Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Federal Department of the Environment, Transport, Energy and Communications DETEC

Swiss Federal Office of Energy SFOE
Energy Research and Cleantech Division

Interim report dated 25 November 2024

AISOP

AI-assisted grid situational awareness and operational planning

Source: AISOP 2021

AISOP

www.aisopproject.com

Date: November 25th, 2024

Location: Bern

Publisher:
Swiss Federal Office of Energy SFOE
Energy Research and Cleantech
CH-3003 Bern
www.bfe.admin.ch

Subsidy recipients:

HSLU Technikumstrasse 21 6048 Horw	ETH Zürich, FEN Sonneggstrasse 28 8006 Zurich	Hive Power Sagl Via Motto Gandioni 17 6582 Pianezzo

Authors:

Braulio Barahona Garzon[1], braulio.barahonagarzon@hslu.ch
C. Yaman Evrenosoglu[2], evrenosoglu@ethz.ch
Adamantios Marinakis[2], marinakis@fen.ethz.ch
Philippe Buchecker[2], bucheckp@ethz.ch
Peter Allenspach[1], peter.allenspach@hslu.ch
Daniel Raimundo[1], daniel.raimundo@hslu.ch

[1] Lucerne University of Applied Sciences and Arts
[2] Research Center for Energy Networks (FEN), ETH Zürich

SFOE project coordinators:
Dr. Michael Moser, michael.moser@bfe.admin.ch

SFOE contract number: SI/502314-01

Summary

AISOP aims to create an **AI-assisted decision support system for the electric distribution system operators** (DSOs) to drive decarbonisation that is underpinned by **advanced digital technology**. The decision-support system securely and privately acquires, processes, interprets, and exploits data for the benefit of DSO operational planning. In this context, AISOP expands data-driven techniques for improved operational planning in distribution grids with high shares of DERs by integrating AI/ML-based solutions, enhanced situational awareness and market incentives. Within the proposed project we combine *(i)* data access and ingestion, *(ii)* distribution grid situational awareness, *(iii)* decision-support for distribution grid management, *(iv)* dynamic tariffs, and *(v)* digital platform integration with exploitation through test and training environments.

In the current reporting period, we focus on practical aspects of situational awareness for decision support, such as unsupervised anomaly detection and user interface. An approach based on dimension reduction and clustering is demonstrated on measurements from a grid sensor. The methodology for ML-based dynamic tariffs and the virtual demonstration are also part of this reporting period. Thus, the current approach based on regression and clustering applied on a rural grid is presented. Finally, an outlook to the final stage of the project is provided.

Contents

List of Figures

List of Tables

Abbreviations

AI	Artificial Intelligence
DER	Distributed Energy Resources
DG	Distributed Generation
DSM	Demand Side Management
DSO	Distribution System Operator
DPT	Digital Process Twin
DT	Digital Twin
GenAI	Generative AI
LLM	Large Language Model
LV	Low Voltage
ML	Machine Learning
MLT	Machine Learning Technique
MV	Medium Voltage
MVDS	Minimum Viable Data Space
NE	Netzebene (Grid level)
OP	Operational planning
OPF	Optimal Power Flow
PMU	Phasor Measurement Unit
PQ	Power Quality
RE	Romande Energie
RES	Renewable Energy Sources
SA	Grid Situational Awareness
SCADA	Supervisory Control and Data Acquisition
SFOE	Swiss Federal Office of Energy
SM	Smart Meter
SSoT	Single Source of Truth
ToU	Time-of-Use
TSO	Transmission System Operator
WWN	Westfalen-Weser Netz

1 Introduction

1.1 Context and motivation

Digitalisation of the electric energy systems creates opportunities to improve grid situational awareness and operational planning. As distribution grids incorporate more renewable energy sources and demand becomes more flexible (i.e., prosumers), more information about the current and future state of the grid becomes vital for operating the grid in a cost-effective way. Digitalization is therefore essential, as it facilitates data acquisition and processing. As distribution system operators (DSOs) explore the use of monitoring solutions, the volume of data and the associated costs increase. Thus, automated processes are required to manage and use energy system data to the advantage of DSOs. However, these processes need to ensure data protection and security and be designed in a way that improves the quality of underlying data sources.

AISOP aims at creating an AI-assisted decision support system for DSOs. The decision-support system securely and privately acquires data using state-of-the-art digital platforms. It then processes and interprets it to generate knowledge for situational awareness and dynamic tariff setting. Using heterogeneous data, the overall objective is to improve operational planning in active distribution grids by integrating AI- or ML-based solutions, enhanced situational awareness, and market incentives. Thus, it combines (i) data access and ingestion, (ii) distribution grid situational awareness, (iii) decision-support for distribution grid management, (iv) dynamic tariffs, and (v) digital platform integration.

Traditionally, operational planning prepares TSOs for real-time operation such that the probability of experiencing unexpected deviations in the balance of supply and demand is minimized. Such operational planning has not been necessary for distribution systems as the end-customers are only consumers of electricity. However, as the distribution systems are preparing for unprecedented levels of prosumers, DSOs will benefit from planning schemes in the long-term (decades time scale), the near-term (multiple years), operational planning schemes (intraday to years). Such operational planning schemes, need good information of the current and future grid situation [1-4]. Specific applications include better control renewable energies taking into account uncertainty [5], and dynamic pricing of electricity to incentivize flexibility of demand and ameliorate grid congestion issues [4,6]. The focus of AISOP lies on tools for situational awareness that serve to design dynamics tariffs and overall support DSO operation planning decisions. These tools are envisioned to inform on intra-day, day-ahead, and yearly timescales.

1.2 Project Objectives

The AISOP project objectives are to:

1. increase grid observability by using data from multiple sources and in different time resolutions,
2. help DSOs operate the grid using data-driven decision support tools,
3. improve the efficiency of network operations,
4. reduce curtailment of renewable energy and distributed energy resources, and to
5. improve options for tariffs for DSO’s and prosumers.

AISOP’s solutions will acquire, process, interpret and exploit data for the benefit of DSO operational planning, integrating AI/ML-based solutions, enhanced situational awareness, and market incentives. The project aims to create actionable, tangible, and applicable outcomes for distribution systems to improve operational planning and support decarbonisation. The outcomes will take the forms outlined in Table 1.

Table 1: AISOP project outcomes.

Methodologies and knowledge	Technologies	Services
Accessing and combining heterogenous, dispersed datasets	Data analytics (forecasting, local optimisation)	Dynamic tariffs
Developing grid situational awareness using edge and embedded network devices	ML-based anomaly detection and fault prediction	DSO congestion management
ML-based risk analysis and risk quantification	Digital process twin for distribution systems	Fault detection and prediction
AI/ML-based identification of dynamic tariffs for congestion management	Embedded and distributed sensors for LV and MV networks	Operational risk management
		Integration of community in digital platforms

In addition to the outcomes described above, AISOP will deliver environmental and socio-economic impacts as described in Figure 1.

Figure 1: Environmental and socio-economic objectives in AISOP.

2 Approach, method, results and discussion

This section describes grid situational awareness (SA), and dynamic tariffs approaches and analyses recently conducted. In Section 2.1 our grid situational awareness model is described, with focus on anomaly detection and user interface. A description of the approach to design dynamic tariffs is provided in Section 2.2 along with illustrations on a selected case study. Finally, Section 2.3 provides the description, plans and the progress in virtual demonstrators.

2.1 ML-based grid situational awareness

The goal of grid SA is to predict risky states of the grid. To achieve this goal, we defined four calculation workflows: sequential power flow simulations, power flow forecasting, anomaly detection, and risk assessment which are summarized below and described in more details in [8].

(a) **Sequential Power Flow Solutions** to characterize the impact of connecting more solar PV, electrical vehicles (EVs), and heat pumps.

(b) **Power Flow Forecasting** to estimate grid conditions in the next day(s).

(c) **Anomaly Detection** to detect faults such as short circuits, equipment failures, or incipient faults, and irregularities at the end-user side including new consumption patterns indicating new demand and generation from EVs and solar PV.

(d) **Risk Assessment** where compliance to EN 50160 is evaluated and risk metrics such as operational Over (or Under) Voltage Risk are calculated.

In this reporting period we focus on unsupervised anomaly detection in Section 2.1.1. Moreover, for these tools to be applicable, we look into how to organise data for visualization, and how to facilitate that users interact with analytics results in Section 2.1.2.

2.1.1 Unsupervised anomaly detection

It is essential to automate the analyses of data and the extraction of information to help operators identify situations where data may inform of potential cyber-attacks and risky operation states (i.e., anomalies with a relevant underlaying cause) or where data itself may be of poor quality (i.e., outliers due to missing or corrupted measurements). These two aspects are typically threated in different parts of the data ingestion and processing chain, once data from a given source has a minimum level of accessibility and quality, data analytics can start. Following, we describe an approach that can be applied to either step of the data analytics chain and demonstrate it with data recorded by a grid sensor [9]. In this way, we illustrate its application for SA where anomalies in voltages measured at the LV side of a distribution transformer with data obtained with a GridEye sensor.

Demonstration site. A district of the community of Rolle which combines commercial and residential customers was equipped, during the project “Romande Energie Electric network in local balance Demonstrator” (REEL Demo) within the context of SCCER-FURIES [9], with grid monitoring sensors and smart meters to develop, amongst other, activities on techno-economic and socio-economic methods for increasing energy flexibility. We tap on a curated data from the REEL demo that concerns a LV feeder in this district. As described in Table 2 we have access to heterogenous sources of data such as grid topology, grid sensor and smart meter data concerning a single feeder, also open data from the Swiss building register and weather data is accessible to enhance our analysis in future iterations.

Table 2: Description of available datasets in Rolle demonstration site.

Data	Description
Grid topology	Describes the LV grid of feeder 13 in Rolle, it is provided in matpower [11] format, and JSON format standard according to the open-source software pandapower. [12]
Grid sensor	Data recorded by a GridEye sensor between October 2018 to December 2019, 10-min values of power quality measurements: 58 quantities that describe voltage, power, current, harmonics, and grid frequency.
Smart meter	Small sample of active power data recorded by RE smart meters installed in the REEL Demo.
Building	Federal Register of Buildings and Dwellings (RBD) provides a view of the current buildings and dwellings stock of Switzerland, continuously updated by communal building departments. It includes energy relevant data such as energy reference area (GEBF), energy source for hot water (GENH1, GENH2) and for heating (GENW1, GENW2).
Weather	Historical and recent weather data is available for a large variety of providers most notably for solar radiation and ambient temperature, historical data from the National Solar Radiation Database (NSRDB) and measurements provided by MeteoSwiss stations.

