Part of

AI-DAPT AI-Ops Framework for Automated, Intelligent and Reliable Data/AI Pipelines Lifecycle with Humans-in-the-Loop and Coupling of Hybrid Science-Guided and AI Models

Description

Demonstrator 3 (Energy) - ZENITH - Individual Historical gas/power consumption

Description

Individual historical gas / power consumption dataset: on the delivery point (house)

Attributes: region, municipality, post code, address, point of delivery, etc

Researchers

Dimitris Bibikas (orcid:0000-0002-2962-5026)

Description

1.1 Brief description of the described research output

1.1.1 What kind of research output are you describing?

Research Data

1.1.2 Is it physical or digital?

Digital

1.1.3 Are you generating or re-using it?

Re-used

This data has been used so far by ZENITH for predictive analysis of gas/power loads, and will be re-used within the scope of the project.

1.1.4 What is the type of the described dataset?

Observational

1.1.5 What is its format?

Excel format

1.1.6 What is its expected size?

Approximately 5 MB

1.1.7 Why are you collecting/generating or re-using it?

- To share information
- To make informed decisions
- To improve a product
- To combine with other data

This information is shared by ZENITH with the AI-DAPT consortium, to be used for experimentation in the development and evaluation of the project's AI-Ops framework, also in the development & validation of specific solutions for the demonstrator. Since the data providers (ZENITH) already use this information for predictive analytics, its re-use aims to the improvement of accuracy in the prediction of gas/energy loads so that the company is assisted to make more informed decisions.

1.1.8 What is its origin / provenance?

Data collected by ZENITH from the company's clients upon their consent

1.1.9 To whom might it be useful ('data utility')?

- Researchers

- Economy
- Industry

Researchers in academia or the energy industry may use this data for the statistical/correlation/predictive analysis of gas/power consumption patterns and influencing factors.

2.1 Publications

2.1.1 Does the described output support any scientific publication?

No

2.1.2 Is there a data availability statement provided along with the publication?

No

2.3 Software

2.3.1 Does the described output use or support any software?

No

3.1.1 Making data findable, including provisions for metadata

3.1.1.1 What type(s) of persistent identifier(s) are used for the described dataset / output?

Data identifiers

3.1.1.2 Will you provide metadata for the described dataset / output?

Yes

3.1.1.3 What type(s) of metadata?

Descriptive

3.1.1.4 Do the metadata use standardised vocabularies?

No

3.1.1.6 Are the metadata searchable?

No

3.1.1.8 Are keywords provided in the metadata?

No

3.1.1.9 Are metadata harvestable?

No

3.2.1 Repository

3.2.1.1 In which repository will the dataset / output be deposited?

AI-DAPT repository/database

3.2.1.2 Is the selected repository a trusted source?

Yes

- Has certification
- Supports authentication and authorization of users
- Has data security mechanisms in place

3.2.1.5 Does the repository(ies) assign datasets / outputs with persistent identifiers?

No

3.2.2 Data

3.2.2.1 What is the described dataset / output title?

Individual Historical gas / power consumption dataset - ZENITH

3.2.2.2 How is the dataset / output shared?

Shared

The dataset is expected to be shared under restricted access policies. Access will be granted to technical support partners of the energy demonstrator.

3.2.2.3 What is the reason of limiting access to the dataset / output?

Access to this dataset is limited due to its contents including personal information of the company's customers.

3.2.2.5 Are there any methods or tools required to access the dataset / output?

No

3.2.2.8 Is the described dataset / output supported by a data access committee?

Yes

3.2.2.9 Please specify how the dataset / output will be accessed during and after the project ends

Restricted access to authorized users, sharing contracts

3.2.3 Metadata

3.2.3.1 Will you provide metadata even if the described dataset / output can not be openly shared?

Yes

3.2.3.3 Do metadata provide information about how to access the described dataset / output?

No

3.2.3.4 Will metadata remain available after the dataset / output is no longer available?

No

3.3 Making data and other outputs interoperable

3.3.1 Does your (meta)data use a controlled vocabulary?

Yes

SAREF4ENER is used, the extension of Smart Applications REference (SAREF) suite of ontologies for the Energy domain.

<https://saref.etsi.org/index.html>

3.3.3 Have you applied a standard schema for your (meta)data?

No

3.3.4 Will you provide a mapping to more commonly used ontologies?

Yes

To be decided

3.3.5 What is the methodology followed?

To be decided

3.3.6 What community-endorsed interoperability best practices are followed?

Use of SAREF4ENER for the energy domain

3.3.7 Does the described dataset / output provide qualified references with other outputs?

Yes

This dataset contains individual historical gas/power consumption data, which are then aggregated into the Large-scale (aggregated historical & geographical) gas/power consumption dataset.

3.4 Increasing data and other outputs reuse

3.4.1 What internationally recognised licence will you use for your dataset / output?

The dataset is proprietary and will not be made publicly available.

3.4.2 What reusability and / or reproducibility methods are followed?

- Readme files
- Variable definitions
- Units of measurement

3.4.3 Will you provide the described dataset / output in the public domain?

No

3.4.4 Do you intend to ensure (re)use by third parties after your project finishes?

No

3.4.5 Is provenance well documented?

No

Not applicable

3.4.6 What documented procedures for quality assurance do you have in place?

Set up of scientific and technical committee

4.1 Allocation of resources

4.1.1 What will be the cost of making the described output FAIR?

The cost is covered by the EC funding of the project

4.1.2 How will this cost be covered?

Infrastructure Grant

The cost is covered by the EC funding of the project

4.1.3 Identify the people who will be responsible and their role(s) in the management of the described output

Dimitris Bibikas (orcid:0000-0002-2962-5026)

5.1 Data Security

5.1.1 What security measures are followed?

Other

Anonymization

5.1.2 What conditions do the security measures meet?

- Data access
- Data storage
- Data sharing

5.1.3 How will you preserve the described dataset / output in the long term?

Data will be stored and maintained in ZENITH data infrastructures.

