UK Energy Research Centre

Data Sharing for Energy Modelling

Workshop held on October 2 October 2024

UKERC Working Paper

Catherine Jones (UKERC, STFC), Elizabeth Newbold (DINI, STFC), Katie Yates (DINI, STFC), Teagan Zoldoske (DINI, STFC), Andy Boston (RedVector), Mike Colechin (Cultivate Ltd), Keelan Colechin (Cultivate Ltd), Joy Davidson (DCC), Clara Lines (DCC) and Agnes Jasinska (DCC)

March 2025

Acknowledgements

This work has been funded by the Department of Science, Innovation and Technology as part of the Data Infrastructure for National Infrastructures project. The Energy Data Centre is funded through the UK Energy Research Centre (EP/S029575/1)

.

Introduction to UKERC

The UK Energy Research Centre (UKERC) carries out world-class, interdisciplinary research into sustainable future energy systems.

It is a focal point of UK energy research and a gateway between the UK and the international energy research communities.

Our whole systems research informs UK policy development and research strategy.

UKERC is funded by the UK Research and Innovation.

Introduction to DINI

The Data Infrastructure for National Infrastructure (DINI) project explores the challenges and opportunities in data sharing within the domain of national infrastructure systems research with a focus on energy, transport and water.

DINI has been funded by Department of Science, Innovation and Technology.

.

Contents

1. Executive Summary

This report presents an analysis of the barriers to data sharing identified by participants of the data sharing for energy modelling workshop, using the DINI barriers and FAIR (Findable, Accessible, Interoperable and Reusable) approach to make a set of recommendations to address those barriers.

Energy research is multi-disciplinary, so there are different expectations, domain standards and cross-over terminology which makes the sharing of data a challenge, especially for models which focus on whole systems analysis. The participants highlighted the importance of community building and knowledge sharing.

The energy sector, both research and operational, is a heavy user of energy models and there is overlap in input data requirements for these models which is reflected in some of the commercial and legal barriers identified.

A particular feature for modelling in this domain is the consideration of the timescales being modelled. It is common for models to address the next 30-50 years. Physical power assets have a decades long lifespan, but new types of assets are coming on stream while old things disappear. So, models need to consider how to deal with both old assets and incorporating new technologies which means that resources need to be invested to keep models active and relevant.

The analysis grouped the workshop messages into the FAIR framing which showed

- **Findable issues**: discovering models for use and data to use in them (*Cultural* and *Technical* DINI barriers)
- **Accessible issues:** rights and licensing (*Commercial, Legal & Privacy* and *Security* DINI barriers)
- **Interoperable issues:** terminology (*Technical* DINI barriers)
- **Reusable issues**: Data quality, documentation, and context (*Technical* DINI barriers)

All of the DINI top level themes (see Appendix II for complete list) were identified by participants during the workshop, however the majority of barriers linked to the key messages were considered to be technical and relating to the description, documentation and context of data.

While the focus of the workshop was on data used and created by models, some of the comments and barriers relate to the models themselves and the issues around software being developed by domain experts. While participants shared experiences of sharing both input and output data, there was an emphasis in the key messages on reusing other people's data rather than on preparing their data for sharing.

Recommendations

Infrastructure

1. Establish requirements for a registry for energy models.

   The first phase should be focused on models created in academia. Undertake further requirements scoping work for a model registry. This is something that could build on work done by UKERC and the Energy Data Centre (EDC) could lead this work within a future phase of DINI.

2. There is an identified need for a central platform to share data, run models and facilitate communities. Ensure that the requirements identified form part of future DINI implementation plans.

Guidance and support

3. Enhance EDC guidance on software citation to explicitly cover citation and referencing of models to support discoverability within the literature.

4. Produce and publicise guidance available from the EDC to support the creation of FAIR data to be used in modelling.

Energy modelling community

5. Create a working group to review classifications of energy data for statistics (IEA; Eurostat; DUKES) and to develop a multi-level classification of energy data that we can use to: (i) compare outputs from different models; (ii) output in international statistical formats; and, (iii) output in formats for international initiatives (e.g. for IPCC comparisons led by IAEA, IIASA). There could be a role for UKERC/EDC/DAFNI as a convener of this activity.

6. Contribute to the AI adoption debate by writing a position paper on this topic, with input from DAFNI, UKERC and other collaborators.

Researchers, research funders and research organisations

7. Creation of appropriate contextual information should be done routinely by researchers as part of good research practice, having received appropriate training and support

8. Research organisations should enable researchers to have the right skill sets or access to professionals to create long-lived research outputs

9. Research funders should consider how to support models created through research funding to gain the maximum benefit from the models over the longer-term.

10. Research organisations who train, support and employ researchers need to ensure that credit for data management and sharing is reflected in the professional development programme and reward structure.

2. Introduction

This workshop was a follow-on from a workshop held by the Centre for Research into Energy Demand Solutions (CREDS) and the UK Energy Research Centre (UKERC) in October 2023 [1], [2] which considered barriers to data sharing focused on energy consortia specifically. One of the recommendations was:

"*The energy community is a large producer and user of models in a wide variety of areas and common standards for what to archive to enable FAIR data and reproducibility have not yet been agreed. Such protocols would be helpful to discuss. The energy research-specific issues for sharing the outputs of energy models should continue to be highlighted*."

The 2024 workshop progressed this recommendation by bringing together key stakeholders who create, use or support energy modelling to develop a better understanding of the key issues to improve data sharing.