Approach and results. Processing large amounts of grid monitoring data and detecting points that are different to the bulk is typically approached with data quality control rules or by means of unsupervised anomaly detection: statistics, thresholds set on the basis of domain knowledge, and unsupervised learning can be applied. Here we demonstrate a classical ML approach to unsupervised anomaly detection. First step is pre-processing the GridEye data by removing the mean and scaling the dataset to have unit variance. Then, removing linear correlations and performing dimension reduction by applying singular value decomposition are performed. A final step is to find those data points that are most different to the majority, in this case by applying density-based clustering method (DBSCAN) and manipulating two hyperparameters: a distance to neighbour threshold, and a number of samples around a cluster centre. Results are shown in Figure 2, where the 58 dimensions of the dataset are reduced to two dimensions that clearly synthetise the variability of the original data and let us, even visually, identify anomalies. On these two dimensions DBSCAN is applied to identify clusters. The results are then mapped back to the original data as shown in Figure 3 where the three anomalies highlighted in Figure 2 are found to show relatively odd power quality values as observed in the homopolar components, voltages, and harmonics metrics. Inverted and homopolar components are negative- and zero-sequence components of the three-phase voltages, thus higher values indicate that the system is not perfectly balanced giving also potential indications of faults. This is also reflected in the voltages as observed in the plot, while the effect is not so noticeable in the total harmonic distortion shown in the bottom subplot. This is consistent with the nature of power where load imbalances and faults are not necessarily linked to harmonics.

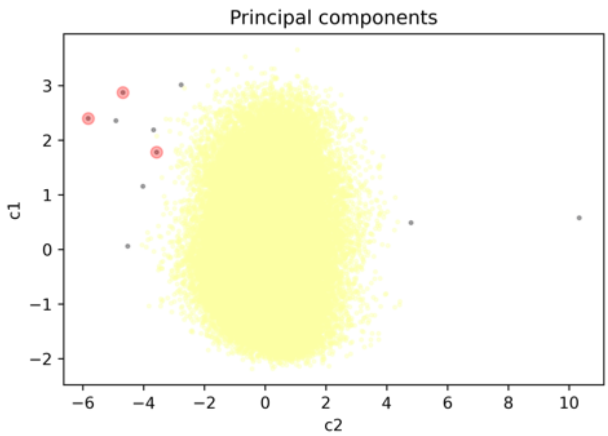

Figure 2: First two principal components of GridEye data and clusters derived with BDSCAN.

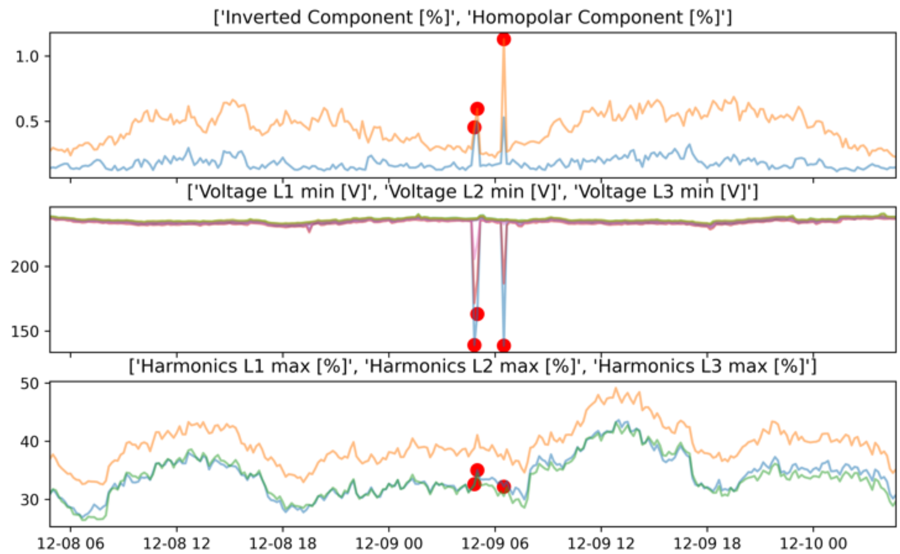

Figure 3: Anomalies detected in the GridEye power quality data by clustering in the first two principal components

Discussion. Benefits of this approach include that it is explainable, very easy to tune, and to implement on embedded hardware. Thus, opening the door to various applications such as interactive use where the user may select different data sources and adjust parameters to filter and select different number of anomalies. Also, it can be used for data compression and anomaly detection at the edge. We use efficient open-source implementations [10], which can be ported or rewritten to run in embedded devices subject mainly to memory constraints.

2.1.2 Data co-pilot concept

A key aspect of a decision support system is how its tools fit into established practices of operators. These practices are expected to adapt as digitisation and automation increases [12,13], particularly for TSOs, but DSO are also expected to benefit from increasing automation. Moreover, preferences for user interfaces and access to information as seen by the actual operators or engineers that make use of a given tool should also play a role in designing user interfaces for decision support tools. Understanding the preferences of users in the different DSO departments is important to design the information content and form of the interfaces to these users, which may include a graphical user interface (GUI), a command line for advance users, a natural language interface (e.g., Chatbot) possibly tapping into the capabilities of large language models (LLMs). As a first step, illustrated in Figure 4, the digital process twin (DPT) structure and main software building blocks applicable to the grid congestion use case were framed as a cyber-physical system to better visualise different concerns, as we aim for a modular AISOP Decision Support System. Details of the proposed architecture for the decision support system itself, are further described in Appendix and Figure 16.

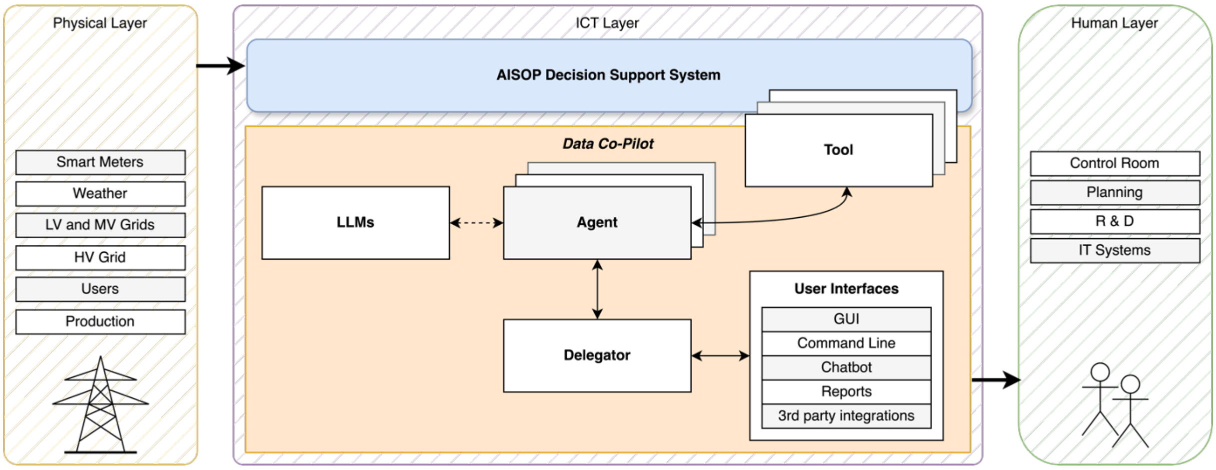

Figure 4: Data co-pilot concept consisting of user interfaces, a delegator that controls the main application flow, agents that abstract functionality by calling on tools and LLMs.

Web application interface. Generative AI, expressed as LLMs and foundation models, has taken a remarkable position in the development of AI applications, even in technical domains numerous simulation and data analytics commercial offers are developing functionalities based on these technologies to automate tasks and enhance user experience. A predominant approach to develop these applications resorts to agentic or multi-agent patterns, where LLM agents determine the application control flow. However, relying on LLMs to control task execution does not guarantee high reliability and increases cyber-security risks. Thus, in our concept LLMs have a limited scope of control and are meant to help with retrieving context data and facilitating a conversational interface. A first architecture consisting of User Interfaces, Delegator, Agent(s), Tool(s) and LLMs blocks is shown in the Data Co-Pilot block in Figure 4.

1. **Delegator.** Takes the form of a rule-based expert system, with a small knowledge base (i.e., definitions of agents, tasks, and tools) and an inference engine (i.e., set of rules). It has a tree structure to address each case depending on the inputs of the user. Its main function is to assign agents and tools to tasks.
2. **Agents.** Software object with states and memory, as well as minimal autonomy steaming mainly from rules. It can act and exchange data with other agents, and call tools within its limited scope (e.g., Hewitt's Actor model).
3. **Tools.** Hard coded functionalities that are part of workflows, or functions that wrap models or smaller analytics pipelines that are commonly used to return grid state information to the user.
4. **LLM.** It has the same hierarchy and more autonomy as an agent. Corresponds to agents in Langchain framework [Ref.StateofAgents] and autonomous agents utilizing LLMs in MetaGPT framework. [Ref.Hong2024] Here, their scope is limited to access tools that provide functions that are not essential for computational workflows.

AISOP Decision Support System. This block consists of several modules that facilitate operational planning by providing information about current and future grid states. They correspond to SA and dynamic tariffs work packages, whose outputs are combined to create a virtual demonstrator as described in Section 2.3.

Discussion. Feedback from technicians and operators involved in operational planning, as well as relevant IT-personnel is key to useful user interfaces. While interactive visualisations may be engaging, in some cases concrete information maybe more desirable. On the other hand, some advanced users may want to go beyond interactive visualisation and be able to perform analysis, for these users, tools as presented here are attractive. In terms of the web application development and implementation, we believe that an expert systems approach in combination with consolidated, machine-readable data could take us a long way, but the potential of generative AI cannot be underestimated and needs to be balanced to costs, information security and cybersecurity. We followed common implementations patterns but kept LLMs with a limited scope. Potential further development could include letting LLMs access data in different ‘security zones’ and if needed run LLMs on DSOs digital infrastructure in this way data never leaves the premises. Another application very popular in other domains is summarisation of company data.