6.1 Ethical aspects

6.1.1 Are there any ethical or legal issues that can have an impact on sharing the described dataset / output?

yes

Dataset is proprietary and sharing contracts need to be in place for usage within and out of the scope of this project.

6.1.2 Does the described dataset / output contain sensitive information?

No

6.1.3 Does the described dataset / output contain personal data?

Yes

7.1 Other

7.1.1 Do you make use of other procedures for data management?

No

Powered by



Part of

AI-DAPT AI-Ops Framework for Automated, Intelligent and Reliable Data/AI Pipelines Lifecycle with Humans-in-the-Loop and Coupling of Hybrid Science-Guided and AI Models

Description

Demonstrator 3 (Energy) - ZENITH - Aggregated Historical gas/power demand

Description

Large-scale (aggregated) gas / power consumption dataset: on a national & historical level.

Collected in Hourly/Daily intervals

Researchers

Dimitris Bibikas (orcid:0000-0002-2962-5026)

Description

1.1 Brief description of the described research output

1.1.1 What kind of research output are you describing?

Research Data

1.1.2 Is it physical or digital?

Digital

1.1.3 Are you generating or re-using it?

Re-used

1.1.4 What is the type of the described dataset?

Derived or compiled

Aggregation of individual gas/power consumption data

1.1.5 What is its format?

Excel format

1.1.6 What is its expected size?

Approximately 300 MB

1.1.7 Why are you collecting/generating or re-using it?

- To share information
- To make informed decisions
- To improve a product
- To combine with other data

This information is shared by ZENITH with the AI-DAPT consortium, to be used for experimentation in the development and evaluation of the project's AI-Ops framework, also in the development & validation of specific solutions for the demonstrator. Since the data providers (ZENITH) already use this information for predictive analytics, its re-use aims to the improvement of accuracy in the prediction of gas/energy loads so that the company is assisted to make more informed decisions.

1.1.8 What is its origin / provenance?

Data collected by ZENITH from the company's clients upon their consent and aggregated on a national level

1.1.9 To whom might it be useful ('data utility')?

- Researchers

- Economy
- Industry

Researchers in academia or the energy industry may use this data for the statistical/correlation/predictive analysis of gas/power consumption patterns and influencing factors.

2.1 Publications

2.1.1 Does the described output support any scientific publication?

No

2.1.2 Is there a data availability statement provided along with the publication?

No

2.3 Software

2.3.1 Does the described output use or support any software?

No

3.1.1 Making data findable, including provisions for metadata

3.1.1.1 What type(s) of persistent identifier(s) are used for the described dataset / output?

None

3.1.1.2 Will you provide metadata for the described dataset / output?

Yes

3.1.1.3 What type(s) of metadata?

Descriptive

3.1.1.4 Do the metadata use standardised vocabularies?

No

3.1.1.6 Are the metadata searchable?

No

3.1.1.8 Are keywords provided in the metadata?

No

To be decided

3.1.1.9 Are metadata harvestable?

No

3.2.1 Repository

3.2.1.1 In which repository will the dataset / output be deposited?

AI-DAPT repository/database

3.2.1.2 Is the selected repository a trusted source?

Yes

- Supports authentication and authorization of users
- Has data security mechanisms in place

3.2.1.5 Does the repository(ies) assign datasets / outputs with persistent identifiers?

No

3.2.2 Data

3.2.2.1 What is the described dataset / output title?

Aggregated historical gas / power consumption dataset - ZENITH

3.2.2.2 How is the dataset / output shared?

Shared

The dataset is expected to be shared under restricted access policies. Access will be granted to technical support partners of the energy demonstrator.

3.2.2.3 What is the reason of limiting access to the dataset / output?

Aggregated demand and price data based on official Energy Exchanges are generally open access. Access to a particular customer dataset is limited due to its contents including personal information.

3.2.2.5 Are there any methods or tools required to access the dataset / output?

No

3.2.2.8 Is the described dataset / output supported by a data access committee?

No

3.2.3 Metadata

3.2.3.1 Will you provide metadata even if the described dataset / output can not be openly shared?

No

3.2.3.3 Do metadata provide information about how to access the described dataset / output?

No

3.2.3.4 Will metadata remain available after the dataset / output is no longer available?

No

3.3 Making data and other outputs interoperable

3.3.1 Does your (meta)data use a controlled vocabulary?

Yes

SAREF4ENER is used, the extension of Smart Applications REference (SAREF) suite of ontologies for the Energy domain.

<https://saref.etsi.org/index.html>

3.3.3 Have you applied a standard schema for your (meta)data?

No

3.3.4 Will you provide a mapping to more commonly used ontologies?

No

3.3.6 What community-endorsed interoperability best practices are followed?

Use of SAREF4ENER

3.3.7 Does the described dataset / output provide qualified references with other outputs?

Yes

This is an aggregated dataset (market level) of the individual gas/power consumption dataset

3.4 Increasing data and other outputs reuse

3.4.1 What internationally recognised licence will you use for your dataset / output?

The dataset is proprietary and will not be made publicly available.

