MODERATE: Marketable Open Data Solutions for Building Energy Optimisation

JASPER VERMAUT, EU Policy & Project Officer, REHVA
FRANCESCA CONSELVAN, Data Scientist Researcher, e-think Energy Research
DANIELE ANTONUCCI, Senior Researcher, EURAC Research
PHILIPP MASCHERBAUER, Researcher Associate, Vienna University of Technology
CRISTIAN POZZA, MODERATE Coordinator & Senior Researcher, EURAC Research

MODERATE is a Horizon Europe funded project that started in June 2022. Its aim is to develop a marketplace platform that improves availability and interoperability between datasets for the building industry, leveraging open data and open-source solutions. The objective is to promote data exchange between different producers and consumers while complying with legal and ethical constraints. The project innovates building data collection, synthesis, and services for management, optimization, and decision-making.

MODERATE aims to create a fully open platform that offers open data and data-driven services. Large and diverse datasets will be available on buildings of different types, such as residential, commercial, and offices. Different data-driven services and software pipelines will also be accessible as open-source solutions, allowing targeted users to test solutions before investing in tailored development. Ultimately, the availability of data and services will bridge the gap between confidential data, low-quality, or unavailable data and informed decision-making, with the final goal of reducing carbon emissions and mitigating climate change in the building sector. This article focuses on two components of the MODERATE platform that contribute to the enhanced use of building data:

1. Improved data collection and exchange through data synthetisation: The platform uses data synthetisation techniques to ensure that sensitive data is anonymised, but most of the data's value is retained for analysing building performance. Data synthetisation makes it more secure for data owners to share their data and complies with the General Data Protection Regulation (GDPR). This way, the MODERATE platform can combine multiple datasets within the building sector and obtain aggregated results.
2. Analysing the data through data-driven services: The platform will include ten data-driven services that transform raw data into meaningful knowledge. These services cover different aspects of the building industry and work with static, dynamic or both types of data.

The platform allows data owners to retain control of their own data while enabling its use in a wider market for experts.

Data synthetisation techniques for secure data exchange

The increasing use of building monitoring and control systems has created an opportunity for data-driven approaches in the construction sector. However, security and privacy concerns retain companies to share their data. Synthetic data can be used as an alternative, since it masks any sensitive information, like household or person's identity, and maintains the original statistical properties. This ensures that no confidentiality or privacy issues are breached, while the quality of data remains very high. Additionally, synthetic data allows for high scalability of data and different data sources can be merged to generate an enriched dataset to analyse data at a more aggregated level.

Synthetic data is increasingly being used in the energy sector for various applications, like a prediction of energy demand [**1**], fault detection in energy systems [**2**] and building stocks [**3**].

Synthetic data is artificially generated data which mimics the real data by using the underlying statistical properties and at the same time masks the most sensitive properties of this data. After the synthetisation process, it's impossible to assign the data a specific person or household entity. This ensures that no confidentiality or privacy issues are breached while the quality of data remains very high since it keeps the most valuable properties of the real data. This process allows for a high scalability of the data, where different data sources can be merged to generate an enriched dataset to analyse data at a more aggregated level.

The synthetised data within the de MODERATE platform will include data coming from EPCs across Europe, electricity load profiles from smart meters, heating profiles and data coming from temperature sensors.

Synthetic data is artificially generated using different statistical and ML techniques, like hidden Markov Chain [**4**], autoregressive models [**5**] and other ML algorithms. The advent of Generative Adversial Network (GAN) has revolutionized this field thanks to its ability to generate realistic and high-quality data and its flexibility to create different variants of GANs using the fundamental piece of work [**6**]. GAN is a type of deep learning algorithm