2.2 ML-based dynamic tariff

2.2.1 Methodology

Objective: Dynamic tariffs are increasingly seen as a means for a utility to influence the power withdrawal and/or injection patterns of its customers, with objective to eventually reduce the grid loading whenever and wherever this might be desired due to otherwise excessive flows.

Designing such a tariff is far from being a trivial task. The utility needs to:

1. **Understand what exactly the tariff should try to influence (the target).** A utility needs to decide the measure (i.e., KPI) according to which it assesses the success of its tariff scheme. Following is a non-exhaustive list of such KPI candidates:
    - The loading of NL6 transformers
    - The loading of specific cables
    - The average voltage throughout the LV grid (NL7)
    - Nodal voltages at specific buses

    In addition to the above-listed spatial KPIs, temporal aspects can also be evaluated as part of the KPI: focusing only on specific moments in time (e.g., hours of date, months, seasons)
2. **Understand how the target can be influenced.**
    - Targeting the behaviour of specific customers or specific device types (spatial aspect)
    - Targeting the behaviour of customers at certain time intervals (temporal aspect)

Once steps 1 and 2 are completed, a utility can devise a tariff that aims at influencing the behaviour of those customers that we identified as the most relevant in step 2.

Depending on the level of **observability of its distribution network**, based on substation measurements, grid sensors, smart meters etc., a utility can have access to a heterogeneous set of data that contains potentially useful information. This data can be combined with **other sources of information**, such as weather measurements (i.e., solar irradiation, temperature). Such data can be collected over lengthy periods of time and grouped into datasets.

It is noted that, in addition to (or alternatively to) the measured data, these datasets can also include data created by means of simulation of various scenarios.

The objective of this work has been to develop methods and processes for a utility to **extract knowledge** from such datasets, which can be used to develop appropriate dynamic tariff rules and schemes.

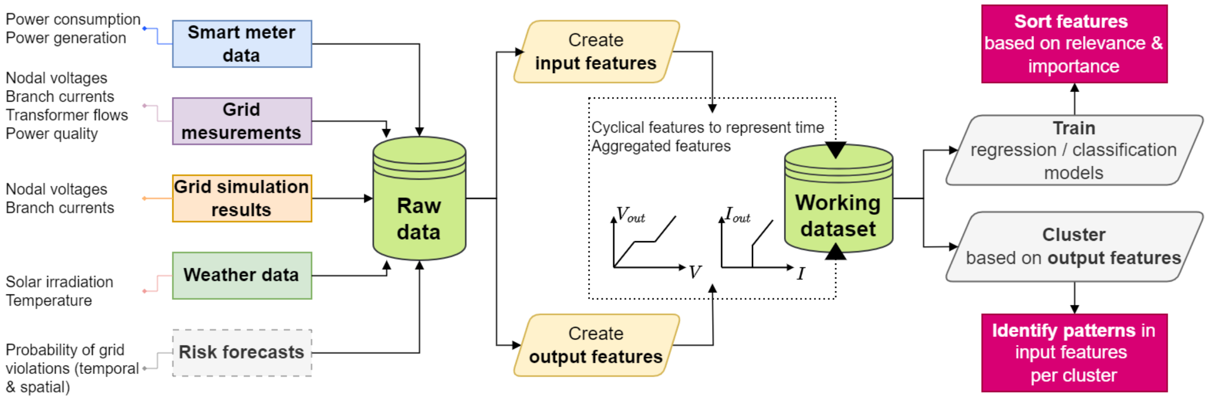

Figure 5: The framework for creating the ML-model to set the dynamic tariffs

Approach: In its essence, the proposed approach, illustrated in Figure 5, consists of two steps.

Processing the input data to create appropriate working datasets.

Extracting valuable knowledge from these datasets, in an automatic manner, by applying appropriate machine learning algorithms.

Step 1. Data processing

Data collection: The required raw input data is collected and/or created by means of simulations. Such data include, but are not limited to, voltage and current measurements at different network nodes and branches, active and reactive power injections and withdrawals at different network connection points, generation, and consumption by various devices (such as PVs, EV chargers, heat pumps), temperature, solar irradiance, and others. They are typically in the form of a time series, at various time resolutions. Other information that can be part of the "raw input data" includes calendar information, such as the hour of the day, the day of the week, month, public holiday information, etc.

The objective is to develop the methodology or a set of methodologies which can be used with various levels and types of data available to the utilities. Therefore, the methodology or the set of methodologies are not dependent on a given set of data.

Input feature creation: At this step, new synthetic input features are created by using the available raw data. The objective is to have features that represent situations not strongly reflected in each variable in the raw data. A representative list of such "data transformations" are the following:

- Creation of cyclical features to represent time
- Aggregate nodal power injections with topological criteria (e.g., along a feeder, downstream from a specific node etc.)
- Aggregate power generation or consumption per type of device

Output feature creation: At this step, different potential "target features" are created. As explained in the previous section, these features shall represent the objective that the utility tries to achieve by applying a dynamic tariff scheme. They will be used by the machine learning algorithms in step 2 (presented in the sequel), to drive the knowledge extraction results.

Two examples of such "target features" are:

- Per node, create a new feature for “nodal voltage in violation of the desired limits,” which is the nodal voltage magnitude (in p.u.) when the voltage exceeds a selected maximum value (e.g., 1.08 p.u.) or falls below a selected minimum value (e.g., 0.92 p.u.) Otherwise, it is 1 p.u. Such a feature allows the ML-model to distinguish over- or under-voltages while treating values in the acceptable range in the same manner, thus allowing the user to focus on cases when the power injections and withdrawals at electrically nearby nodes have to be influenced so that the voltage falls back into the desired interval.
- Per branch, create a new feature for “branch current in violation of desired thermal limits,” which is the branch loading (in %) when the loading is above a selected maximum value (e.g., 80%), and it is 0 otherwise. Such a feature allows the ML-model to distinguish branch currents that approach the limit while treating all the acceptable values in the same manner, thus allowing the ML-model to focus on cases when the power injections and withdrawals at electrically nearby nodes have to be influenced so that the branch current is under the selected limit.

Step 2. Knowledge extraction

Two techniques have been developed to enable automated processing of the data in order to eventually identify the input features that shall be the “targets to influence” by means of a dynamic tariff scheme.

Approach I. Identification of the most important input features by means of training of a regression model

General approach: This approach relies on the fact that a side-outcome of the training of certain types of regression models allow the ML-model to identify the candidate input features that turned out to be the most critical for achieving a high-quality model, which is as accurate as possible.

A suitable machine learning model that is selected to identify the most important input features is the “random forest.” Random forests are ensemble models. They are created by training many decision trees. The random forest model consists of all the trained decision trees. Its prediction is the average of the individual decision tree predictions. Different decision trees are obtained by repeatedly sampling the training dataset and creating diverse (different from each other) subsets.

During the training of each decision tree, the algorithm uses a metric to identify the feature to use to make the split at each tree node. A metric such as the “Gini importance index” is utilized to select the feature (and the feature value) for which the split at a node maximizes the decrease in impurity (i.e., randomness) of the data in its leaf below that node. Hence, a side-result of the process of training a random forest is that the value of each feature in splitting the data has been estimated many times, as the various trees are being built. Based on these calculations, a by-product of the training process of a random forest is a value per feature indicating its importance in efficiently splitting the data. Typically, this feature’s importance value is in a range from zero to one.

<u>Application</u>: We use this technique to identify the candidate input features, created in step 1, that are the most relevant for each target output feature created in step 1. For example, a random forest is trained as a predictor of the “voltage outside limits” feature (see step 1) for a given node in the network. A by-product of this training is that each candidate input feature (such as those described in step 1) will be assigned a value indicating its relevance. Obviously, the most relevant features are these that the utility shall aim at influencing via a dynamic tariff scheme.

Approach II. Utilization of clustering to identify a range of values of the input features associated with the target output features

Clustering is a powerful unsupervised machine learning technique. It splits a dataset into subsets, such that the data within each subset are “as similar to each other as possible” and “as different from the data in the other subsets as possible.”

<u>Application</u>: First, one or more target output features, among those computed in step 1, are selected. Following, the data are clustered based on the values of these target features. If the number of clusters is selected properly, some clusters will contain the data samples where one or more of the target output

features take non-desirable values (e.g., a cluster with nodal over voltages). Finally, the input features of each cluster are analysed, e.g., by performing basic statistical calculations, to identify the desired ranges of the input features, to avoid violations. This information is used to devise appropriate dynamic tariff schemes.

2.2.2 Case study

Example

A rural grid with 103 nodes and 105 branches is used for prototyping, testing and validating. For each grid connection point (i.e., Hausanschlusskasten – HAK), the electric heat pump demand time-series, the conventional household demand time-series, the solar PV time-series and EV charging time-series are created for nine (9) representative days (i.e., Workday, Saturday and Sunday in Winter, Summer and Transition seasons) in 15-minute resolution. The HP, EV and PV proliferation levels correspond to Energy Perspective scenarios for 2050. The methodology used to create the time-series for each grid connection point is described in the final report of TDFlex project[1].

Voltage violations (over- or under-) per each node and loading violations per each branch, defined according to the rule described in Step 1 above (see "output feature creation"), are used as the output features. Per node and per branch, the algorithm identifies the most influential input features. As a general rule, these input features are dependent on the output feature, i.e., different input features impact different violations.

Figure 7 illustrates the ten input features that are observed to be, on average, the most influential i.e., the input features that seem to influence most of the grid violations. Nine of these features refer to the aggregated net injection / withdrawal downstream of a give node. These are denoted with yellow colour arrows in the network diagram in Figure 6, and one feature refers to the net injection / withdrawal across a given feeder.

As an example, Figure 8. Five most relevant input features for the over-loading of the branch between nodes 16 and 17. DS-N-X stands for the aggregated net power injection/withdrawal downstream of node X. NI-N-X stands for the net power injection/withdrawal at node X. Min-Sin denotes a cyclical feature that was created by representing the time by means of two variables, a sine and a cosine, thus being able to express the time cyclicality (otherwise, for example, 23:00 would be interpreted by the algorithm as being very different from 00:00, while, in reality, they are equally close as, for example, 02:00 is to 03:00). The y axis indicates the relative importance of each feature. shows the five input features that are the most relevant for the (over-)loading of the branch connecting the nodes 16 and 17. In addition to the net injection downstream of given nodes and the net injection at a given node, the time turned out to also be a relevant feature for this overloading.