3.4.2 What reusability and / or reproducibility methods are followed?

- Readme files
- Variable definitions
- Units of measurement

3.4.3 Will you provide the described dataset / output in the public domain?

No

3.4.4 Do you intend to ensure (re)use by third parties after your project finishes?

No

3.4.5 Is provenance well documented?

No

Not applicable

3.4.6 What documented procedures for quality assurance do you have in place?

Set up of scientific and technical committee

4.1 Allocation of resources

4.1.1 What will be the cost of making the described output FAIR?

The cost is covered by the EC funding of the project

4.1.2 How will this cost be covered?

Infrastructure Grant

The cost is covered by the EC funding of the project

4.1.3 Identify the people who will be responsible and their role(s) in the management of the described output

Dimitris Bibikas (orcid:0000-0002-2962-5026)

5.1 Data Security

5.1.1 What security measures are followed?

- Firewall
- Passwords
- Other

Passwords, Private Networks, Firewall, anonymization, etc.

5.1.2 What conditions do the security measures meet?

- Data access
- Data storage
- Data sharing

5.1.3 How will you preserve the described dataset / output in the long term?

Data will be stored and maintained in ZENITH data infrastructures.

6.1 Ethical aspects

6.1.1 Are there any ethical or legal issues that can have an impact on sharing the described dataset / output?

yes

Dataset is proprietary and sharing contracts need to be in place for usage within and out of the scope of this project.

6.1.2 Does the described dataset / output contain sensitive information?

No

6.1.3 Does the described dataset / output contain personal data?

Yes

7.1 Other

7.1.1 Do you make use of other procedures for data management?

No

Powered by



Part of

AI-DAPT AI-Ops Framework for Automated, Intelligent and Reliable Data/AI Pipelines Lifecycle with Humans-in-the-Loop and Coupling of Hybrid Science-Guided and AI Models

Description

Demonstrator 3 (Energy) - ZENITH - Aggregated Historical gas/power prices

Description

Large-scale (aggregated) gas / power prices dataset: on a historical and national level.

Collected in Hourly/Daily intervals.

Researchers

Dimitris Bibikas (orcid:0000-0002-2962-5026)

Description

1.1 Brief description of the described research output

1.1.1 What kind of research output are you describing?

Research Data

1.1.2 Is it physical or digital?

Digital

1.1.3 Are you generating or re-using it?

Re-used

1.1.4 What is the type of the described dataset?

Derived or compiled

Aggregation of individual gas/power prices data

1.1.5 What is its format?

Excel format

1.1.6 What is its expected size?

Approximately 300 MB

1.1.7 Why are you collecting/generating or re-using it?

- To share information
- To make informed decisions
- To improve a product
- To combine with other data

This information is shared by ZENITH with the AI-DAPT consortium, to be used for experimentation in the development and evaluation of the project's AI-Ops framework, also in the development & validation of specific solutions for the demonstrator. Since the data providers (ZENITH) already use this information for predictive analytics, its re-use aims to the improvement of accuracy in the prediction of gas/energy loads so that the company is assisted to make more informed decisions.

1.1.8 What is its origin / provenance?

Data collected by ZENITH from the company's clients upon their consent and aggregated on a national level.

Also, data from Hellenic Energy Exchange S.A.

1.1.9 To whom might it be useful ('data utility')?

- Researchers

- Economy
- Industry

Researchers in academia or the energy industry may use this data for the statistical/correlation/predictive analysis of gas/power consumption patterns and influencing factors.

2.1 Publications

2.1.1 Does the described output support any scientific publication?

No

2.1.2 Is there a data availability statement provided along with the publication?

No

2.3 Software

2.3.1 Does the described output use or support any software?

No

3.1.1 Making data findable, including provisions for metadata

3.1.1.1 What type(s) of persistent identifier(s) are used for the described dataset / output?

None

3.1.1.2 Will you provide metadata for the described dataset / output?

Yes

3.1.1.3 What type(s) of metadata?

Descriptive

3.1.1.4 Do the metadata use standardised vocabularies?

No

3.1.1.6 Are the metadata searchable?

No

3.1.1.8 Are keywords provided in the metadata?

No

To be decided

3.1.1.9 Are metadata harvestable?

No

3.2.1 Repository

3.2.1.1 In which repository will the dataset / output be deposited?

AI-DAPT repository/database

3.2.1.2 Is the selected repository a trusted source?

Yes

- Supports authentication and authorization of users
- Has data security mechanisms in place

3.2.1.5 Does the repository(ies) assign datasets / outputs with persistent identifiers?

No

3.2.2 Data

3.2.2.1 What is the described dataset / output title?

Aggregated historical & geographical gas / power prices dataset - ZENITH

3.2.2.2 How is the dataset / output shared?

Shared

The dataset is expected to be shared under restricted access policies. Access will be granted to technical support partners of the energy demonstrator.

3.2.2.3 What is the reason of limiting access to the dataset / output?

Aggregated demand and price data based on official Energy Exchanges are generally open access. Access to a particular customer dataset is limited due to its contents including personal information.

3.2.2.5 Are there any methods or tools required to access the dataset / output?

No

3.2.2.8 Is the described dataset / output supported by a data access committee?

Yes

3.2.3 Metadata

3.2.3.1 Will you provide metadata even if the described dataset / output can not be openly shared?

No

3.2.3.3 Do metadata provide information about how to access the described dataset / output?

No

3.2.3.4 Will metadata remain available after the dataset / output is no longer available?

No

3.3 Making data and other outputs interoperable

3.3.1 Does your (meta)data use a controlled vocabulary?

Yes

SAREF4ENER is used, the extension of Smart Applications REference (SAREF) suite of ontologies for the Energy domain.