It was organised by UKERC's Energy Data Centre (EDC) and the DINI project, held on 2nd October 2024. The organising committee consisted of Catherine Jones (UKERC, STFC), Katie Yates (DINI, STFC), Elizabeth Newbold (DINI, STFC), Andy Boston (RedVector) and Mike Colechin (Cultivate Innovation Ltd). We were supported on the day by Joy Davidson, Agnes Jasinska, Clara Lines and Linda Freimane (Digital Curation Centre), Teagan Zoldoske (DINI, STFC) and Keelan Colechin (Cultivate Innovation Ltd). All have contributed to this report.

Contributors and Contributions:

Author	Role
Andy Boston	Investigation (equal), Methodology (equal), Project administration (equal), Writing – review & editing (equal)
Keelan Colechin	Investigation (equal), Writing – original draft (equal)
Mike Colechin	Investigation (equal), Methodology (equal), Project administration (equal), Writing – review & editing (equal)
Joy Davidson	Investigation (equal), Writing – review & editing (equal)
Clara Lines	Investigation (equal), Writing – review & editing (equal)
Linda Freimane	Project administration (equal)
Agnes Jasinska	Investigation (equal), Writing – review & editing (equal)
Catherine Jones	Conceptualization (lead) Investigation (equal), Project administration (lead), Writing – original draft (lead), Writing – review & editing (equal)
Elizabeth Newbold	Investigation (equal), Writing – original draft (equal), Writing – review & editing (lead)
Jordan Willis	Writing – review & editing (equal)
Katie Yates	Investigation (equal), Project administration (equal), Writing – original draft (equal), Writing – review & editing (equal)
Teagan Zoldoske	Investigation (equal), Writing – original draft (equal), Writing – review & editing (equal)

2.1 Data sharing and FAIR data

Sharing energy data used for research is good practice for responsible and reproducible research and for publicly funded research it is expected that data generated will be as open as possible and as closed as necessary. Data can be hard to define but this report will use the OECD definition of data [3]:

"Data or research-relevant digital objects, such as metadata, algorithms, code, software and workflows, used as sources for research or generated as a result of research."

The FAIR Data principles [4] encourage researchers to make their data Findable, Accessible, Interoperable and Reusable and it is this framing that we will be using in this report. In addition, the FAIR Principles for Research Software [5] would also apply to the models themselves. Both sets of principles use the same overarching framing.

Many models used in UK energy research are funded by UK Research and Innovation (UKRI) and subject to their data management and sharing expectations. The UKRI guidance [6] addresses data used in modelling:

"For example, in the case of simulated data or outputs of models, it may be more effective to preserve the means to recreate the data by preserving the generating code and environment, rather than preserving the data themselves."

Challenges in the energy research sector include the multidisciplinary nature of the research where different data management and sharing cultures, expectations and terminologies come together. With the increase in emphasis on reproducibility and making data behind publications available, it is an opportune time to consider data sharing issues for energy modelling data.

There is significant work in the community on the openness of models. The UKERC Modelling Hub, funded in UKERC4 (2019-2024)was focussed on improving the transparency and openness of models. It did a survey of models which identified 76 models, analysed the results and produced four publications [7], [8],[9],[10]. Key findings of this work were:

- All modellers wrestled with how complex the model should be, how to fund and maintain the model, how transparent to make its inputs and outputs and how to use them for decision-making.
- There was a spread of modelling methodologies in use covering all sectors in the energy transition space.
- No single model will cover all the elements required to understand the energy transition, so linking models and input/output data was recommended. The authors considered time, space, technologies, and behaviour.
- Energy models are expensive to build, and maintenance is unglamourous and time consuming.
- The 2021 survey found that around half of the surveyed models are not transparent (open documentation, open access or open source).

- Models rely on historical data, data from other models, expert opinion from stakeholders, and calculation by modellers themselves all as key inputs.
- UK energy models are both built by academics, consultants and government modellers and applied to policy, business and research questions.

Many of these points were again identified by participants in the workshop being reported here.

2.2 DINI & barriers

The Data Infrastructure for National Infrastructure (DINI) project explores the challenges and opportunities in data sharing within the domain of national infrastructure systems research with a focus on energy, transport, and water.

The DINI project team have identified potential barriers to data sharing categorised into five broad areas: privacy and legal, security, commercial, cultural, and technical. More details of these are in Appendix II. These have been used as part of the workshop outputs analysis.

2.3 Framing of the modelling problem space

Energy modelling as a technique is used in many parts of energy research, government, regulation and commercial operations. This workshop was focused on the use of models for research purposes, although it is acknowledged that this is a blurred boundary. Modelling is used to understand and explore a wide range of topics from the power network, to political and economic impacts. Most modelling needs real-world data as the input to the model and this is often generated/gathered by other researchers or commercial companies.

Mike Colechin and Andy Boston designed the framing for the workshop as shown in the diagram below.