[1] C.Y. Evrenosoglu, J. Garrison, A. Fuchs and T. Demiray, "TDFlex – TSO-DSO Flexibility: towards integrated grid control and coordination in Switzerland," Swiss Federal Office of Energy, Final Report SI/501735, 2022.

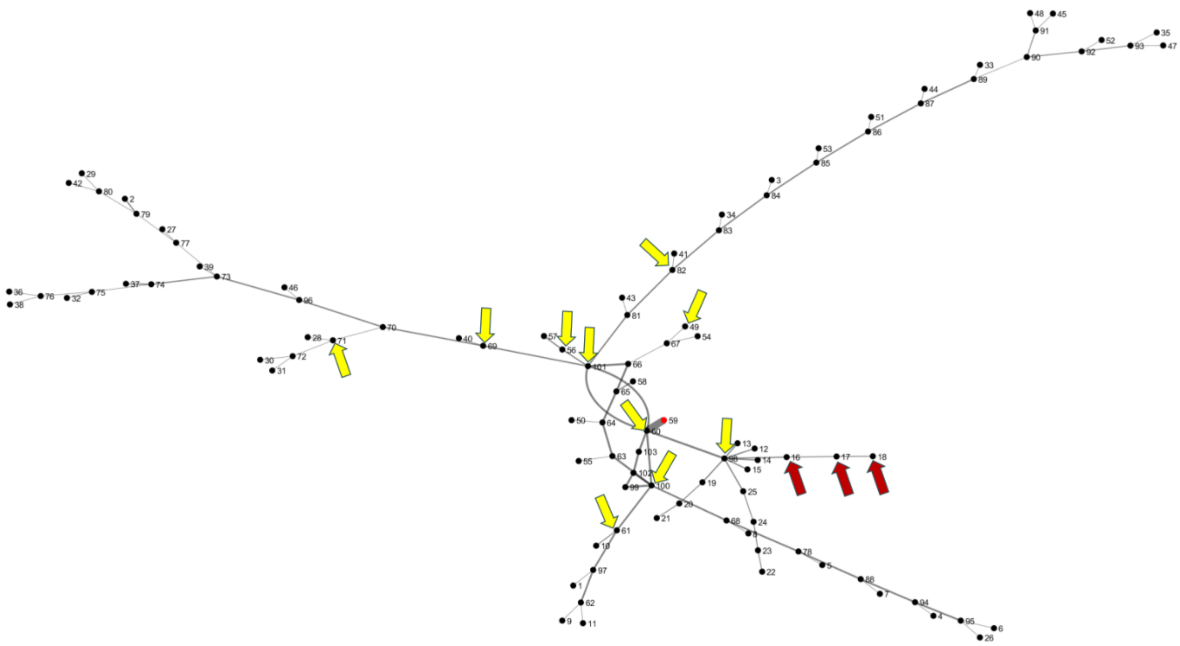

Figure 6. A rural grid used for prototyping ML-based tariff methodologies. Yellow colour arrows denote the nodes relating to the ten most influential features on average. Red colour arrows denote the nodes relating to the features that are more influential for the over-loading of the branch between the nodes 16 and 17.

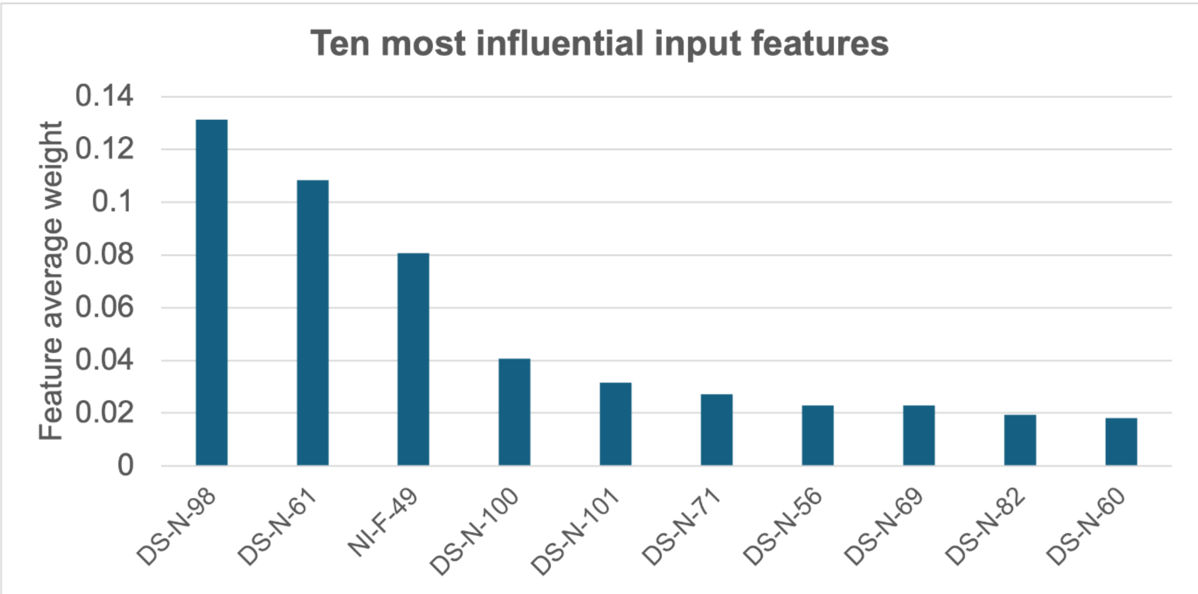

Figure 7. The ten most relevant features on average. DS-N-X stands for the aggregated net power injection/withdrawal downstream of node X. NI-F-X stands for the aggregated net power injection/withdrawal of all nodes across feeder X. The y axis indicates the relative importance of each feature.

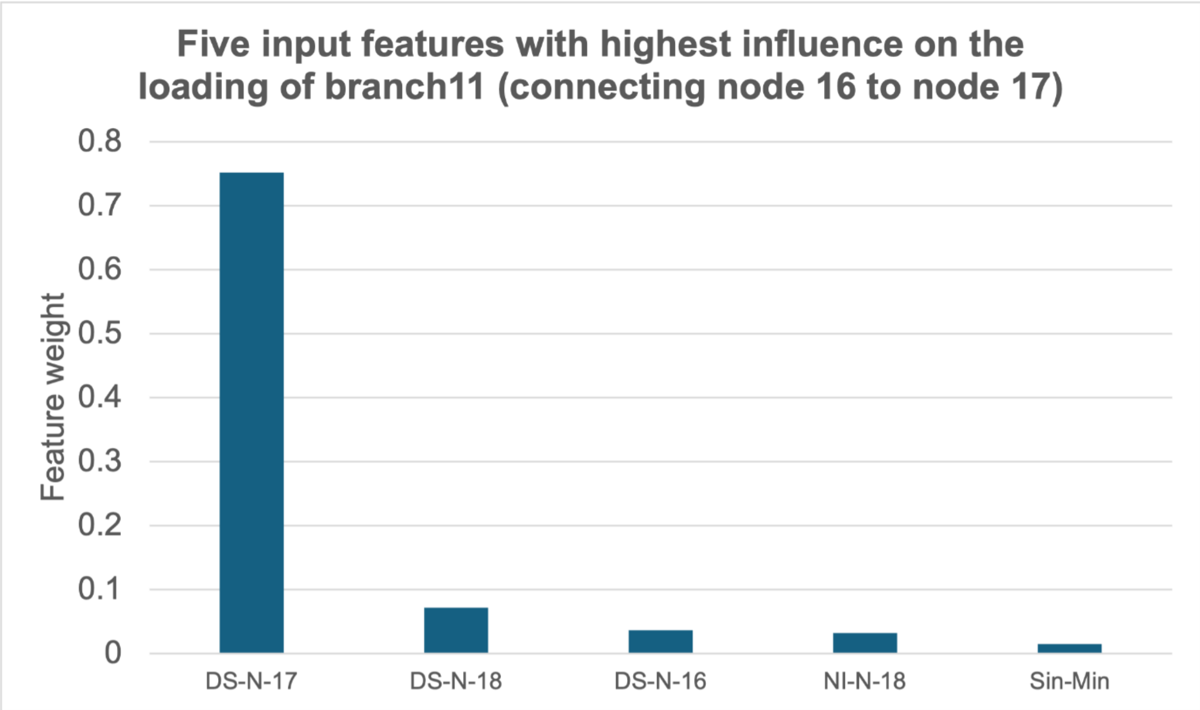

Figure 8. Five most relevant input features for the over-loading of the branch between nodes 16 and 17. DS-N-X stands for the aggregated net power injection/withdrawal downstream of node X. NI-N-X stands for the net power injection/withdrawal at node X. Min-Sin denotes a cyclical feature that was created by representing the time by means of two variables, a sine and a cosine, thus being able to express the time cyclicality (otherwise, for example, 23:00 would be interpreted by the algorithm as being very different from 00:00, while, in reality, they are equally close as, for example, 02:00 is to 03:00). The y axis indicates the relative importance of each feature.

2.3 Virtual demonstrator

In this section we describe our activities towards a virtual demonstrator. First, we present the current file-based concept of AISOP Decision Support, and the concept for federated data assimilation. Following, a use-case is presented that is designed to combine multiple modules in a closed-loop operation: grid simulations, dynamic tariffs, end-user optimization and risk forecasting.

2.3.1 Data federation based on data spaces

An approach to facilitate access to data in a secured and federated way is a data space, which the International Data Spaces Association (IDS) defines as: " ... a virtual space that provides a standardized framework for data exchange, based on common protocols and formats, as well as secure and trusted data sharing mechanisms. The IDS data space is designed to support data sovereignty, meaning that data owners retain control over their data and can determine who can use it and under what conditions." Moreover, a data space in a given domain, say transport is intended to be compatible with data spaces in other domains. Specific to the energy domain, [15] describes high level goals and use cases such as coordination of TSO-DSO for congestion management. Also, uses cases in the interface between local communities and energy utilities (e.g., DSOs, Stadwerke), such as grid the facilitation of grid connection processes and maintenance services are mentioned as having potential benefits from accessing heterogeneous data sets within an energy data space.