<https://saref.etsi.org/index.html>

3.3.3 Have you applied a standard schema for your (meta)data?

No

3.3.4 Will you provide a mapping to more commonly used ontologies?

No

3.3.6 What community-endorsed interoperability best practices are followed?

Use of SAREF4ENER

3.3.7 Does the described dataset / output provide qualified references with other outputs?

Yes

This is an aggregated dataset (market level) of gas/power consumption dataset

3.4 Increasing data and other outputs reuse

3.4.1 What internationally recognised licence will you use for your dataset / output?

The dataset is proprietary and will not be made publicly available.

3.4.2 What reusability and / or reproducibility methods are followed?

- Readme files
- Variable definitions
- Units of measurement

3.4.3 Will you provide the described dataset / output in the public domain?

No

3.4.4 Do you intend to ensure (re)use by third parties after your project finishes?

No

3.4.5 Is provenance well documented?

No

Not applicable

3.4.6 What documented procedures for quality assurance do you have in place?

Set up of scientific and technical committee

4.1 Allocation of resources

4.1.1 What will be the cost of making the described output FAIR?

The cost is covered by the EC funding of the project

4.1.2 How will this cost be covered?

Infrastructure Grant

The cost is covered by the EC funding of the project

4.1.3 Identify the people who will be responsible and their role(s) in the management of the described output

Dimitris Bibikas (orcid:0000-0002-2962-5026)

5.1 Data Security

5.1.1 What security measures are followed?

- Firewall
- Passwords
- Other

Passwords, Private Networks, Firewall, anonymization, etc.

5.1.2 What conditions do the security measures meet?

- Data access
- Data storage
- Data sharing

5.1.3 How will you preserve the described dataset / output in the long term?

Data will be stored and maintained in ZENITH data infrastructures.

6.1 Ethical aspects

6.1.1 Are there any ethical or legal issues that can have an impact on sharing the described dataset / output?

yes

Dataset is proprietary and sharing contracts need to be in place for usage within and out of the scope of this project.

6.1.2 Does the described dataset / output contain sensitive information?

No

6.1.3 Does the described dataset / output contain personal data?

Yes

7.1 Other

7.1.1 Do you make use of other procedures for data management?

No

Powered by



Part of

AI-DAPT AI-Ops Framework for Automated, Intelligent and Reliable Data/AI Pipelines Lifecycle with Humans-in-the-Loop and Coupling of Hybrid Science-Guided and AI Models

Description

Demonstrator 4 (Manufacturing) OHS - Equipment

Description

Dataset holding information on equipment utilized during maintenance processes.

Attributes include:

name, category, equipmentNumber, instanceNumber, sequenceNumber, manufacturerPartNumber

Researchers

Carl Hans

Description

1.1 Brief description of the described research output

1.1.1 What kind of research output are you describing?

Research Data

1.1.2 Is it physical or digital?

Digital

1.1.3 Are you generating or re-using it?

Re-used

Generated by OHS

1.1.4 What is the type of the described dataset?

Derived or compiled

Data on equipment used in maintenance processes.

1.1.5 What is its format?

Can be acquired in CSV/Excel/JSON format

1.1.6 What is its expected size?

The size will be defined upon export

1.1.7 Why are you collecting/generating or re-using it?

- To obtain information
- To share information
- To make informed decisions
- To develop a product
- To combine with other data

1.1.8 What is its origin / provenance?

Dataset owned by OHS

1.1.9 To whom might it be useful ('data utility')?

Researchers

2.1 Publications

2.1.1 Does the described output support any scientific publication?

No

2.1.2 Is there a data availability statement provided along with the publication?

No

2.3 Software

2.3.1 Does the described output use or support any software?

Yes

Smartmaintain

3.1.1 Making data findable, including provisions for metadata

3.1.1.1 What type(s) of persistent identifier(s) are used for the described dataset / output?

Data identifiers

TBD

3.1.1.2 Will you provide metadata for the described dataset / output?

No

3.2.1 Repository

3.2.1.1 In which repository will the dataset / output be deposited?

AI-DAPT repository/database

3.2.1.2 Is the selected repository a trusted source?

Yes

- Has certification
- Supports authentication and authorization of users
- Has data security mechanisms in place

3.2.1.5 Does the repository(ies) assign datasets / outputs with persistent identifiers?

No

3.2.2 Data

3.2.2.1 What is the described dataset / output title?

OHS - Equipment

3.2.2.2 How is the dataset / output shared?

Shared

Exclusively shared by OHS with the AI-DAPT consortium in the context of Demonstrator 4 (Manufacturing).

3.2.2.3 What is the reason of limiting access to the dataset / output?

Personal data of the end customers are confidential.

3.2.2.5 Are there any methods or tools required to access the dataset / output?

No

3.2.2.8 Is the described dataset / output supported by a data access committee?

No

3.2.2.9 Please specify how the dataset / output will be accessed during and after the project ends

Restricted access to authorized users, sharing contracts

3.2.3 Metadata

3.2.3.1 Will you provide metadata even if the described dataset / output can not be openly shared?

No

3.2.3.3 Do metadata provide information about how to access the described dataset / output?

No

3.2.3.4 Will metadata remain available after the dataset / output is no longer available?

No

3.3 Making data and other outputs interoperable

3.3.1 Does your (meta)data use a controlled vocabulary?

No

3.3.3 Have you applied a standard schema for your (meta)data?

No

3.3.4 Will you provide a mapping to more commonly used ontologies?

No

3.3.7 Does the described dataset / output provide qualified references with other outputs?

Yes

References to Equipment categories, Maintenance processes, Additional efforts, Reporting datasets provided by OHS.