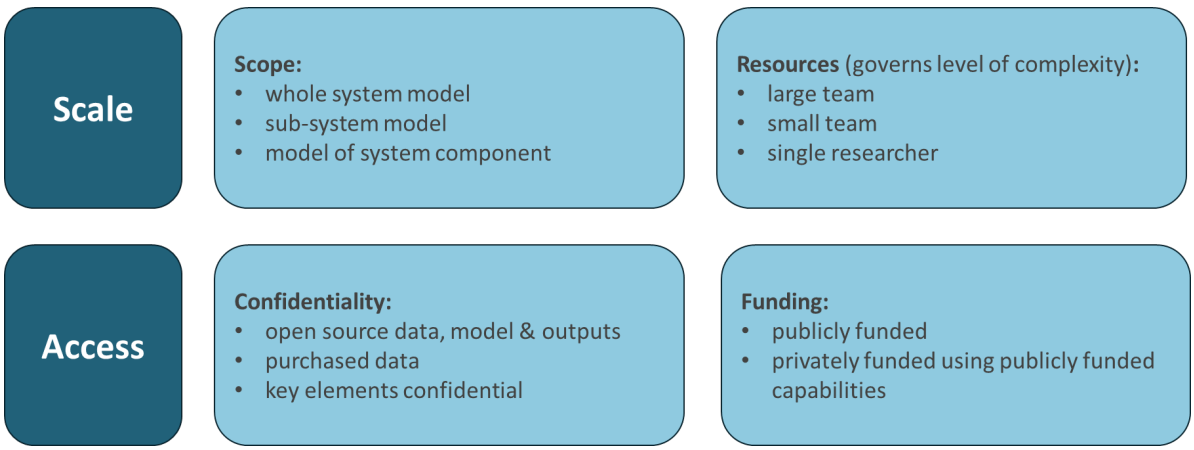


***Figure 1**: Energy modelling dimensions.*

Energy models are defined in this framing by two axes: scale of the model and access to the model and data. Taking these in turn.

The scope defines what problems the model seeks to answer, it may be addressing a whole energy system, specific subsystems or an individual component. The wider the scope the greater the potential complexity and the likelihood that the input data comes from a wider range of sources with issues of compatibility and data conversion.

The resources needed to build and maintain the model reflect and govern the complexity and maintainability of the model. These can range from a large, distributed team to a single researcher. The team size affects take up of the model (community building takes effort), the reliability of the outputs and the ability to adapt when the model needs updating.

Access to the data needed and generated by the model can be determined by the levels of confidentiality and IP involved. These range from open-source models using open-access data, through models that are open but need data purchased for input, to models where significant elements are confidential and may not be shared beyond the modelling team. Linked to this is the distribution method, some models are open source and can be downloaded or forked for independent modifications, some are closed, and some may have a single source. Managing communities around models is a complex activity. Energy modelling by its nature can be affected by restrictions on information on critical national infrastructure, which can make academic modelling more challenging.

The funding for the model and/or data can affect the ability to share input or output data or the model itself.

2.4 Workshop description and outputs

The outputs will be:

- This report containing an analysis of the workshop and identification of key messages and recommendations.
- A blog (to be shared with the wider research community via the UKERC websites and newsletters).
- The summary of key messages and recommendation to contribute to the DINI final report.

This workshop explored data sharing for energy modelling to identify barriers and key messages to overcome them. In breakout groups, participants were asked to:

- Articulate specific issues or concerns with sharing data in the energy modelling space.
- Consider the landscape from a specific perspective and suggest some models that fit that perspective.
- Explore the attributes and issues of the examples.
- Develop some key messages.
- Compare key messages with another breakout group to put forward the most important.

The five groups were asked to address the issues from the following perspectives:

- Group 1: Small team
- Group 2: Large team
- Group 3: Publicly funded
- Group 4: Privately funded
- Group 5: Whole system modelling

A visual tool (Mural) was used to capture responses from participants during the group discussions. The energy modelling landscape framing (see section 2.3) was used for this process. The outputs of this process were analysed and are discussed in sections 3 and 4 and the recommendations are in section 5.

While the workshop was focused on data used and generated by models in the energy domain, some of the discussion and recommendations refer to the models themselves.

3. Initial Issues Raised

The first session started with the opportunity for participants to raise issues they had encountered in sharing data relating to energy modelling. These have been grouped using the FAIR (Findable, Accessible, Interoperable and Reusable) principles, noting which DINI barrier areas apply.

3.1 Findable

Key concerns were the discoverability of data and the longevity of the availability of data. It was noted that organisational change, especially in the commercial sector, can lead to an increased risk of data loss/disappearance.

It was acknowledged, especially due to the increases in data volume, there may be difficulties in finding the funding of resources for storing the data.

These issues all broadly fall into the **DINI cultural barriers.**

Additionally, another concern that was not covered by the DINI barriers, was how data about emerging technology, such as costs and environmental impact, becomes findable and available due to the fact that it is new and a novelty; this is a live issue which needs addressing within the sector.

These are all linked to the longevity of access especially if the data is about infrastructure components/real-life monitoring data which are needed for a model - these infrastructure assets last a long time in real life. There is also a need to ensure that data about new technology which will become new infrastructure assets is available and discoverable in a timely fashion.

These concerns are mostly related to input data, the only reference to output data was about funding for storage.

Locating additional data for the inputs of models is time-consuming due to the need to identify what is in the dataset, coverage and restrictions and determine if they are compatible with the model and of good quality. Whilst this is related to the **DINI technical barrier** there is some overlap with the **DINI cultural barrier** and having appropriate skills.

Resolving some of the discoverability issues by the provision of an integrated platform to provide access to models and appropriate data would be beneficial.

3.2 Accessible

Data availability, especially from commercial companies, was identified as an issue, specifically the type of data (time series and special event data from the electricity supply industry). Privacy concerns and confidentiality being a barrier to open data due to the conservative industry attitudes were mentioned by all groups at different levels (mixture of **DINI Legal and privacy & cultural barriers**). It was noted that data gatekeepers can be reluctant to share. However, the legal challenges and public investment considerations need to be understood when looking to share with private companies.