Minimum Viable Data Space (MVDS). As data spaces aim at being fully interoperable and standardized, along with their definitions reference implementations are provided, for example by IDS, by Eclipse Foundation, and by private companies. In order to investigate data federation in AISOP, we look into advancing the current approach outlined in Figure 9 into that described in Figure 10, where a data space and a data extraction system operated by the DSO or by a 3rd party facilitate the access to sensor data, instead of a simple file exchange, or the forwarding of data via ftp as done in the Swiss Energy Data Hub.

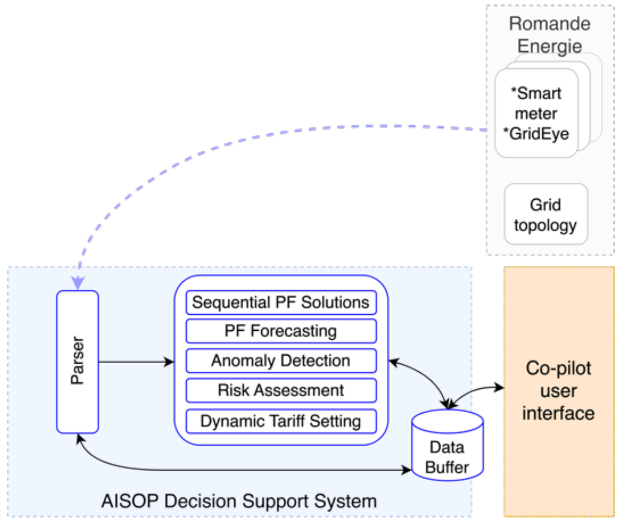

Figure 9: Current file-based implementation of AISOP software modules using data from the REEL Demo provided by Romande Energie.

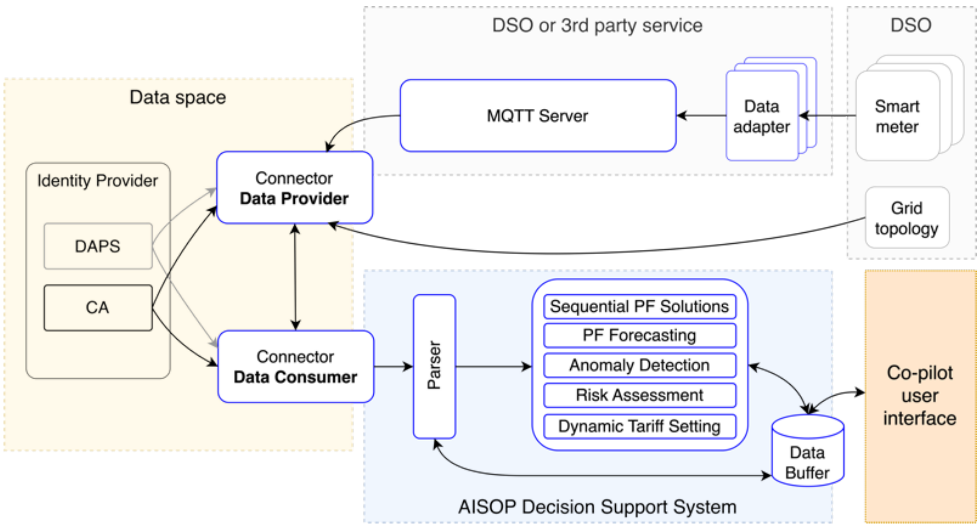

Figure 10: Concept for data federation in AISOP illustrating the extraction of data from sensors from DSOs and a MVDS consisting of identity management and data connectors.

The main difference between the current file-based implementation is that the Parser module is responsible of accessing data and convert them to AISOP data models. Thus, this data management module keeps a data catalogue and data models, and it would establish SSoT when formalising the implementation of a DPT. Details of an architecture pattern are given in Figure 15. Whereas, when data is federated via a data space, a Data Consumer module needs to be created to pass the data to AISOP Parser module. The MVDS illustrated in Figure 15 consist of a minimum set of components with enough features to experiment the provision of identities and the creation of data connectors. We experimented with the deployment of a MVDS with open, pre-configured implementations contained in the IDS Testbed git repository that include a Certificate Authority (CA), Dynamic Attribute Provisioning Service (DAPS), Dataspace Connector(s) (DSC), and MetadataBroker (not shown in Figure 10).

The main learning from deploying the IDS Testbed is the need for more powerful ICT infrastructure as we had been operating. Namely, we use two virtual machines with 4 and 12 GB of random access memory (RAM) for the implementations in Figure 9, whereas for testing the basic data space components we had to use a separate machine with 18 GB RAM and 500 GB of free disk space.

Discussion. Although data spaces have recently gained much relevance, their standardization is on-going and there are several challenges when looking into their implementation. Some of them are related to the maturity of the technologies, the learning curve, and costs. Other, are related to the business models and data governance that needs to be in place. Moreover, data spaces as per their definition and standardisation do not consider computation itself. This brings them to some extent in conflict with SSoT, which is a fundamental characteristic of a DT, as the data space federates the access to data. Therefore, implementing SSoT in data spaces is another challenge that does not have standardised practices. Looking further into the future, AI marketplaces, where users can access data consolidation workflows, securely share data across stakeholders, and ultimately deploy analytics workflows including state-of-the-art AI models [16]. AI marketplaces are less established as data space, data virtualisation, data as a service (data products), software as a service, or (application) platform. They hinge on the value of data that can be monetized, [16] presented archetypes of business models focused on a demand response use case. Thus, AI marketplaces could have benefits over data spaces, as they aim at an ecosystem of digital products and services, where users can trade datasets, provide labelling and curation services, download or run AI models in the cloud. On the other hand, they do not necessarily promote data sovereignty. In the context of AISOP, the creation of a fully fledge data space is beyond scope, but the MVDS provided a basic understanding of the technologies behind and generated synergies with research initiatives that are focused on data spaces within the energy domain.

2.3.2 Closed-loop use-case

The concept and the interactions among the selected modules in the project are trialled and demonstrated inside the ReSIM[2] simulation tool, which originated from a former project, ReMaP, funded by the SFOE, and is described in final report in detail[3].

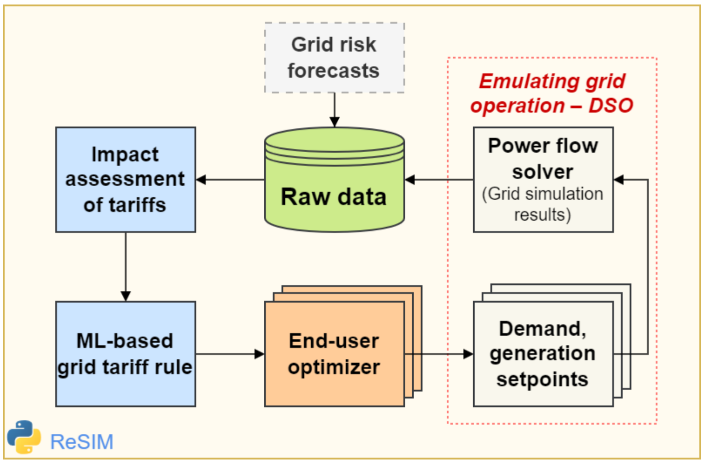

Figure 11: Visualizing the basic layout for the virtual demonstration with ReSIM

[2] fen/resim
[3] ReMaP – Renewable Management and Real-Time Control Platform, Swiss Federal Office of Energy, Final Report SI/501810-1, 2023. sfoe/remap.

The goal of the virtual demonstration is to show the feasibility of a closed-loop simulation combining a grid model and power flow data with a dynamic tariff algorithm based on power flow results and several end-users distributed throughout the grid that react to the changing tariffs. Figure 11 illustrates the setup for this virtual demonstration in ReSIM, where the individual components are modelled as follows:

- End-user optimizer:
    - The end-user models and the optimizer are provided by the HSLU team and integrated into ReSIM via a wrapper class designed by the FEN team. The optimizer is set up to minimize the household's energy cost throughout the day.
    - There are three end-user models defined in the virtual demonstration setup, each of them placed at a separate node of the selected grid. While each end-user receives a different electricity retail tariff signal, all other parameters are identical and designed to represent an electrified household:
        - The grid connection capacity of each end-user is assumed to be 15 kW which corresponds to a typical single-family house[4],
        - The base (conventional) electricity demand, excluding the EV charging and the HP, is 89.6 kWh per day with a peak of 7 kW.
        - A 10-kWp PV system along with a 5-kWh / 3-kWp battery is installed.
        - The battery-electric car is assumed to have 50-kWh battery and is assumed to arrive at home at 18:00 with a state-of-charge of 50%. The EV-charger has a maximum power of 11 kW.
        - The house is assumed to be heated with a 3-kWp (el.) heat pump connected to a 200-L heat storage tank that serves 30.2 kWh per day heat demand peaking at 4 kW.
        - **End-user 1** receives a static retail tariff of 26.36 Rp/kWh, **End-user 2** a stepped tariff based on the common high/low tariff structure with the low tariff being 18.76 Rp/kWh (before 6 AM and after 9 PM) and the high tariff 30.16 Rp/kWh (between 6 AM and 9 PM) and the dynamic tariff for **End-user 3** follows a sinusoidal form with a maximum value of 30.6 Rp/kWh and an average of 22.6 Rp/kWh.
        - The feed-in tariff for solar PV excess generation is assumed to be 10 Rp/kWh for all end-users.
    - When setting up the virtual demonstration, the type of new demand or generation (e.g., PV+BESS first, followed by EV charging and HP and thermal storage) and the amount of installed infrastructure for each end-user was increased step-by-step (in the order described above, apart from the tariffs which were always present) to check the behaviour of the end-users.

- Grid Model:
    - An open-source synthetic network from Simbench[5] is used for designing and testing. The "1-LV-urban6--0-sw" network model represents an urban low-voltage grid. The Simbench dataset provides the timeseries of load data at each node and these time-series are used whenever "end-users" introduced above are not used.

---

[4] Branchenempehlung: Werkvorschriften CH, Technische Anschlussbedingungen für den Anschluss von Verbraucher- Energieerzeugungs- und elektrischen Energiespeicheranlagen and das Niederspannungsnetz, VSE, 2021.