3.4 Increasing data and other outputs reuse

3.4.1 What internationally recognised licence will you use for your dataset / output?

To be determined

3.4.2 What reusability and / or reproducibility methods are followed?

- Readme files
- Variable definitions

3.4.3 Will you provide the described dataset / output in the public domain?

No

3.4.4 Do you intend to ensure (re)use by third parties after your project finishes?

No

3.4.5 Is provenance well documented?

No

3.4.6 What documented procedures for quality assurance do you have in place?

Set up of scientific and technical committee

4.1 Allocation of resources

4.1.1 What will be the cost of making the described output FAIR?

The cost is covered by the EC funding of the project

4.1.2 How will this cost be covered?

Infrastructure Grant

The cost is covered by the EC funding of the project

4.1.3 Identify the people who will be responsible and their role(s) in the management of the described output

Carl Hans, Managing Director OHS Engineering GmbH

5.1 Data Security

5.1.1 What security measures are followed?

- Encryption

- Firewall
- Passwords

To be determined

5.1.2 What conditions do the security measures meet?

- Data access
- Data storage
- Data sharing

5.1.3 How will you preserve the described dataset / output in the long term?

OHS data infrastructures (postgres DBMS)

6.1 Ethical aspects

6.1.1 Are there any ethical or legal issues that can have an impact on sharing the described dataset / output?

yes

Proprietary data of OHS clients

6.1.2 Does the described dataset / output contain sensitive information?

Yes

6.1.3 Does the described dataset / output contain personal data?

Yes

6.1.4 What are the methods used for processing and accessing sensitive/personal information?

Anonymising data where necessary

7.1 Other

7.1.1 Do you make use of other procedures for data management?

No

Powered by



Part of

AI-DAPT AI-Ops Framework for Automated, Intelligent and Reliable Data/AI Pipelines Lifecycle with Humans-in-the-Loop and Coupling of Hybrid Science-Guided and AI Models

Description

Demonstrator 4 (Manufacturing) OHS - Equipment category

Description

Dataset holding information on the categories of equipment utilized during maintenance processes.

Attributes include: Field Name, name, description, remark, subCategory, assetsOfCategory, owningProgramme, numberOfAssets, maintenaceEffortClass, criticality, activities

Fields with missing values: activities

Researchers

Carl Hans

Description

1.1 Brief description of the described research output

1.1.1 What kind of research output are you describing?

Research Data

1.1.2 Is it physical or digital?

Digital

1.1.3 Are you generating or re-using it?

Re-used

Generated by OHS

1.1.4 What is the type of the described dataset?

Derived or compiled

Data on equipment categories used in maintenance processes.

1.1.5 What is its format?

Can be acquired in CSV/Excel/JSON format

1.1.6 What is its expected size?

The size will be determined upon export of the file

1.1.7 Why are you collecting/generating or re-using it?

- To obtain information
- To share information
- To make informed decisions
- To develop a product
- To combine with other data

1.1.8 What is its origin / provenance?

Dataset owned by OHS

1.1.9 To whom might it be useful ('data utility')?

Researchers

2.1 Publications

2.1.1 Does the described output support any scientific publication?

No

2.1.2 Is there a data availability statement provided along with the publication?

No

2.3 Software

2.3.1 Does the described output use or support any software?

Yes

Smartmaintain

3.1.1 Making data findable, including provisions for metadata

3.1.1.1 What type(s) of persistent identifier(s) are used for the described dataset / output?

Data identifiers

To be determined

3.1.1.2 Will you provide metadata for the described dataset / output?

No

3.1.1.5 Please provide URL/Description of used vocabularies

To be determined

3.2.1 Repository

3.2.1.1 In which repository will the dataset / output be deposited?

AI-DAPT repository/database

3.2.1.2 Is the selected repository a trusted source?

Yes

- Has certification
- Supports authentication and authorization of users
- Has data security mechanisms in place

3.2.1.5 Does the repository(ies) assign datasets / outputs with persistent identifiers?

No

3.2.2 Data

3.2.2.1 What is the described dataset / output title?

OHS - Equipment category

3.2.2.2 How is the dataset / output shared?

Shared

Exclusively shared by OHS with the AI-DAPT consortium in the context of Demonstrator 4 (Manufacturing).

3.2.2.3 What is the reason of limiting access to the dataset / output?

Personal data of the end customers are confidential.

3.2.2.5 Are there any methods or tools required to access the dataset / output?

No

3.2.2.8 Is the described dataset / output supported by a data access committee?

No

3.2.2.9 Please specify how the dataset / output will be accessed during and after the project ends

Restricted access to authorized users, sharing contracts

3.2.3 Metadata

3.2.3.1 Will you provide metadata even if the described dataset / output can not be openly shared?

No

3.2.3.3 Do metadata provide information about how to access the described dataset / output?

No

3.2.3.4 Will metadata remain available after the dataset / output is no longer available?

No

3.3 Making data and other outputs interoperable

3.3.1 Does your (meta)data use a controlled vocabulary?

No

3.3.3 Have you applied a standard schema for your (meta)data?

No

3.3.4 Will you provide a mapping to more commonly used ontologies?

No

3.3.7 Does the described dataset / output provide qualified references with other outputs?

Yes

References to Equipment, Maintenance processes, Additional efforts, Reporting datasets provided by OHS.