Working in the energy modelling space around infrastructure means that there are tensions between concerns of export controls, protecting IP and protecting critical national infrastructure (**DINI security barrier)** versus making data open and FAIR. Making sharing a requirement of funding was considered to be beneficial. IP or privacy restrictions can sometimes mean that data can only be shared with project partners or those who collected the original data used in the modelling.

This led to a discussion on risk appetites for sharing UK systems power data, where some things can be inferred from available data and the importance for longevity of commercial data for infrastructure components which are usually supplied as a “black box”. These were additional examples that could be classified as relating to the **DINI legal and privacy & cultural barriers.**

The lack of knowledge of appropriate licensing leads to fewer licenses being applied to data and software which is a barrier to further sharing (**DINI commercial and DINI cultural barrier**).

All the accessibility concerns referred to input data, and this is to be expected as some areas of energy modelling depend on commercial data, and this can be a complex area as there are many organisations in this domain. The focus of the participants tended to gravitate to the barriers they experienced in accessing data that they wanted to use and not on making the data that they produced accessible to others.

3.3 Interoperable

Key issues include:

- Using open source and commercially sensitive data together as inputs is a challenge.
- Data formats vary between different software and there may need to be a lot of data cleaning before it can be used (which may require specialised domain knowledge).
- Using data for inputs to a model from a variety of sources can be challenging due to format differences and level of documentation, unified formats were proposed as a solution.
- People collect and think about data in different ways and have different training, so the same kind of data may be recorded differently or kept in multiple domain related places.
- Lack of standards, especially to describe model outputs to enable comparison, were raised by all groups.
- The need for documentation for example if there was not existing metadata and other standards to be able to reuse, it was felt that these should be specified at the start of a project so that the data generated was interoperable by design.

All these issues contribute to the recognition that combining data from multiple sources is complex and comparing outputs is difficult due to the lack of semantic agreement. Interoperability issues apply to both input and output data.

All the issues raised in this section relate to and map against the **DINI technical barrier characteristics.** Additionally, the DINI characterisation of **legal and privacy** issues was raised in relation to the use of open source and commercially sensitive data.

3.4 Reusable

A number of points raised related to the reusability of research data and mostly these could be considered to be **technical barriers.**

- Data quality was raised by most groups, either regarding consistency of data or that open data may be low quality and this needs to be addressed.
- Related to data quality was the issue of inadequate documentation and metadata, especially for historic datasets.
- The increasing size of data sets was an issue, both for using them and assessing the quality of them.
- A participant identified Equality, Diversity and Inclusivity challenges within the energy modelling research area.

Some industry related data, especially for specific components, can be considered to be a “data black box” which is prevalent in private sector interactions. This can make it difficult to assess the longevity of a model as it isn’t clear how long that black box will be valid for, which when assets in the energy sector have lifetimes of decades is an issue. They may also become an issue with AI (for explainability). Whilst this could be considered a technical barrier it is an emerging use case that needs to be considered as a “new” barrier/challenge to be considered and assessed.

Understanding the data and how it fits into the reuse anticipated is a large area and documentation and common terminology would assist in the process.

In addition to the technical barriers to reuse that were identified and discussed there was an emerging theme around ethical issues (**DINI cultural barrier**). For example, there was reported use of AI to transform data which the participants of the specific group felt needed to be documented well enough to understand any bias that this type of technology might introduce. Use of synthetic data to replace sensitive data, such as smart meters was identified as a possible solution to some of the challenges. Ethical issues, such as those raised by smart meter data, need to be carefully considered and managed effectively; however, this can lead to it being more difficult to aggregate responses from many models. The importance of context for data collected by others was highlighted, specifically data history.

3.5 Sustainability and maintaining models

A key issue was the funding landscape that academic modellers work in which gives the perception of a lack of support or incentives for creating models with longevity. It is also difficult to centralise funding for community developed models maintained with project funding.

This leads to a duplication of model done by a transient (PhD) workforce who are focused on the research, not on good software development practices, which makes building on existing work hard. They are created with short term research goals, not for making something reusable.

A significant difference between well-resourced models with longevity is the quality of software engineering. Models generated by a small group, or a PhD student, are less likely to consider software engineering concepts of reusability and error handling as they are created by domain specialists who may not have formal programming training. This can leave these smaller models vulnerable to becoming lost or incompatible with current software. Methods/models may be written up in a thesis and no longer used. Models based in a research group may be developed by a series of PhD students, building on each other's work.

The importance of communities around specific models and the effort needed to run and participate in them was raised. This perceived lack of an organised community can lead to not enough notice of changes to models and a loss of control for researchers using these models.

4. Examples and Discussion

The participants identified models that either fitted the perspectives of the group or that they had knowledge of and were able to discuss. Models that have a public presence are described below together with the discussion points made in considering these models, followed by a section on the types of closed models.

Three common themes emerged from the discussions on these specific models.

Standards and documentation: as previously highlighted the need for standards and documentation are increasingly important to the modelling community to support reuse, irrespective of whether a model is publicly available or limited to a closed community. Documentation could also be expanded to include the need for data sharing agreements to be available, it is not always clear where these are available and in place.

Data availability: a variety of issues related to data availability and quality were identified, which limit the usefulness of models. Issues range from lack of standardisation of data to specific sector data not being available. This links back to the need for good documentation to understand what limitations there may be.

Community and collaboration: all the participants highlighted the need for community and collaboration to maintain models, noting that this is not straightforward especially when working across sectors and boundaries for example. academia/industry/public sector.