[5] S. Meinecke, D. Sarajlić, S. R. Drauz, A. Klettke, L.-P. Lauven, C. Rehtanz, A. Moser, and M. Braun, "Simbench—a benchmark dataset of electric power systems to compare innovative solutions based on power flow analysis," Energies, vol. 13, no. 12, p. 3290, Jun. 2020

- Power Flow Solver:
  - The commercial Adaptricity power flow solver, integrated as part of the ReMaP project was replaced by an open-source power system analysis package, pandapower [Ref.Pandapower], which is now integrated into the core ReSIM code.
- Data collection:
  - The power flow results (including transformer and line loadings, nodal voltages, net loads etc.) and the results of the end-user optimizers are collected and stored by ReSIM's internal data management, which makes the data available throughout the simulation and takes care of storing the data at the end.
- Assessment of tariff impact and network tariff rule:
  - Key performance indicators (loadings and voltages) are extracted from the power flow results and stored to create a historical trend of grid congestion which serves to both measure tariff impact and guide network tariff rules.
  - For network tariff rules, both randomized day-ahead tariffs and grid-performance based tariffs can be generated. To properly benchmark the end-user behaviour, the results in the next section are generated using the predefined retail tariffs described above.

Example: Figure 12, Figure 13, and Figure 14 serve as examples for how the end-user optimizer reacts to different tariff structures for a fully electrified residential building (PV + BESS, HP, EV) according to the parameters described above for the three end-users. If we focus on each device one at a time, we can make the following observations:

The EV, which arrives at home at 18:00 with a half empty battery, needs to charge 20 kWh before the end of the optimization horizon (e.g., in this case 1 day is selected from 0:00 to 23:45). With a static retail tariff, the EV starts charging immediately upon arrival and charges most of the required energy before 20:00. Although the PV panels are still producing power at that time, this is already being used by the conventional electricity demand which peaks around this time. Thus, the EV causes a large additional peak demand at a time, when the electricity grid is highly loaded. If a variable retail tariff (either stepped or sinusoidal) is assumed, almost the entire EV demand is shifted to the very end of the day (starting at 21:45), which is desirable from a grid perspective.

For any of the discussed tariff structures, the HP operation is scheduled for the afternoon, when plenty of PV power is available and it is thus contributing to an increased degree of self-consumption, unaffected by the tariff. When the HP operation occurs at or near the peak PV production, which can be seen for End-user 3, this is not only beneficial from the financial perspective of the end-user but also performing "PV peak-shaving" for the grid.

In case of a static retail tariff, the battery energy storage system (BESS) is purely used to increase the degree of self-consumption by charging on excess PV power and discharging after PV stops producing. It, however, does not synchronize its discharge with the highest demand peak as there is no cost for the peak load. For End-users 2 & 3, which are subjected to varying retail tariffs, the battery still stores some excess PV power for the evening. Additionally, it adjusts its charging/discharging based on the variable tariffs, charging when electricity prices are low and discharging to minimize the net load at the moments with the highest tariffs.

The combined efforts of the HP-flexibility and BESS work towards reducing PV feed-in but do not achieve a significant effect due to their limited capacity. However, the example is designed to demonstrate the architecture and the principles behind the developed modules and the approach.

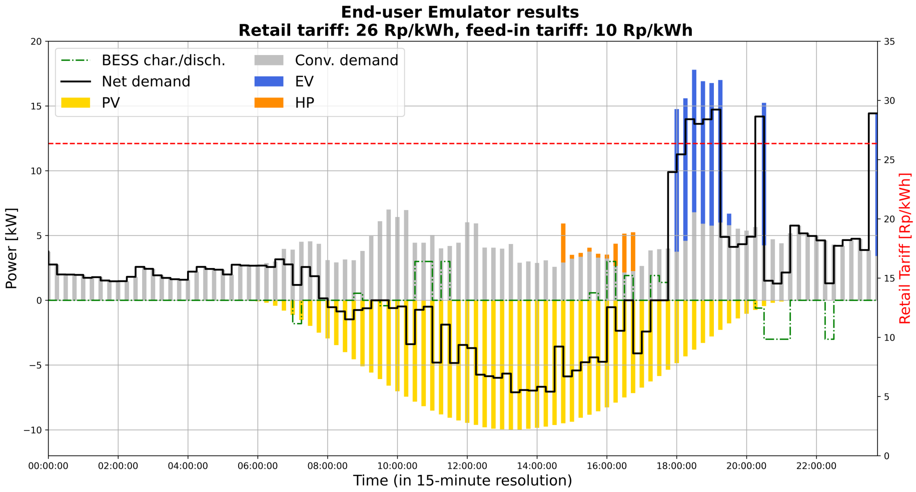

Figure 12: End-user 1; configured as a fully electrified household, optimized for a constant retail tariff of 26.36 Rp/kWh.

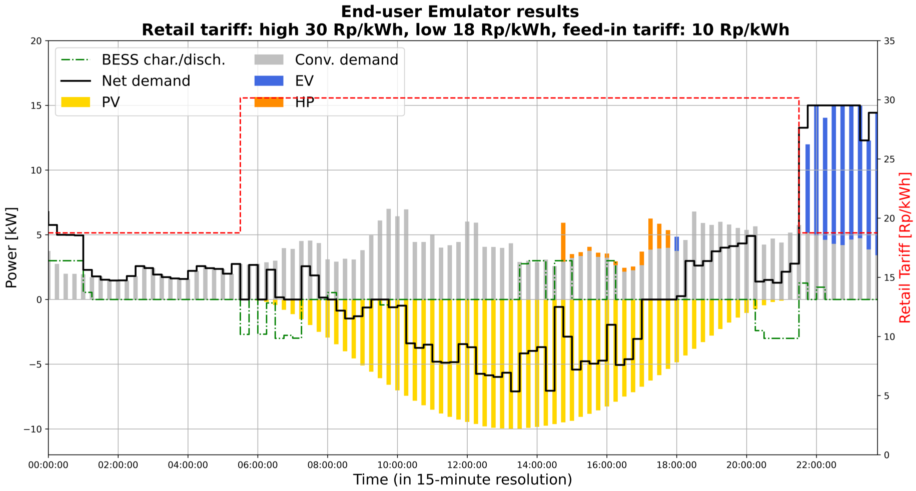

Figure 13: End-user 2; configured as a fully electrified household, optimized for a stepped retail tariff of 30.16 Rp/kWh (high, during the day) and 18.76 Rp/kWh (low, at night).

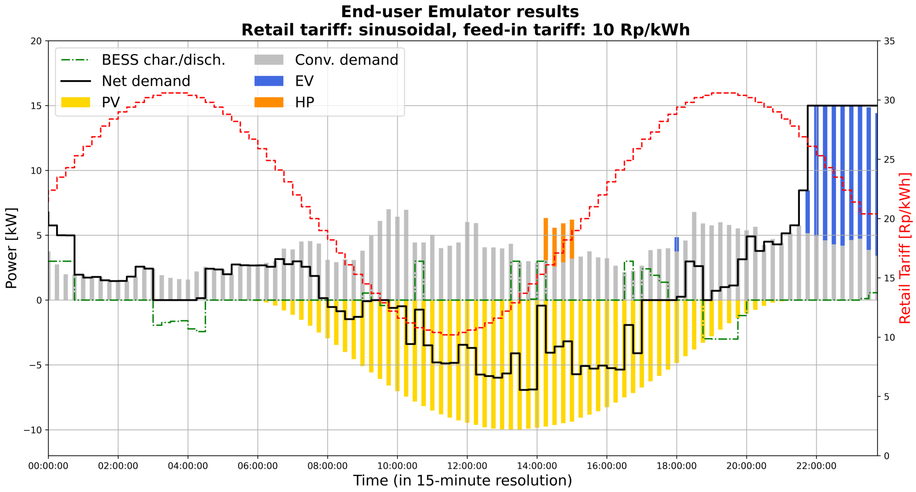

Figure 14: End-user 3; configured as a fully electrified household, optimized for a sinusoidal retail tariff with a minimum of 10.2 Rp/kWh and a maximum of 30.6 Rp/kWh.

3 Conclusions and outlook

Results on grid situational awareness for decision support, such as detection of anomalies using only grid sensor data, demonstrated the use of unsupervised ML techniques which are potentially easy to deploy for DSOs without requiring them to invest large efforts to create datasets for training and testing supervised learning algorithms. In this way, supervised approaches that were developed in the first stage of the project, on the basis of physics-based simulations, are complemented with unsupervised approaches of different level of complexity and applied to different data that are representative of various levels of grid digitalisation. Furthermore, for grid SA tools to be effective they need to be well integrated within the DSOs digital infrastructure and be suitable and informative in the context of human workers. Thus, we conceptualize and work towards a prototype of an interactive user interface, that can be used as a starting point for future decision support systems.

An approach for ML-based dynamic tariffs is introduced along with current development of software modules and the concept for data federation in a virtual demonstrator. Dynamic tariffs have the potential to incentivise changes of consumption patterns to benefit the grid, with relatively low effort from the grid operator, but methods and tools to investigate and design tariff rules need to be further developed. In this reporting period, an approach based on regression and clustering is described for ML-based dynamic tariffs and illustrated using a real grid data while the customers are represented with synthetic time-series for HPs, EV charging and solar PV generation to reflect future scenarios. In addition, as part of the virtual demonstrator development, mock dynamic tariffs are integrated to the collaboration platform, ReSIM, to demonstrate, in closed loop, how the selected end-user change their behaviour based on tariffs resulting in changes in the grid loadings. A synthetic grid was used for the demonstration purpose.

Current activities and next steps include the following.

Load Forecasting. Documentation of ensemble models for load forecast, and evaluation of a subset with multiple metrics error metrics that resemble more closely the DSO use cases such as prediction of peak loads.

Power flow. Linking forecasts and sequential power flow simulations to data from demonstration sites.

Risk. Extend situational awareness with risk estimation by using data obtained from REEL Demo sites. Risk metrics that capture time-varying operating conditions of a given asset are to be calculated to support grid operation by forecasting risks in the next day(s).

ML-based dynamic tariff. The identification of importance of input features and clustering framework will be supplemented by a rule-based algorithm to translate the knowledge extracted from the identification and clustering process to temporal and spatial tariffs. The output of the risk forecasting module will be incorporated as an input to the ML-based dynamic tariff. Hive Power will provide temperature and solar irradiation forecasts, which will also be used as input to the ML model. The relevance and importance of the risk forecasting results, and the weather forecasts will be assessed and the final list of relevant data to develop a meaningful ML-model will be determined.

Virtual demonstrator. The developed ML-based dynamic tariff will be integrated and tested with the end-user optimizer module. Part of the grid provided by Romande Energie will be used, replacing the synthetic grid used for prototyping.