3.4 Increasing data and other outputs reuse

3.4.1 What internationally recognised licence will you use for your dataset / output?

To be determined

3.4.2 What reusability and / or reproducibility methods are followed?

- Readme files
- Variable definitions

3.4.3 Will you provide the described dataset / output in the public domain?

No

3.4.4 Do you intend to ensure (re)use by third parties after your project finishes?

No

3.4.5 Is provenance well documented?

No

3.4.6 What documented procedures for quality assurance do you have in place?

Set up of scientific and technical committee

4.1 Allocation of resources

4.1.1 What will be the cost of making the described output FAIR?

The cost is covered by the EC funding of the project

4.1.2 How will this cost be covered?

Infrastructure Grant

The cost is covered by the EC funding of the project

4.1.3 Identify the people who will be responsible and their role(s) in the management of the described output

Carl Hans

Managing Director OHS Engineering GmbH

5.1 Data Security

5.1.1 What security measures are followed?

- Encryption
- Firewall
- Passwords

To be determined

5.1.2 What conditions do the security measures meet?

- Data access
- Data storage
- Data sharing

5.1.3 How will you preserve the described dataset / output in the long term?

OHS data infrastructures (postgres DBMS)

6.1 Ethical aspects

6.1.1 Are there any ethical or legal issues that can have an impact on sharing the described dataset / output?

yes

Proprietary data of OHS clients

6.1.2 Does the described dataset / output contain sensitive information?

Yes

6.1.3 Does the described dataset / output contain personal data?

Yes

6.1.4 What are the methods used for processing and accessing sensitive/personal information?

Anonymising data where necessary

7.1 Other

7.1.1 Do you make use of other procedures for data management?

No

Part of

AI-DAPT AI-Ops Framework for Automated, Intelligent and Reliable Data/AI Pipelines Lifecycle with Humans-in-the-Loop and Coupling of Hybrid Science-Guided and AI Models

Description

Demonstrator 4 (Manufacturing) OHS - Maintenance processes

Description

Dataset holding operational data, one record per process, updates during maintenance process.

Attributes include: name, asset, equipmentNumber, category, repair, estimatedDelivery, location, expectedDeliveryDate, etc.

Fields with missing values: location, expectedDeliveryDate

Researchers

Carl Hans

Description

1.1 Brief description of the described research output

1.1.1 What kind of research output are you describing?

Research Data

1.1.2 Is it physical or digital?

Digital

1.1.3 Are you generating or re-using it?

Re-used

Generated by OHS

1.1.4 What is the type of the described dataset?

Derived or compiled

Operational data on maintenance processes.

1.1.5 What is its format?

Can be acquired in CSV/Excel/JSON format

1.1.6 What is its expected size?

Approximately 3 MB

1.1.7 Why are you collecting/generating or re-using it?

- To obtain information
- To share information
- To make informed decisions
- To develop a product
- To combine with other data

1.1.8 What is its origin / provenance?

Dataset owned by OHS

1.1.9 To whom might it be useful ('data utility')?

Researchers

2.1 Publications

2.1.1 Does the described output support any scientific publication?

No

2.1.2 Is there a data availability statement provided along with the publication?

No

2.3 Software

2.3.1 Does the described output use or support any software?

Yes

Smartmaintain

3.1.1 Making data findable, including provisions for metadata

3.1.1.1 What type(s) of persistent identifier(s) are used for the described dataset / output?

Data identifiers

TBD

3.1.1.2 Will you provide metadata for the described dataset / output?

No

3.1.1.5 Please provide URL/Description of used vocabularies

N/A

3.2.1 Repository

3.2.1.1 In which repository will the dataset / output be deposited?

AI-DAPT repository/database

3.2.1.2 Is the selected repository a trusted source?

Yes

- Has certification
- Supports authentication and authorization of users
- Has data security mechanisms in place

3.2.1.5 Does the repository(ies) assign datasets / outputs with persistent identifiers?

No

3.2.2 Data

3.2.2.1 What is the described dataset / output title?

OHS - Maintenance processes

3.2.2.2 How is the dataset / output shared?

Shared

Exclusively shared by OHS with the AI-DAPT consortium in the context of Demonstrator 4 (Manufacturing).

3.2.2.3 What is the reason of limiting access to the dataset / output?

Personal and corporate data of the end customers are confidential.

3.2.2.5 Are there any methods or tools required to access the dataset / output?

No

3.2.2.8 Is the described dataset / output supported by a data access committee?

No

3.2.2.9 Please specify how the dataset / output will be accessed during and after the project ends

Restricted access to authorized users, sharing contracts

3.2.3 Metadata

3.2.3.1 Will you provide metadata even if the described dataset / output can not be openly shared?

No

3.2.3.3 Do metadata provide information about how to access the described dataset / output?

No

3.2.3.4 Will metadata remain available after the dataset / output is no longer available?

No

3.3 Making data and other outputs interoperable

3.3.1 Does your (meta)data use a controlled vocabulary?

No

3.3.3 Have you applied a standard schema for your (meta)data?

No

3.3.4 Will you provide a mapping to more commonly used ontologies?

No

3.3.5 What is the methodology followed?

N/A

3.3.6 What community-endorsed interoperability best practices are followed?

N/A

3.3.7 Does the described dataset / output provide qualified references with other outputs?

Yes

References to Equipment, Equipment categories, Additional efforts, Reporting datasets provided by OHS.

