

## Ofgem Consultation: Governance of data-sharing infrastructure

### Centre for Net Zero response

Sent by email to: [digitalisation@ofgem.gov.uk](mailto:digitalisation@ofgem.gov.uk)

Centre for Net Zero (CNZ) is part of the world-leading group of organisations that comprise Octopus Energy Group. We are an autonomous, not-for-profit organisation that delivers pioneering research to transform future energy systems, in particular by leveraging the Octopus Energy customer dataset.

A digitalised and data-driven energy sector is of central importance to our mission. As this consultation focuses on governance, we have limited this to a short overall response aligned with our research, where we think we can add most value. We are happy to discuss and share any further information directly.

We agree with the rationale for Ofgem's work in this area. Data is a critical enabler for a smart, digital and low-carbon energy system with flexibility at its centre. There will be growing opportunities to maximise the use of energy data across all services and markets. In particular, as energy systems electrify and rely on renewable generation from more varied sources, it is increasingly important to accurately profile and actively manage demand. Good quality demand data is therefore highly valuable to researchers, grid operators and innovators - and, ultimately, net zero.

We also agree with Ofgem in highlighting that certain sensitive data, such as data which can be used to identify an individual or poses a risk to grid infrastructure, cannot be shared openly. This, of course, applies to personal consumption data. Given the diversity in consumers, buildings, technologies, and grid conditions, to be useful this demand data often needs to be sufficiently granular, i.e. household-level, with at least half-hourly resolution and combined with useful metadata such as low-carbon technology ownership. Real smart meter data cannot be released at this level of granularity without unacceptable privacy risks.

While we agree with the overall direction of travel towards a decentralised data-sharing model, compared to a central repository, we think the data-sharing solution will entirely depend on the use case. We urge Ofgem to consider specific use cases for the vastly different types of energy data used across the system, and suitable solutions for each of these. The best solution may not be the sharing of real data in every case, especially when privacy or cyber security is at risk.

When it comes to smart meter data, AI-generated synthetic data can provide realistic profiles for consumer archetypes that cannot be attributed to individuals. CNZ has begun to generate and share this data ourselves - see the annex for further detail. This approach avoids privacy concerns of releasing raw data, and mitigates delivery risks around developing lengthy or complex data-sharing agreements. Even accepting the treatment of aggregated smart meter data as presumed open - as set out in Ofgem's Data Best Practice Guidance - GDPR principles, security controls and technical barriers are likely to continue to limit the smart meter data that suppliers

and networks can share. Synthetic data has the potential to democratise smart meter data access more easily, safely and quickly.

While some uses of smart meter data are inherently specific to an individual consumer, such as a household understanding its own consumption, for many use cases we simply need realistic profiles for consumer archetypes. We have not yet explored the full potential of synthetic data, but our research already shows immense value for two broad purposes:

- a) Grid modelling and demand forecasting with accurate and granular smart meter data - from improved energy system modelling, to strategic network planning, to sizing grid connections for new housing developments.
- b) Innovation to develop new products, services and business models using smart meter data - for example, new smart tariffs or green finance products.

More broadly than synthetic smart meter data, advances in generative AI have the potential to transform the use of data across the system. This includes changing the way datasets are processed, which may mean existing protocols to standardise data format become outdated. Given the growing role of AI, data-sharing arrangements should give it specific focus, ensuring that developing infrastructure is both resilient to developments and embraces its transformative potential.

In summary, the implications for governance of data-sharing infrastructure are that:

1. the interim DSI coordinator should consider that secure exchange of real data is not the only solution for each use case. Specifically, it should have oversight of common standards for synthetic data to ensure it is used for the public good and meets standards for protecting the raw training data;
2. governance arrangements must be as agile as possible, and resilient to changing and emerging technologies and data-sharing practices. The expertise that the interim DSI coordinator draws on should include those with capabilities in generative AI as well as digital tools.

Beyond governance, CNZ believes there is a more strident role the public sector can play in managing and the use of synthetic data in the future energy system - including using it in grid modelling, overseeing its regulation, and enabling its use for innovation and research.

## Annex: Existing work to generate and scale synthetic smart meter data

CNZ has developed a generative AI model, [Faraday](#), trained on the Octopus Energy customer dataset to produce synthetic smart meter data. Based on Conditional Variational Auto-encoder and Gaussian Mixture Model, it generates daily load profiles consisting of half-hourly kWh consumption for a given set of user-specified inputs, which currently include low carbon technology, property type, and season.

In collaboration with Linux Foundation Energy, CNZ has also established an international data community, [Open Synth](#), to facilitate the sharing of synthetic smart meter data and algorithms so it can scale quickly, particularly in areas with limited data access. This is an open-source one-stop-shop for synthetic smart meter data, including:

- a) Model repository to host open-source algorithms for generating synthetic data
- b) Data repository to host synthetic data sets contributed by data owners (trained on their own raw proprietary data)
- c) A community for quality assurance, in particular to agree a common evaluation framework against which all data is vetted to ensure privacy is protected.

To date, Centre for Net Zero has:

- Released Faraday to alpha testers for research (e.g. [TEED Digitisation Project](#) by University of Birmingham, Better Home Leeds Project by ARUP, other academic projects by University of Manchester and King's College London)
- Published a short [paper of our own Faraday model](#) ([presented](#) at the “Tackling Climate Change with Machine Learning” Workshop at the 12th International Conference on Learning Representation 2024).
- Published ["Defining 'Good': Evaluation Framework for Synthetic Smart Meter Data](#) (jointly with leading academics from Georgia Tech, MIT, and Oxford University) proposing a common evaluation framework to benchmark algorithms for synthetic smart meter data based on three tests: fidelity, utility and privacy.
- Set up the [OpenSynth Model Repository](#) and released our own source code for Faraday's algorithm, with a basic notebook tutorial on how to use the code.

In the next 6 months CNZ plans to:

- Develop Faraday - continue to improve the model, including to retrain it on a national representative dataset; version 4 will be released shortly.
- Open up access - ‘OpenSynth’ data community will generate, share and improve synthetic data; we will start by releasing a sample of our synthetic Octopus Energy dataset to the community, which will be publicly available.
- Demonstrate real-world applications - release a paper on the potential role and specific use cases for synthetic smart meter data.