4 National and international cooperation

The collaboration between the team at ZEDO e.V. (TU Dortmund) is ensured by means of virtual synchronization meetings (monthly or quarterly), and advisory board meetings where the activities in both countries are presented to the Advisory Board consisting of representatives from BKW, Amprion and Elia Grid. The deliverables in the form of conference papers, presentations and reports are shared.

The ZEDO e.V. activities focus on anomaly detection with very limited data, i.e., power flow measurements at the MV-LV transformers. The unexpected and unregistered PV generation, EV charging, new heat-pumps, meter failures, and drastic changes in consumption patterns at the customer end are classified and identified as anomalies, and the methodologies are developed to extract knowledge from the MV-LV transformer measurements to identify such instances. The methodologies are documented in conference publications [17] and [18] as well as in a report which will be published on the project website soon.

Communication channels are established with groups in two ERA-Net funded projects: Lasagne, OWGRE, and Digicities. Industry has shown interest on our approaches, the exchanged with two companies has help to gain perspective on innovative products and the gap between commercial offers and current research.

5 Publications and other communications

This section lists publications, but also participations in events, as well as upcoming activities. The research activities in the current reporting period produced published results in the form of a master thesis report, a journal publication, and multiple conference papers where project participants disseminated results and ongoing work in international events such as AMLD, CIRED, IEEE ISGT. Details of these activities and upcoming ones are shown in Table 3. Newsletters are prepared and disseminated via social media (i.e., Linkedin) and the project website (www.aisopproject.com), which is updated frequently to disseminate the activities and the project progress. Most of these documents, are accessible in the project website under Resources (www.aisopproject.com/resources/). Moreover, AISOP members participate in the ERA-Net Smart Energy Systems working groups on System Architecture and Modelling, and Regional Matters as well as in CETPartnership TRI5 & JPP ERA-Net SES Knowledge Community Meeting.

Table 3: Dissemination activities by the Swiss team, reporting period December 2023 to November 2024.

Activity	Description
Journal paper	R. Khatami, S. Nowak and Y. C. Chen, "Measurement-Based Locational Marginal Prices for Real-Time Markets in Distribution Systems," in *IEEE Trans. on Pow. Syst.*, vol. 39, no. 6, pp. 6974-6985, Nov. 2024.
Organized Event	Applied Machine Learning Days (AMLD) EPFL 2024 Track for AI for Energy Utilities • Matthias Bucher, Swissgrid AG. *Where AI could help to keep operating the transmission grid in a safe and efficient way*. (Mar. 26, 2024). • Stefanos Delikaraoglou, Axpo Group. *AI for energy trading.* (Mar. 26, 2024). • Arthur Cherubini, Romande Energie. *Data-driven generation of synthetic load curves for grid planning.* (Mar. 26, 2024). Accessed: Nov. 18, 2024. [Online Video]. Available: https://youtu.be/BqYRFtYlwPA?si=wD6Tb41lN4p17CA3 • Max Zurkinden, SwissLLM. *Enhancing LLM performance with Retrieval-Augmented-Generation.* (Mar. 26, 2024). Accessed: Nov. 18, 2024. [Online Video].Available: https://youtu.be/4EnsIOhEKcI?si=W_r1RqUA0e8xI4A7
Data	B. Barahona, Mar. 2024, “CKW Smart Meter Data,” Zenodo, doi: 10.5281/zenodo.13304499.
Report	D. Papadopoulos, “Low Voltage Load Forecasting Using Ensemble Methods,” M.Sc. thesis, School of Business, HSLU, Lucerne, 2024.
Presentation	B. Barahona et al., “A data co-pilot for electric distribution utilities to support grid situational awareness”, AMLD EPFL 2024, April 2024.
Conference paper	B. Barahona et al., “A framework for data-driven decision support for operational planning in active distribution networks,” in the Proc. of CIRED 2024 Vienna Workshop, June 2024.
Presentation	C.Y. Evrenosoglu, AISOP Project in ERA-NET Energy Systems Peer-to-Peer Feedback Sessions 2024, SESSION 2 - AI and ML for Energy Systems, June 2024.
Presentation	B. Barahona, et al., AISOP Project in CET Partnership TRI5 & JPP ERA-Net SES Knowledge Community Meeting, November 2024.

CIRED 2025	Abstracts submitted: • Risk metrics for guiding decisions in operational planning of active distribution networks • Evaluation of ensemble methods for low-voltage load forecasting using multiple metrics upon acceptance participation takes place in Lausanne, June 2025.

6 References

[1] P. van den Oosterkamp, P. Koutstaal, A. van der Welle, J. de Joode, J. Lenstra, K. van Hussen and R. Haffner, “The role of DSOs in a Smart Grid environment”, Ecorys, Amsterdam/Rotterdam, 2014.

[2] C. Basu, M. Padmanaban, S. Guillon, L. Cauchon, M. De Montigny and I. Kamwa, "Situational awareness for the electrical power grid," in *IBM Journal of Research and Development*, vol. 60, no. 1, pp. 10:1-10:11, Jan.-Feb. 2016.

[3] “Modern Distribution Grid”, U.S. Department of Energy’s (DOE) Office of Electricity (OE), 2020.

[4] G. de Almeida Terça, A. Delnooz, A. Sanjab, K. Kessels and M.U. Hashmi, “Market Enabling Interface to Unlock Flexibility Solutions for Cost-Effective Management of Smarter Distribution Grids,” EUniversal UMEI, H2020 – LC-ES-1-2019, 2022.

[5] S. Karagiannopoulos, L. Roald, P. Aristidou and G. Hug, "Operational planning of active distribution grids under uncertainty." In *IREP 2017, X Bulk Power Systems Dynamics and Control Symposium*, 2017.

[6] X. He, L. Hancher, I. Azevedo, N. Keyaerts, L. Meeus and J.-M. Glachant, “*Shift, not drift: towards active demand response and beyond*,” THINK, 2013.

[7] L. Thurner et al., "Pandapower—An Open-Source Python Tool for Convenient Modeling, Analysis, and Optimization of Electric Power Systems," in IEEE Transactions on Power Systems, vol. 33, no. 6, pp. 6510-6521, Nov. 2018.

[8] B. Barahona et al., "A framework for data-driven decision support for operational planning in active distribution networks," in *IET Conf. Proc. CP876*, 2024, pp. 954-959, doi: 10.1049/icp.2024.1979

[9] J. P. Achara et al., "SCCER-FURIES Romande Energie ELectric network in local balance Demonstrator (REEL Demo)," Swiss Federal Office of Energy SFOE, Energy Research and Cleantech Division, Final Report, Oct. 14, 2022.

[10] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," *J.Mach. Learn.Res.*, vol. 12, pp. 2825-2830, 2011.

[11] R. D. Zimmerman, C. E. Murillo-Sánchez and R. J. Thomas, "MATPOWER: Steady-State Operations, Planning, and Analysis Tools for Power Systems Research and Education," in *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 12-19, Feb. 2011, doi: 10.1109/TPWRS.2010.2051168.

[12] A. M. Prostejovsky *et al.*, “The future role of human operators in Highly Automated Electric Power Systems,” *Electric Power Systems Research*, vol. 175, p. 105883, Oct. 2019. doi:10.1016/j.epsr.2019.105883

[13] A. Marot, A. Rozier, M. Dussartre, L. Crochepierre, and B. Donnot, “Towards an AI assistant for Power Grid Operators,” *Frontiers in Artificial Intelligence and Applications*, Sep. 2022. doi:10.3233/faia220191

[14] VDE Association for Electrical, Electronic & Information Technologies: The Digital Twin in the Network and Electricity Industry, VDE Study, Offenbach am Main, Germany, May 2023

[15] M. Gouriet, H. Barancourt, M. Boust, P. Calvez, M. Laskowski, A.S. Taillandier, L. Tilman, M. Uslar, and O. Warweg, "The Energy Data Space: The Path to a European Approach for Energy," in *Designing Data Spaces: The Ecosystem Approach to Competitive Advantage*, Cham: Springer International Publishing, 2022, pp. 535-575.

[16] B. Barahona, B. Bowler, C. Gounden, and A. Papaemmanouil, “Business Models of an AI Marketplace for Energy Systems with Focus on Demand Response,” in *Smart Services Summit*, 2022, doi: 10.1007/978-3-031-36698-7_8

[17] R. Balouchi Anaraki, R. Palaniappan, U. Häger, C. Rehtanz, “Data-driven Approaches for Anomaly Detection in Low-Voltage Grid Net Power,” in Proc. IEEE PES ISGT, Dubrovnik, Croatia, 2024.

[18] R. Balouchi Anaraki, R. Palaniappan, U. Häger, C. Rehtanz, “Anomaly Detection in Low-Voltage Grids with LSTM Autoencoders: A Study on Future Scenario Impacts,” in *Proc. IEEE PES ISGT*, Dubrovnik, Croatia, 2024.

[19] International electrotechnical commission. “IEC 61850 Standard Series.” Accessed: Nov. 15, 2024. [Online.] Available: https://iec61850.dvl.iec.ch.

[20] ENTSO-E. “Common Grid Model Exchange Standard (CGMES) Library.” Accessed: Nov. 15, 2024. [Online.] Available: https://www.entsoe.eu/data/cim/cim-for-grid-models-exchange.

[21] Modellica Association. “Functional Mock-up Interface.” Accessed: Nov. 15, 2024. [Online.] Available: https://fmi-standard.org/.

[22] Open Knowledge. “Data Package.” Accessed: Nov. 15, 2024. [Online.] Available: https://datapackage.org.

7 Appendix: Architecture of SA tool

Part of the recent activities consisted in defining specific data management approaches and defining the organization of the modules and workflows. The following sections describe the current results of these activities that shape the architecture of the SA decision support tools describe in Section 2.1.

Data management. The following aspects were investigated with the objective to define specific implementation paths:

- data models and standards to facilitate data interoperability and ingestion, and
- data spaces solutions or ETL tools

Clearly, multiple possibilities arise, and no single combination of technologies rules every other option. In our case, we are creating software modules (i.e., digital twin) and then combining them in workflows (i.e., digital process twin) to prototype an application (i.e., TRL 6) for specific use cases such as anomaly detection. We resort to standards and popular opensource tools that are well maintained and in active development or well stablished. For data models, standards such as IEC 61850 [19], CIM [20], or FMI [21] are often referred to in data models of smart grid components and grid simulation models respectively. However, at this stage in the development we find the minimal datapackage [22] specification is suitable given that we are developing with files as the data source, its simplicity, and its extensibility. Therefore, we selected it as the basis model for describing data in a catalogue; for extracting these data and then validating it we resort to the Frictionless Framework [22].