4.1 Energy Transition Model

The Energy Transition Model (ETM) [11] is a free-to-use interactive online simulation tool for energy systems, allowing users to explore potential future energy systems in detail. Users can simulate supply, demand, emissions and more across the entire energy system for a given region. Allowing users to predict the potential impacts of factors such as, increasing energy demands, transitioning to greener forms of energy production and more.

Framing: whole system / large team / open source / delivered by private sector organisation (with public sector partners).

Developed and maintained by Quintel, the model is **open-source** and uses **openly available input data**. By not including industry data, the model remains openly available for all to use but may not be truly representative.
This model was identified by two different groups. Specific issues identified was the need to understand who is **accountable**, that **standards** and **documentation** are important and the fact that the model only uses open data is a **limitation** of the model as it means that **not all the data needed for an area being modelled is available**.

4.2 MARCO UK

MARCO-UK (UK MAcroeconometric Resource COnsumption) [12] is a macroeconometric (ME) model based on post Keynesian economic theory. It is well suited for capturing demand-side stimulus to the economy, as well as the role of energy efficiency in the economy. However, like other ME models, the model can struggle to study how a structural change in the economy would respond in the future and may forecast historical trends to continue.

Framing: whole system / large team / licensed model / publicly funded

This was identified by a single group. They noted that it uses **publicly available data** but the model is available through a licensing model. **Collaboration** is key to the success of this model but this is difficult across sectors and needs resources to sustain both the collaboration and the model themselves. The open data may be of **inadequate resolution**, **some sector data is missing** and there may be issues with **standardisation** of metadata and the lack of information on assumptions, which also impacts on the ability to work with the uncertainty.

4.3 PyPSA

PyPSA (Python for Power System Analysis) [13] is an open-source toolbox for simulating and optimising modern power and energy systems. The framework contains multiple packages that can be used to conduct a variety of studies from calculating renewable energy potentials to creating complex electricity networks and cost-optimizing the pathway of a whole energy system. Each package is built in a modular form so that they can be used independently of each other but also interact easily.

Framing: sub-system / large team / open source / public & institutional funding

This was identified by a single group. This toolbox has thousands of international users, each creating their own instances parameterised with data for their own energy problem. This highlights a distinction between well-resourced modelling frameworks which are well maintained and documented whereas the parameterised local models are used by smaller teams and may be **less well documented** and maintained.

4.4 TIMES Model

The TIMES (The Integrated MARKAL-EFOM System) [14] is an economic model generator for local, national, multi-regional, or global energy systems. It is a bottom-up model generator that uses linear-programming to produce a least-cost energy system, optimised according to a number of user constraints. Users create scenarios using four types of inputs: energy service demand curves, primary resource supply curves, policy setting, and the descriptions of a complete set of technologies. Once all the inputs, constraints and scenarios have been entered, the model will attempt to determine an energy system that meets the energy service demands at the least cost. The TIMES frameworks have been used by numerous institutions **around the world** to develop national, regional, and global models to analyse the impact of energy, environment and climate policies.

Framing: whole system / large team / open source / publicly funded

This model was identified by two groups. The TIMES model has a large international community, the UK-TIMES is the main curated version, but individual institutions/government department will have their own local variants based on

different data (privacy/security/IP restrictions). This complex usage pattern means that there is **not consistency** between modellers from different groups as the inputs may not be the same and that it is difficult to disentangle this restricted data from the core model to be able to publish it. The benefit is that it allows PhD students to conduct whole systems modelling without having to start from scratch. The participants noted that the model acts as a means for knowledge sharing between government/academia/industry and a sense of **community is important**. It was noted that some types of data should be similar across models.

4.5 SOLPS

SOLPS (Scrape-Off Layer Plasma Simulation) [15] is used to model and study plasma boundary physics within the scrape-off layer region of a fusion reactor. Within a fusion reactor, the periphery of the plasma (the scrape-off layer) is a critical buffer region between the hot core and the walls of the reactor. Understanding the behaviour of plasma within the scrape-off layer is critical to ensure the correct functioning of the reactor.

Framing: system component / large team / open source / publicly funded

This was identified by a single group. They noted that it was an example of a model that has diverged through forking and that there **is not a data sharing policy** associated with it.

4.6 Closed models

There were a series of models which are not openly available. They covered energy demand predictions, a commercial project including Hartree researchers and academics, a heating decarbonisation optimisation model and an optimisation model looking at choice of energy providers. For all these models, the input data (or part of it) is **openly available data** but the output data may or may not be.

5. Key Messages and Recommendations

Using the FAIR and intervention framing, the key messages identified by the participants have been brought together, analysed and recommendations made.

5.1 Findable

Message	DINI Barrier	Affecting
Effort needed to find previous modelling work to avoid duplication of effort but need to balance this against the costs of updating technical obsolescence.	Cultural (discoverability)	Models

This message has identified a piece of discovery infrastructure which is currently missing and would be addressed by creating a catalogue or registry for models (not a service to store models). This could build on the work done in the UKERC Modelling Hub previously discussed.

Scoping would be required to establish at what point a model would be ready to be registered, and the updating process to reflect changes in model status such as entry when a model is no longer actively maintained.

This registry should be defined and implemented to enable federated searching across a landscape of model repositories.

Recommendation: Establish requirements for a registry for energy models.

The first phase should be focused on models created in academia. Undertake further requirements scoping work for a model registry. This is something that could build on work done by UKERC and the Energy Data Centre (EDC) could lead this work within a future phase of DINI.

Recommendation: Enhance EDC guidance on software citation to explicitly cover citation and referencing of models to support discoverability within the literature.