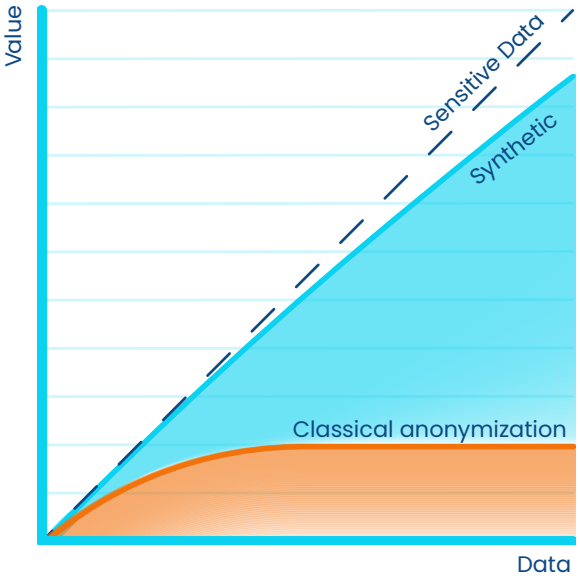

Figure 1. Data Synthetisation anonymises the most sensitive data while keeping most of the data's value through machine-learning techniques, making it much more useful than classical anonymisation.

that uses two neural networks to generate new data. The two networks, the generator and the discriminator, compete with each other in a game of cat and mouse. The generator creates new data, while the discriminator determines whether the samples are real or fake (**Figure 2**). The ultimate goal is to create synthetic data that is indistinguishable from the real.

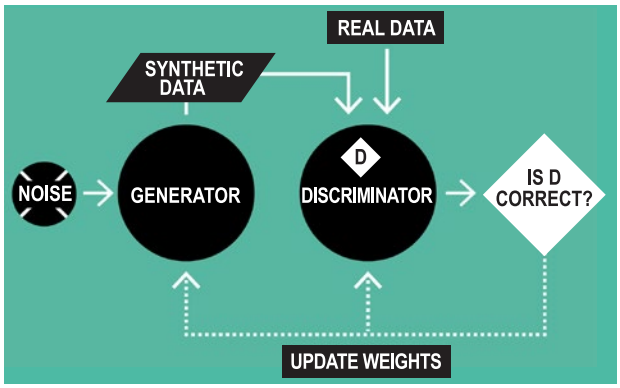

Figure 2. GAN's workflow (picture by SquareUp).

Turning raw data into meaningful knowledge

In addition to improved data collection and harmonization, the MODERATE platform will offer 10 data-driven services that enable users to analyse the performance of buildings and gain meaningful insights from the raw (synthesized) data available on the platform.

As can be seen on **Figure 3**, the MODERATE platform will offer three distinct categories of data-driven services for monitoring building performance:

System management services: These services can be used for monitoring and controlling the building performance through data analysis, providing real-time insights into energy consumption and potential inefficiencies through fault detection. The services included in this category are:

- *Fault detection & forecasting*: This service ensures a reliable monitoring and control system that can be used by utility companies, ESCOs, and facility managers to identify faults in real-time and enable predictive maintenance of the buildings systems they manage. Through a user-friendly web application users will be able to predict and understand possible failures in their systems.
- *Energy system optimization*: This service uses machine learning techniques to generate possible optimizations in building systems based on the available data available in the platform. This will facilitate the work

of energy and facility managers, ensuring that the potential of the building defined in the design phase is respected and identifying the right set-up that guarantees the best level of comfort at the lowest cost.

- *Energy Conservation Measure (ECM)*: The ECM application assesses potential energy savings through requalification measures like improved insulation of building facade or upgraded HVAC systems. It calculates and rank. The possible ECMs identifying the best efficiency intervention, leading to economic savings and CO_2 reduction. The calculation of each ECM is performed using a dynamic building model created using the ISO 52016 standard. The required inputs can be provided manually or through a direct connection to a BIM model. The ECM is a measure of energy efficiency that compares the additional costs of energy efficiency measures to maintenance and building stock valorisation costs.