3.4 Increasing data and other outputs reuse

3.4.1 What internationally recognised licence will you use for your dataset / output?

N/A

3.4.2 What reusability and / or reproducibility methods are followed?

- Readme files
- Variable definitions

3.4.3 Will you provide the described dataset / output in the public domain?

No

3.4.4 Do you intend to ensure (re)use by third parties after your project finishes?

No

3.4.5 Is provenance well documented?

No

3.4.6 What documented procedures for quality assurance do you have in place?

Set up of scientific and technical committee

4.1 Allocation of resources

4.1.1 What will be the cost of making the described output FAIR?

The cost is covered by the EC funding of the project

4.1.2 How will this cost be covered?

Infrastructure Grant

The cost is covered by the EC funding of the project

4.1.3 Identify the people who will be responsible and their role(s) in the management of the described output

Carl Hans

Managing Director OHS Engineering GmbH

5.1 Data Security

5.1.1 What security measures are followed?

- Encryption
- Firewall
- Passwords

TBD

5.1.2 What conditions do the security measures meet?

- Data access
- Data storage
- Data sharing

5.1.3 How will you preserve the described dataset / output in the long term?

OHS data infrastructures (postgres DBMS)

6.1 Ethical aspects

6.1.1 Are there any ethical or legal issues that can have an impact on sharing the described dataset / output?

yes

Proprietary data of OHS clients

6.1.2 Does the described dataset / output contain sensitive information?

Yes

6.1.3 Does the described dataset / output contain personal data?

Yes

6.1.4 What are the methods used for processing and accessing sensitive/personal information?

Anonymising data where necessary

Names of workers responsible for specific tasks. Will be anonymized.

7.1 Other

7.1.1 Do you make use of other procedures for data management?

No

Powered by



Part of

AI-DAPT AI-Ops Framework for Automated, Intelligent and Reliable Data/AI Pipelines Lifecycle with Humans-in-the-Loop and Coupling of Hybrid Science-Guided and AI Models

Description

Demonstrator 4 (Manufacturing) OHS - Additional efforts

Description

Dataset holding information on additional efforts during maintenance, including spare parts, unplanned repair activities, updates with new efforts and activities therein.

Attributes include: Field Name, name, description, remark, subCategory, assetsOfCategory, owningProgramme, numberOfAssets, maintenanceEffortClass, criticality, activities etc.

Fields with missing values: activities

Researchers

Carl Hans

Description

1.1 Brief description of the described research output

1.1.1 What kind of research output are you describing?

Research Data

1.1.2 Is it physical or digital?

Digital

1.1.3 Are you generating or re-using it?

Re-used

Generated by OHS

1.1.4 What is the type of the described dataset?

Derived or compiled

Data on additional efforts required during maintenance processes.

1.1.5 What is its format?

Can be acquired in CSV/Excel/JSON format

1.1.6 What is its expected size?

The size will be defined upon file export

1.1.7 Why are you collecting/generating or re-using it?

- To obtain information
- To share information
- To make informed decisions
- To develop a product
- To combine with other data

1.1.8 What is its origin / provenance?

Dataset owned by OHS

1.1.9 To whom might it be useful ('data utility')?

Researchers

2.1 Publications

2.1.1 Does the described output support any scientific publication?

No

2.1.2 Is there a data availability statement provided along with the publication?

No

2.3 Software

2.3.1 Does the described output use or support any software?

Yes

Smartmaintain

3.1.1 Making data findable, including provisions for metadata

3.1.1.1 What type(s) of persistent identifier(s) are used for the described dataset / output?

Data identifiers

To be determined

3.1.1.2 Will you provide metadata for the described dataset / output?

No

3.2.1 Repository

3.2.1.1 In which repository will the dataset / output be deposited?

AI-DAPT repository/database

3.2.1.2 Is the selected repository a trusted source?

Yes

- Has certification
- Supports authentication and authorization of users
- Has data security mechanisms in place

3.2.1.5 Does the repository(ies) assign datasets / outputs with persistent identifiers?

No

3.2.2 Data

3.2.2.1 What is the described dataset / output title?

OHS - Additional efforts

3.2.2.2 How is the dataset / output shared?

Shared

Exclusively shared by OHS with the AI-DAPT consortium in the context of Demonstrator 4 (Manufacturing).

3.2.2.3 What is the reason of limiting access to the dataset / output?

Personal data of the end customers are confidential.

3.2.2.5 Are there any methods or tools required to access the dataset / output?

No

3.2.2.8 Is the described dataset / output supported by a data access committee?

No

3.2.2.9 Please specify how the dataset / output will be accessed during and after the project ends

Restricted access to authorized users, sharing contracts

3.2.3 Metadata

3.2.3.1 Will you provide metadata even if the described dataset / output can not be openly shared?

No

3.2.3.3 Do metadata provide information about how to access the described dataset / output?

No

3.2.3.4 Will metadata remain available after the dataset / output is no longer available?

No

3.3 Making data and other outputs interoperable

3.3.1 Does your (meta)data use a controlled vocabulary?

No

3.3.3 Have you applied a standard schema for your (meta)data?

No

3.3.4 Will you provide a mapping to more commonly used ontologies?

No

3.3.5 What is the methodology followed?

To be determined

3.3.6 What community-endorsed interoperability best practices are followed?

To be determined

3.3.7 Does the described dataset / output provide qualified references with other outputs?

Yes

References to Maintenance processes, Equipment, Equipment category, Reporting datasets provided by OHS.

3.4 Increasing data and other outputs reuse

3.4.1 What internationally recognised licence will you use for your dataset / output?

To be determined

3.4.2 What reusability and / or reproducibility methods are followed?

- Readme files
- Variable definitions

3.4.3 Will you provide the described dataset / output in the public domain?

No

3.4.4 Do you intend to ensure (re)use by third parties after your project finishes?

No

3.4.5 Is provenance well documented?

No

3.4.6 What documented procedures for quality assurance do you have in place?

Set up of scientific and technical committee

4.1 Allocation of resources

4.1.1 What will be the cost of making the described output FAIR?

The cost is covered by the EC funding of the project

4.1.2 How will this cost be covered?

Infrastructure Grant

The cost is covered by the EC funding of the project

4.1.3 Identify the people who will be responsible and their role(s) in the management of the described output

Carl Hans

Managing Director OHS Engineering GmbH

5.1 Data Security

5.1.1 What security measures are followed?

- Encryption
- Firewall
- Passwords

To be determined

5.1.2 What conditions do the security measures meet?

- Data access
- Data storage
- Data sharing

5.1.3 How will you preserve the described dataset / output in the long term?

OHS data infrastructures (postgres DBMS)