5.2 Accessible

Message	DINI Barrier	Affecting
The need for a central platform that stores data in a unified format enabling access to the data and the model in the same place. This platform would enable the integration of data from a variety of sources and provide High Performance Computing (HPC) resource. This HPC resource would lower the barrier to participation for research groups that do not have access to this.	Technical (Standards and Quality)	Data and Models
Machine access, such as API, to make data more accessible.	Technical (Computing Support)	Data
Build/facilitate community(s) that share resources and training across research groups and industry. A platform might help become a community hub. This takes time and resource.	Cultural (Discoverability & Skills)	Models
Need to break boundaries between academia and industry – get people to move across the boundary.	Cultural (Discoverability & Skills)	Data and Models

These messages from the participants contribute to demonstrating that there is a need for research data clouds in general and an implementation of Data Infrastructure for National Infrastructures in particular. Together with the possibility of building a community around the platform and tools & services provided. This is a visionary idea which would take a significant amount of resource.

The central platform and computational resource are provided in part by the Data Analysis Facility for National Infrastructure (DAFNI) but would need to be extended in collaboration with other partners. There has been work in other disciplines on data integration and standardisation such as CLOSER for epidemiological data, and work undertaken by the NERC funded centres on common terminology, and within the DSIT-funded EDS-BOOST project [16]. There is scope for further work in Trusted Research Environments building on the SDR Registry work funded by DSIT.

Recommendation: There is an identified need for a central platform to share data, run models and facilitate communities. Ensure that the requirements identified form part of future DINI implementation plans.

5.3 Interoperable

Message	DINI Barrier	Affecting
Work towards standardisation of terminologies to create shared understanding of terms used in energy research.	Technical (Standards)	Data and Models
Standardised metadata for domains and common understanding where domains meet.	Technical (Metadata)	Data
Encourage interoperability between model outputs at differing system levels to form whole-system modelling.	Technical (Standards)	Data and Models
As AI becomes more widely adopted, there is the potential for standardisation/recommendations at this early stage.	Technical, Privacy & Commercial	Data and Models
Need to consider how to accommodate emerging energy vectors into models.	Technical (Standards)	Data and Models

Energy research is a multi-disciplinary field and with the emergence of a whole system approach within the community there is an increasing requirement for common understanding of terminology and the ability to link models together. These messages reflect the practical need to standardise in this area.

There are many bodies, such as professional bodies and EU research projects that have sought to produce work on standardised metadata and ontologies.

The DINI project is reviewing the status of energy ontologies and will produce a report.

Recommendation Create a working group to review classifications of energy data for statistics (IEA; Eurostat; DUKES) and to develop a multi-level classification of energy data that we can use to: (i) compare outputs from different models; (ii) output in international statistical formats; and, (iii) output in formats for international initiatives (e.g. for IPCC comparisons led by IAEA, IIASA).

There could be a role for UKERC/EDC/DAFNI as a convener of this activity.

Recommendation: Contribute to the AI adoption debate by writing a position paper on this topic, with input from DAFNI, UKERC and other collaborators.

5.4 Reusable

Message	DINI Barrier	Affecting
Metadata should be able to recognise the uncertainty in the data.	Technical (Quality)	Data
Metadata should describe the data quality so others can make informed decisions whether to use.	Technical (Metadata and Quality)	Data
Assumptions used to create synthetic data should be clear.	Technical (Context)	Data
Lack of clarity about governance of models and change control.	Legal (Ownership) and Cultural (Governance)	Models

An important facet of being able to reuse data is an understanding of the quality, coverage and assumptions used to create or generate the data and these messages highlight this topic. There are resources available for creating and sharing FAIR data, but none specifically (in English) for energy modellers.

Recommendation: Produce and publicise guidance available from the EDC to support the creation of FAIR data to be used in modelling.

Recommendation: Creation of appropriate contextual information should be done routinely by researchers as part of good research practice, having received appropriate training and support.

5.5 Interventions: specialist support and training

Message	DINI Barrier	Affecting
Many small models are written by technical experts who are not trained to do software development which leads to maintenance issues. Additionally, it is difficult to sustain models as the current research funding model does not support the maintenance costs of models.	Cultural (Skills)	Models
Responsibility for documentation should be clear and governance and policy for this should be in place.	Technical (Context)	Models
Lack of knowledge about appropriate licensing for model outputs.	Commercial (Licensing) and Cultural (Skills)	Data

These messages, while covering different topics in the data and software management space, are all related to the boundary between research domain experts and research technical professionals such as Research Software Engineers and Data Stewards. The increase in these professions within universities and the ability to add these types of roles to grant applications may lead to improvements over time. There is a case to improve the basic software development capabilities of PhD students.

There is an opportunity for funding to support sustainability in innovative ways. An example of this is the MRC exploratory funding call that supported software maintenance linked to improving methods. Part of the Better methods, better research funding call enabled the developers to receive funding as part of method improvement: "*Costs to support open-source or researcher-generated open-software resources to facilitate the use of methods are eligible under usual MRC guidance*." This approach might benefit the energy modelling community.

Recommendation: Research organisations should enable researchers to have the right skill sets or access to professionals to create long-lived research outputs.

Recommendation: Research funders should consider how to support models created through research funding to gain the maximum benefit from the models over the longer-term.

5.6 Interventions: research design

Message	DINI Barrier	Affecting
Data cleaning takes effort and should be rewarded.	Cultural	Data
More attention to EDI issues in the data collection for true representation.	Technical (Quality)	Data

These messages reflect the importance of various aspects of research design and should be covered in the professional development programme of researchers.