Building optimization and assessment: These services use AI models to evaluate building performance and identify areas for improvement in terms of solar roof potential, cost savings, and overall building performance. The services included in this category are:

- *Solar cadastre*: it is a tool that allows you to visualize and assess the photovoltaic potential in different areas by looking at the roof surface and irradiance in that area. The tool easily calculates the energy savings and the payback time of a solar PV investment in a certain area.
- *Local energy communities assessment*: This service is used in tandem with the solar cadastre and is more at municipality level. The tool can identify the feasibility to develop an energy community based on the availability of rooftop PV in a specific area. It compares the energy demand with the potential of solar PV and can help identify the optimal areas to setup an energy community.
- *Measurement & Verification for building assessment*: The M&V for Building Assessment service uses a standard procedure (IPMVP) to verify actual energy savings achieved through improvement actions. It helps ESCOs to define accurate Energy Performance Contracts by identifying savings and providing guidelines to choose the best option.
- *Benchmarking tool*: The benchmarking tool compares building performance for energy and comfort, facilitating comparison and energy accounting with similar buildings to assess improvement opportunities and verify energy savings. It benefits real estate, municipalities, policy makers, building managers, and planners. The tool will be an open-source web app, using synthetic data generated as a reference for specific analyses.

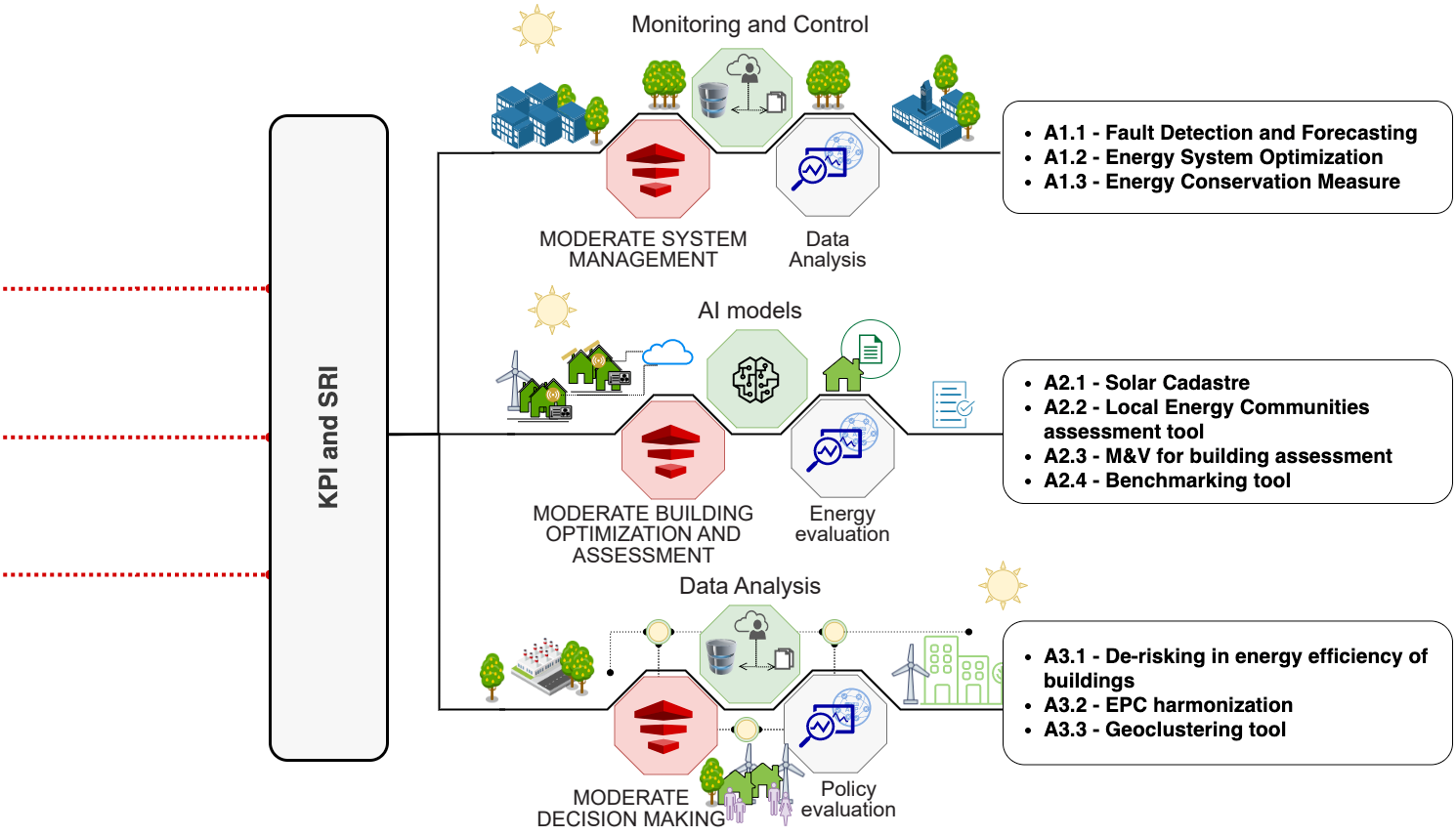

Figure 3. Overview of the MODERATE data-driven analytical services.