6.1 Ethical aspects

6.1.1 Are there any ethical or legal issues that can have an impact on sharing the described dataset / output?

yes

Proprietary data of OHS clients

6.1.2 Does the described dataset / output contain sensitive information?

Yes

6.1.3 Does the described dataset / output contain personal data?

Yes

6.1.4 What are the methods used for processing and accessing sensitive/personal information?

Anonymising data where necessary

7.1 Other

7.1.1 Do you make use of other procedures for data management?

No

Powered by



Part of

AI-DAPT AI-Ops Framework for Automated, Intelligent and Reliable Data/AI Pipelines Lifecycle with Humans-in-the-Loop and Coupling of Hybrid Science-Guided and AI Models

Description

Demonstrator 4 (Manufacturing) OHS - Reporting

Description

Dataset with reporting spreadsheets, automatically generated on a weekly/quarterly basis.

Researchers

Carl Hans

Description

1.1 Brief description of the described research output

1.1.1 What kind of research output are you describing?

Research Data

1.1.2 Is it physical or digital?

Digital

1.1.3 Are you generating or re-using it?

Re-used

Generated by OHS

1.1.4 What is the type of the described dataset?

Derived or compiled

Reporting spreadsheets on maintenance processes.

1.1.5 What is its format?

Can be acquired in CSV/Excel/JSON format

1.1.6 What is its expected size?

The size will be defined upon file export.

1.1.7 Why are you collecting/generating or re-using it?

- To obtain information
- To share information
- To make informed decisions
- To develop a product
- To combine with other data

1.1.8 What is its origin / provenance?

Dataset owned by OHS

1.1.9 To whom might it be useful ('data utility')?

Researchers

2.1 Publications

2.1.1 Does the described output support any scientific publication?

No

2.1.2 Is there a data availability statement provided along with the publication?

No

2.3 Software

2.3.1 Does the described output use or support any software?

Yes

Smartmaintain

3.1.1 Making data findable, including provisions for metadata

3.1.1.1 What type(s) of persistent identifier(s) are used for the described dataset / output?

Data identifiers

To be determined

3.1.1.2 Will you provide metadata for the described dataset / output?

No

3.2.1 Repository

3.2.1.1 In which repository will the dataset / output be deposited?

AI-DAPT repository/database

3.2.1.2 Is the selected repository a trusted source?

Yes

- Has certification
- Supports authentication and authorization of users
- Has data security mechanisms in place

3.2.1.5 Does the repository(ies) assign datasets / outputs with persistent identifiers?

No

3.2.2 Data

3.2.2.1 What is the described dataset / output title?

OHS - Reporting

3.2.2.2 How is the dataset / output shared?

Shared

Exclusively shared by OHS with the AI-DAPT consortium in the context of Demonstrator 4 (Manufacturing).

3.2.2.3 What is the reason of limiting access to the dataset / output?

Personal data of the end customers are confidential.

3.2.2.5 Are there any methods or tools required to access the dataset / output?

No

3.2.2.8 Is the described dataset / output supported by a data access committee?

No

3.2.3 Metadata

3.2.3.1 Will you provide metadata even if the described dataset / output can not be openly shared?

No

3.2.3.3 Do metadata provide information about how to access the described dataset / output?

No

3.2.3.4 Will metadata remain available after the dataset / output is no longer available?

No

3.3 Making data and other outputs interoperable

3.3.1 Does your (meta)data use a controlled vocabulary?

No

To be checked with OHS

3.3.3 Have you applied a standard schema for your (meta)data?

No

3.3.4 Will you provide a mapping to more commonly used ontologies?

No

3.3.6 What community-endorsed interoperability best practices are followed?

To be determined

3.3.7 Does the described dataset / output provide qualified references with other outputs?

Yes

References to Maintenance processes, Equipment, Equipment category, Additional efforts datasets provided by OHS.

3.4 Increasing data and other outputs reuse

3.4.1 What internationally recognised licence will you use for your dataset / output?

To be determined

3.4.2 What reusability and / or reproducibility methods are followed?

- Readme files
- Variable definitions

3.4.3 Will you provide the described dataset / output in the public domain?

No

3.4.4 Do you intend to ensure (re)use by third parties after your project finishes?

No

3.4.5 Is provenance well documented?

No

3.4.6 What documented procedures for quality assurance do you have in place?

Set up of scientific and technical committee

4.1 Allocation of resources

4.1.1 What will be the cost of making the described output FAIR?

The cost is covered by the EC funding of the project

4.1.2 How will this cost be covered?

Infrastructure Grant

The cost is covered by the EC funding of the project

4.1.3 Identify the people who will be responsible and their role(s) in the management of the described output

Carl Hans

Managing Director OHS Engineering GmbH

5.1 Data Security

5.1.1 What security measures are followed?

- Encryption
- Firewall
- Passwords

To be determined

5.1.2 What conditions do the security measures meet?

- Data access
- Data storage
- Data sharing

5.1.3 How will you preserve the described dataset / output in the long term?

OHS data infrastructures (postgres DBMS)

6.1 Ethical aspects

6.1.1 Are there any ethical or legal issues that can have an impact on sharing the described dataset / output?

yes

Proprietary data of OHS clients

6.1.2 Does the described dataset / output contain sensitive information?

No

6.1.3 Does the described dataset / output contain personal data?

Yes

7.1 Other

7.1.1 Do you make use of other procedures for data management?

No

Powered by