Recommendation: Research organisations who train, support and employ researchers need to ensure that credit for data management and sharing is reflected in the professional development programme and reward structure.

6. References

[1] Colechin M and Quigley C. (2023) Improving Data Sharing in Energy Consortia: Summary of Workshop Outputs. https://doi.org/10.5286/UKERC.EDC.000970

[2] Jones, C., and Higginson, S. (2023) CREDS/UKERC Data sharing workshop briefing note. https://doi.org/10.5286/UKERC.EDC.000971

[3] OECD. (2021) Recommendation of the Council on Enhancing Access to and Sharing of Data. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0463.

[4] Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018. https://doi.org/10.1038/sdata.2016.18

[5] Chue Hong, N. P., Katz, D. S., Barker, M., Lamprecht, A.-L., Martinez, C., Psomopoulos, F. E., Harrow, J., Castro, L. J., Gruenpeter, M., Martinez, P. A., Honeyman, T., Struck, A., Lee, A., Loewe, A., van Werkhoven, B., Jones, C., Garijo, D., Plomp, E., Genova, F., RDA FAIR4RS WG. (2022) FAIR Principles for Research Software (FAIR4RS Principles) (1.0). https://doi.org/10.15497/RDA00068

[6] UKRI. (2018) Guidance on best practice in the management of research data. https://www.ukri.org/wp-content/uploads/2020/10/UKRI-020920-GuidanceBestPracticeManagementResearchData.pdf.

[7] Strachan, N. and Li, P-H. (2021) Energy modelling in the UK: Briefing paper 1: The modelling landscape, UKERC. https://doi.org/10.5286/ukerc.edc.000943

[8] Li., P-H. and Strachan, N. (2021) Energy modelling in the UK: Briefing paper 2: Strengths and weaknesses, UKERC. https://ukerc.rl.ac.uk/publications/policy_briefing_paper/UKERC_BN_Modelling-strengths-weaknesses.pdf

[9] Li., P-H. and Strachan, N. (2021) Energy modelling in the UK: Briefing paper 3: Construction, maintenance and transparency, UKERC. https://doi.org/10.5286/ukerc.edc.000949

[10] Li., P-H. and Strachan, N. (2021) Energy modelling in the UK: Briefing paper 4: Decision making in government and industry. UKERC. https://doi.org/10.5286/ukerc.edc.000950

[11] Energy Transition Model (ETM) https://energytransitionmodel.com/.

[12] MARCO-UK (UK MAcroeconometric Resource Consumption) https://www.researchgate.net/publication/339018267_Socio-macroeconomic_impacts_of_meeting_new_build_and_retrofit_UK_building_energy_targets_to_2030_a_MARCO-UK_modelling_study.

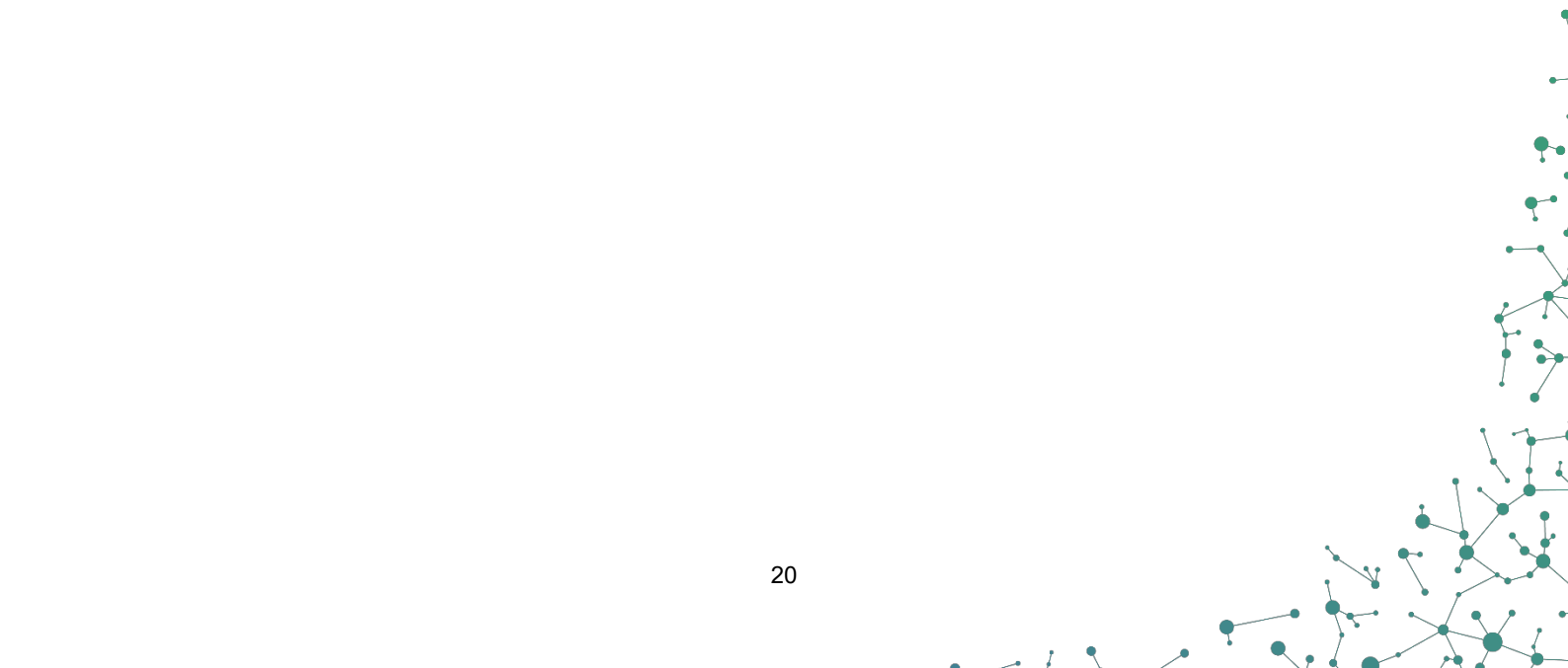
[13] PyPSA (Python for Power System Analysis) https://pypsa.org/.