Decision-making services: This category is focused on analysing data to support policy evaluation and decision-making, helping building managers make informed choices about resource allocation and energy management.

- *De-risking investments in building energy*: National policies are driving investment in energy-efficient building renovations for reduced consumption and improved indoor comfort. This service will use synthetic data to guide investors on what are the most financially attractive investments to improve the performance of their buildings, impacting policy by enabling better decision-making for energy-efficient building investments.
- *Energy Performance Certificate Harmonization*: The EPC harmonization service aims to standardize EPCs across regions and countries based on a European ontology. It will associate energy consumption data to help define policies and promotions for reducing thermal consumption and CO_2 emissions.
- *Geo-clustering tool*: The geo-clustering tool evaluates building stock performance using cluster techniques. It compares the performance of buildings in the same or different areas to assess possible savings in refurbishment. Building designers can use it to determine the best technology for a specific use and climate, while researchers can analyse the impact of building features on energy consumption and thermal comfort. ■

▶ For more information, see the project website: https://moderate-project.eu/

▶ Follow us on LinkedIn & Twitter: https://www.linkedin.com/showcase/moderate-he/ & https://twitter.com/MODERATE_HE

References

[1] Bilgi Yilmaz and Ralf Korn, 'Synthetic Demand Data Generation for Individual Electricity Consumers: Generative Adversarial Networks (GANs)', *Energy and AI* 9 (August 2022): 100161, https://doi.org/10.1016/j.egyai.2022.100161.

[2] Neil H. W. Eklund, 'Using Synthetic Data to Train an Accurate Real-World Fault Detection System', in *The Proceedings of the Multiconference on 'Computational Engineering in Systems Applications'* (Multiconference on "Computational Engineering in Systems Applications, Beijing, China: IEEE, 2006), 483–88, https://doi.org/10.1109/CESA.2006.4281700.

[3] Zhe Wang and Tianzhen Hong, 'Generating Realistic Building Electrical Load Profiles through the Generative Adversarial Network (GAN)', Energy and Buildings 224 (October 2020): 110299, https://doi.org/10.1016/j.enbuild.2020.110299.
Shashank Asre and Adnan Anwar, 'Synthetic Energy Data Generation Using Time Variant Generative Adversarial Network', Electronics 11, no. 3 (24 January 2022): 355, https://doi.org/10.3390/electronics11030355.

[4] Ziegler, Felix et al., 'A Probabilistic Modelling Approach for Residential Load Profiles', 27 February 2020, https://doi.org/10.5281/ZENODO.3689339.

[5] Rafał Czapaj, Jacek Kamiński, and Maciej Sołtysik, 'A Review of Auto-Regressive Methods Applications to Short-Term Demand Forecasting in Power Systems', *Energies* 15, no. 18 (14 September 2022): 6729, https://doi.org/10.3390/en15186729.

[6] Ian Goodfellow et al., 'Generative Adversarial Nets', in *Advances in Neural Information Processing Systems*, ed. Z. Ghahramani et al., vol. 27 (Curran Associates, Inc., 2014), https://proceedings.neurips.cc/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf.
Ankan Dash, Junyi Ye, and Guiling Wang, 'A Review of Generative Adversarial Networks (GANs) and Its Applications in a Wide Variety of Disciplines -- From Medical to Remote Sensing' (arXiv, 1 October 2021), http://arxiv.org/abs/2110.01442.

MODERATE

Funded by the European Union

This article has been prepared under the MODERATE project which has been funded by the European Union's Horizon Europe innovation program under Grant Agreement No. 101069834. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting authority can be held responsible for them.