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 compromizzing 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

 DOI: - 27/06/2024

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

DOI: - 27/06/2024

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 compromizzing 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)
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

Contact: Eleni Lavasa

 DOI: - 27/06/2024

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

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

 DOI: - 27/06/2024

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

 DOI: - 27/06/2024

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

 DOI: - 27/06/2024

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

 DOI: - 27/06/2024

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

 DOI: - 27/06/2024

- 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

 DOI: - 27/06/2024

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.

 DOI: - 27/06/2024

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

 DOI: - 27/06/2024

- 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?

 DOI: - 27/06/2024

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?

 DOI: - 27/06/2024

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

 DOI: - 27/06/2024

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

 DOI: - 27/06/2024

- 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

 DOI: - 27/06/2024

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)

 DOI: - 27/06/2024

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

 DOI: - 27/06/2024

• 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?

 DOI: - 27/06/2024

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.

 DOI: - 27/06/2024

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

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 in1 hour batches with a sampling frequency around 100Hz.

Researchers

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

 DOI: - 27/06/2024

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

 DOI: - 27/06/2024

- 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

 DOI: - 27/06/2024

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

 DOI: - 27/06/2024

7.1 Other

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

No

Powered by

 DOI: - 27/06/2024

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)

 DOI: - 27/06/2024

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

 DOI: - 27/06/2024

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?

 DOI: - 27/06/2024

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

 DOI: - 27/06/2024

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)

 DOI: - 27/06/2024

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?

 DOI: - 27/06/2024

No

Powered by

 DOI: - 27/06/2024

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)

DOI: -

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 &

 DOI: - 27/06/2024

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 impovement 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 concent 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

 DOI: - 27/06/2024

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

 DOI: - 27/06/2024

• 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

 DOI: - 27/06/2024

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?

 DOI: - 27/06/2024

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

 DOI: - 27/06/2024

• 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

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

 DOI: - 27/06/2024

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

 DOI: - 27/06/2024

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 concent 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

 DOI: - 27/06/2024

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

 DOI: - 27/06/2024

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

 DOI: - 27/06/2024

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

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)

 DOI: - 27/06/2024

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 concent for the research purposes of the project.

 DOI: - 27/06/2024

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?

 DOI: - 27/06/2024

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?

 DOI: - 27/06/2024

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?

 DOI: - 27/06/2024

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

 DOI: - 27/06/2024

• 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

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?

 DOI: - 27/06/2024

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 impovement 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

 DOI: - 27/06/2024

- 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?

 DOI: - 27/06/2024

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

 DOI: - 27/06/2024

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

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?

 DOI: - 27/06/2024

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 impovement 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

 DOI: - 27/06/2024

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

 DOI: - 27/06/2024

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

 DOI: - 27/06/2024

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 impovement 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

 DOI: - 27/06/2024

- 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

 DOI: - 27/06/2024

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

 DOI: - 27/06/2024

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

 DOI: - 27/06/2024

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?

 DOI: - 27/06/2024

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?

 DOI: - 27/06/2024

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.

 DOI: - 27/06/2024

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 - 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?

 DOI: - 27/06/2024

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 maintenace 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?

 DOI: - 27/06/2024

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

 DOI: - 27/06/2024

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?

 DOI: - 27/06/2024

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

 DOI: - 27/06/2024

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

 DOI: - 27/06/2024

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?

 DOI: - 27/06/2024

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

 DOI: - 27/06/2024

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?

 DOI: - 27/06/2024

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

 DOI: - 27/06/2024

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?

 DOI: - 27/06/2024

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

 DOI: - 27/06/2024

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

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 maintenace, including spare parts, unplanned repair activities, updates with new efforts and activities therein.

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

 DOI: - 27/06/2024

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 maintenace 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?

 DOI: - 27/06/2024

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

 DOI: - 27/06/2024

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

 DOI: - 27/06/2024

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?

 DOI: - 27/06/2024

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

 DOI: - 27/06/2024

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?

 DOI: - 27/06/2024

Re-used

Generated by OHS

1.1.4 What is the type of the described dataset?

Derived or compiled

Reporting spreadsheets on maintenace 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 e 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?

 DOI: - 27/06/2024

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.

 DOI: - 27/06/2024

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?

 DOI: - 27/06/2024

- 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