[14] The TIMES (The Integrated MARKAL-EFOM System) https://times.readthedocs.io/en/latest/part-1/01-introduction.html.

[15] SOLPS (Scrape-Off Layer Plasma Simulation) https://www.iter.org/node/20687/new-solps-iter-code-version-launched.

[16] Blair, Gordon S.; Baker, Garry; Bee, Emma; Bowie, Jo; Byrne, James; Darroch, Louise; Fremand, Alice; Garland, Wendy; Harrison, Kathryn; Heaven, Rachel; Ten Hoopen, Petra; Juckes, Martin; Kingdon, Andy; Kingston, Richard; Kokkinaki, Alexandra; Lewis, Edd; MacRae, Molly; Mobbs, Stephen; Moncoiffe, Gwenaelle; Pascoe, Charlotte; Passmore, James; Peat,

Helen; Pepler, Sam; Shepherd, Andy; Siddorn, John; Snaith, Helen; Tate, Alex; Townsend, Poppy; Walsh, Colm; Watkins, John. 2024 Contributions to the development of the next-generation NERC Environmental Data Service: Building Interoperability - a NERC Data Commons RoadMap. UK Centre for Ecology and Hydrology. Unpublished.



7. Appendix I: Workshop Agenda

09:30 Welcome & Introductions

09:55 **Session 1** – Exploring the Energy Modelling Space

10:40 **Session 2** – Data Sharing Challenges

11:30 **Session 3** – Recommendations

12:00 Feedback & Discussion

12:30 FINISH

For session 1 and 2, the participants were split into five groups, for session 3 participants were brought into combined groups. There were 24 attendees from a range of academic and government organisations. The career stage for participants from academic organisations ranged from PhD student to professor. The majority of participants were active modellers. For some organisations more than one person attended. The list of organisations are shown below:

Centre for Net Zero
Department for Science, Innovation & Technology
Department of the Economy, Northern Ireland
Energy Demand Research Centre
Imperial College London
National Grid ESO
Science and Technology Facilities Council – Energy Research Unit
Science and Technology Facilities Council Hartree Centre
The University of Manchester
UK Atomic Energy Authority
UK Research & Innovation – Innovate UK
UK Research & Innovation, Open Research
University College, London
University of Bath
University of Birmingham (Energy Demand Research Centre)
University of Durham
University of Herriot Watt
University of Strathclyde
University of Surrey
University of York
York St John University

8. Appendix II: List of Barriers Identified by DINI

	Barrier	Description
1	**Privacy and Legal Barriers**	Relating to legal and personal restrictions
1.1	Personal Sensitivity	The data has additional restrictions due to personal sensitivity.
1.2	Commercial Sensitive	The data has additional restrictions due to commercial sensitivity.
1.3	Legal Ownership	It is not clear who owns the data. The ownership of the data restricts access.
2	**Security Barriers**	Relating to security concerns
2.1	Organisational Security	Is the data restricted due to organisational security concerns?
2.2	National Security	Is the data restricted due to national security concerns?
3	**Commercial Barriers**	Relating to commercial concerns
3.1	Cost/Economics	There are financial costs to access the data. Do the resources needed to make the data useable for the intended purpose outweigh the benefits of using this data?
3.2	Licensing	The license assigned to the data means that it can't be used for the intended purpose. There is no license assigned so the data is riskier to use.
3.3	Contractual Issues (DSA's)	The Data Sharing agreement is not acceptable to potential organisations. It takes too long for the DSA to be agreed for it to be used for the intended purpose.
4	**Cultural barriers**	Relating to people and approaches
4.1	Discoverability	Is it difficult to find or access available data?
4.2	Organisational Resistance	Is there indifference or resistance within the organisation that obstructs data sharing? Does this come from a lack of understanding as to the benefits of data sharing?
4.3	Ethics	Are there ethical consequences of the data being shared e.g. unfair treatment or adverse consequences to particular stakeholder groups.
4.4	Lack of appropriate skills	Are the skills not available at the data source or by the data user to access and use the data?
4.5	Governance	Is there a lack of strategic oversight or governance on the data within an organisation?
5	**Technical barriers**	Relating to processes and procedures

	Barrier	Description
5.1	Reliability	Is the data accurate and relevant? Is it out of date?
5.2	Interoperability/ Standards	The data is in file formats that are difficult to share (proprietary) or there are no standards applied to the data.
5.3	Metadata	The data is not well described and so it is difficult to find and use.
5.4	Quality Measures	The data has not undergone any quality control. The quality control measures are not documented.
5.5	Lack of suitable Computing Support	Are computing systems insufficient to support the effective sharing and reuse of data?
5.6	Contextual Information	There is not enough information associated with the data resource to enable someone outside the data creator/generator group to use it.
5.7	Coverage	Is the data accurate, but does it not give the coverage required? This may be coverage in terms of time frame, geographic coverage, or granularity.