The second point, data spaces or ETL tools, refers to how the data is to be access. In Section 2.3, we refer to data spaces as a solution for federated data access, here we document the more established extract load transform (ETL) tools, such as Airbyte or dbt and workflow orchestrators such as Airflow, dbt, Prefect, or argo workflows. The later can also be applied to the orchestration of computational workflows which are at the core of our SA tool. At this stage, we see that a combination of Airbyte for data ingestion, dbt for transformation and quality, and Prefect for orchestration could be suitable for our prototype. Ultimately, with this toolchain, we aim at Single Source of Truth (SSoT) that procures data consistency and reliability through the various layers, the implementation concept is illustrated in Figure 15.

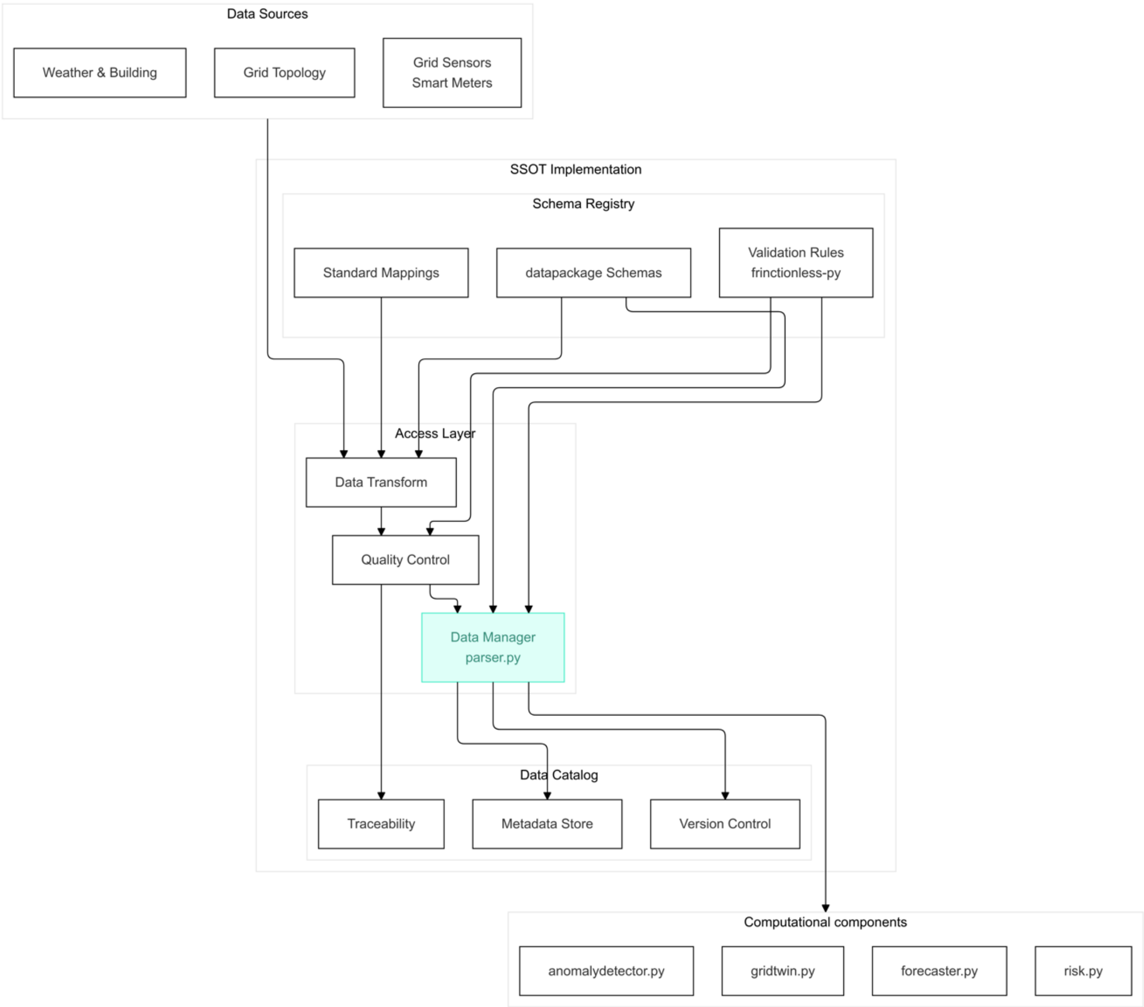

Figure 15: Illustration of Single Source of Truth (SSoT) implementation.

Note that SSoT is a process that, although may be highly automated, like other ICT processes it needs monitoring and maintenance. Moreover, although the motivation to implement SSoT is clear (i.e., big data, heterogeneous data sources, interoperability), and SSoT is considered a fundamental process to maintain attribute values of DTs, there is no encompassing implementation. Comprehensive definitions of DTs that emphasise SSoT, such as [14], provide no information about the actual practical implementation patterns and tools.

Architecture. Beyond the virtual demonstrator diagrams shown in Figure 10 and Figure 11, and the implementation of SSoT shown in Figure 15, a software architecture is needed to guide the implementation by linking desired functionality to specific tools and software modules. Figure 16 shows the main components of the proposed architecture: Data Sources, ETL Layer, Orchestration, Core Components and the Interface to the user. The core parts in the scope of AISOP project are the orchestration of computation workflows and the software modules themselves which implement load forecasting, power flow forecasting, sequential power flow simulations, anomaly detection and risk assessment. For these we selected the main software libraries and started implementations (i.e., *.py scripts). Regarding data sources, a comprehensive data catalogue was created, and a few data sources are in use corresponding to grid topology, grid sensors, smart meters, weather and building data.

However, integration into a SSoT implementation is part of future work. Note that this also reflects the fact that in practice there always needs to be some data integration step before being able to apply new modelling and forecasting tools in an operational set up.

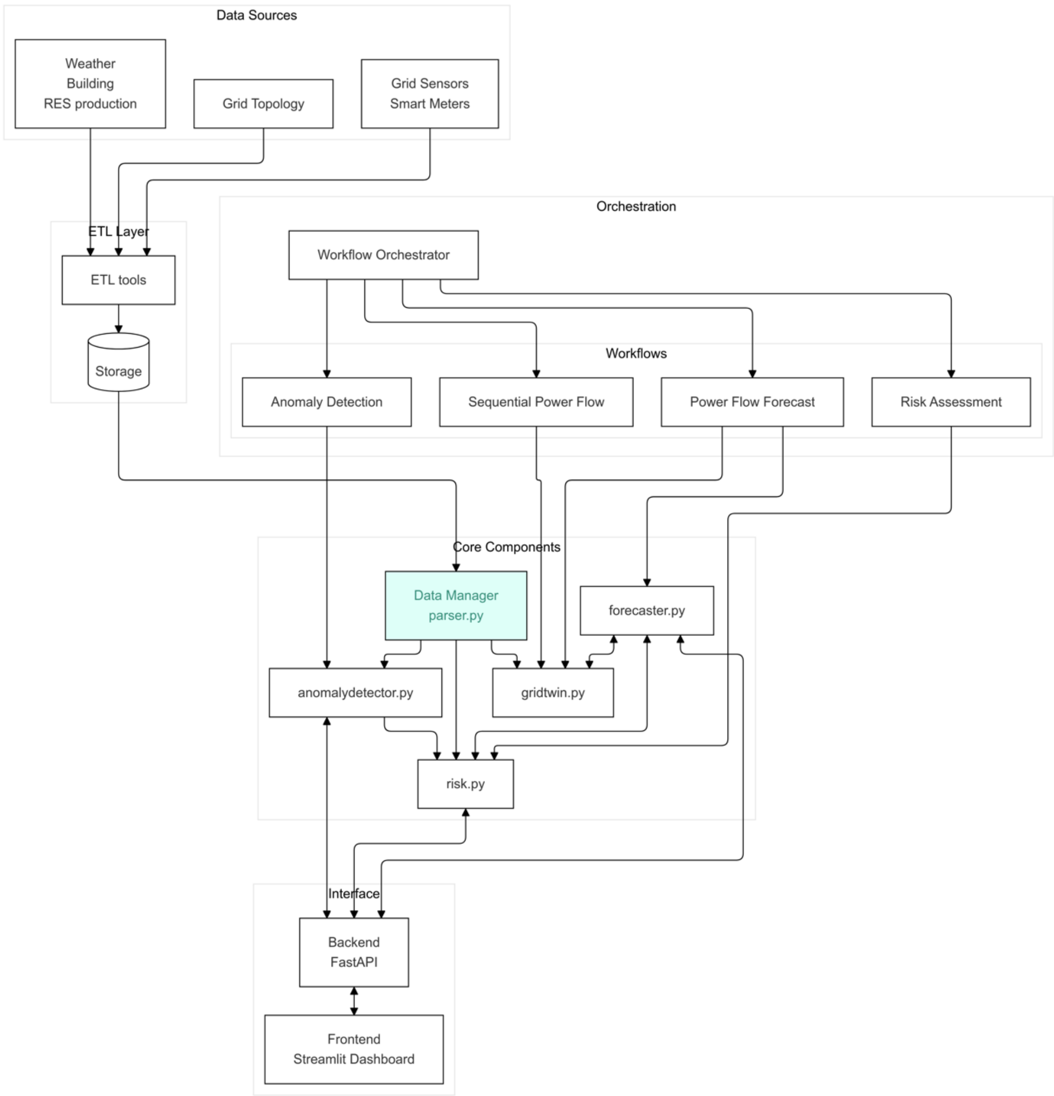

Figure 16: SA tool implementation illustrating the data assimilation, workflow orchestration, core components and user interface.

Moving on to the ETL Layer, this can be implemented as a data space (Section 2.3) or via ETL tools. Finally, the interface to the application user is shown, which can be implemented as a Web application as it was illustrated in the co-pilot example.

Note that with this architecture we aim at clarity and viability of a DPT in the specific context of AISOP. Unlike architecture concepts and definitions such as those based on SGAM [14] that serve as a vision that aims to comprehensive interoperability integrating current and future commercial solutions.