Science and Technology Facilities Council

Technical Report
STFC-TR-2025-004

Data Infrastructure for National Infrastructure. A UK Research Data Cloud Pilot: Final Report

BM Matthews, K Cartmell, CM Jones, E Newbold

April 2025

Executive Summary

National infrastructures, such as energy, water, and transport systems, are critical to society. Exploring the effect of investments in infrastructure is an active area of research with a high impact on the well-being of the UK. Data is an essential pre-requisite for good analysis and good decision making, but there are many barriers to the effective use of data.

Data Infrastructure for National Infrastructure (DINI) is a pilot study within the Department for Science, Innovation and Technology's UK Research Data Cloud Pilot programme. Its aim is to explore the potential of data to drive research and its impact on policy. Our scope is National Infrastructure Systems within the UK with a focus on energy, water and transport. The project was coordinated by the Data and Analytics Facility for National Infrastructure, working with the Energy Data Centre and the JASMIN Facility, with contributions from Icebreaker One, the Digital Curation Centre, and the UK Collaboratorium for Research on Infrastructure and Cities.

We reviewed the potential impacts of sharing infrastructure data for research, and the barriers which can get in the way. We also considered some key technologies needed to build a data cloud, and sponsored demonstrator use cases to illustrate the use of data in research practice.

Research into the evolution of infrastructure systems can have a real impact on the delivery of the strategic goals of government. Our work in use cases in this project has demonstrated that there can be multiple impacts of research; we highlight the potential contribution to the strategic goals of building a secure and resilient society, of boosting economic investment and innovation, and of ensuring the health of our environment.

However, we have also seen the extent that problems in accessing and sharing data have impeded the interaction between researchers and partners in industry and government. Our participants have highlighted barriers to data sharing, particularly the priority that data suppliers give to sharing data safely and securely.

We have made sixteen recommendations which address these challenges; these are summarised below. Together, these characterise a research data cloud to coordinate and sustain the management of data so it can be shared with the academic sector while satisfying the concerns of data suppliers.

Recommendations

1. **Data providers and policy makers** must invest in co-designing appropriate and open governance for data sharing within and across sectors.
2. **A DINI** should manage trusted access and broker data sharing agreements between data suppliers and researchers.
3. **A DINI** should broker access to industry and government data for use within research.
4. **A DINI** should support the identification and creation of reference data with clear access conditions and agreed data sharing arrangements.
5. **A DINI** should provide guidance for data suppliers and researchers on best data publication practices.
6. **A DINI** should build a network of communities of practice through guidance, training and setting standards.
7. **Research funders** should advocate for the practical and strategic benefits of sharing national infrastructure data with researcher.
8. **A DINI** should have a long-term plan and sustained investment on its maintenance.
9. **Funders of research and other** data sharing programmes should invest in best-practice case studies with commercial organisations to demonstrate the tangible benefits of academic re-use of data.
10. **A DINI** should provide access to data from a federated infrastructure of trusted repositories.
11. **A DINI** should provide access to federated trusted computing resources to perform analysis.
12. **A DINI** should support a common data interoperability framework in conjunction with wider community efforts to set data sharing standards.
13. **Data suppliers and users** should work to establish and apply, domain expectations and standards on data quality.
14. **A DINI** should provide services to discover and access research data, and other associated research outputs**.**
15. **A DINI** should provide data stewardship expertise to work with researchers and other data suppliers to support long-lived research products.
16. **Researchers** should be encouraged to manage and describe their research outputs, such as data, models, research software and protocols, to ensure application of FAIR principles and to support interoperability, transparency and reproducibility.

Further, we have developed a vision for a future Data Infrastructure for National Infrastructure would deliver such a research data cloud which satisfy these recommendations; this is outlined below. We would propose that future work should consider how a DINI might be established.

A Data Infrastructure for National Infrastructure should coordinate digital research infrastructure and expertise to:

- **broker** with industry and government for the use of data within research**;**
- **provide access** to data via a federated infrastructure of trusted repositories**;**
- **support** a common data interoperability framework;
- **provide access** to trusted computing resources to perform analysis;
- **foster** a network of communities of practice through guidance, training and setting standards;
- and **collaborate** with cross-sector initiatives to build consensus.

The value which can return from data is recognised in the AI Opportunities Action Plan[1], which calls to *"build public sector data collection infrastructure and finance the creation of new high-value datasets that meet public sector, academia and startup needs"* and *"incentivise and reward researchers and industry to curate and unlock private datasets*. The National Data Library initiative is considering how to bring together existing research programmes and help deliver data-driven public services. A DINI would complement this goal, forming a bridge to the Infrastructure Systems Engineering research sector.

Infrastructure systems are inherently complex and multi-disciplinary, so the solutions require partnership, systems thinking and interoperability. We propose that a DINI should continue to build the community connections to meet the complexities that are faced by organizations across industry, government and academia.

[1] AI Opportunities Action Plan, Independent Report, Department for Science Innovation and Technology, 13 January 2025, https://www.gov.uk/government/publications/ai-opportunities-action-plan

Table of Contents

Introduction

National infrastructures, such as energy, water, and transport systems, are critical to the functioning of society. Exploring the potential effect of investments in infrastructure systems is an active area of research within Engineering and related disciplines with a high impact on the social, economic and environmental well-being of the UK. Computational modelling and machine learning are increasingly important methods used in this research, allowing researchers and policy makers to explore different scenarios and assess the potential impacts on society. Data is an essential pre-requisite for good analysis and good decision making, but there are many barriers to the effective use of data in the infrastructure systems engineering domain.

Data Infrastructure for National Infrastructure (DINI) is a pilot study within the Department for Science, Innovation and Technology's (DSIT) UK Research Data Cloud Pilot programme (November 2023-March 2025). The overall aim of the DINI project is to explore the potential of the huge amount of data to drive research and the impact of that research on policy making. Our scope is National Infrastructure Systems within the UK, including the devolved nations, with a focus on energy, water and transport, and the related built, social and economic, and natural environments.

We reviewed the potential impacts of sharing infrastructure data for research, and also the barriers which can get in the way. We used this review to recommend a research data cloud environment be established to coordinate and sustain the management of data is needed to make data more coherent and sharable with the academic sector. We also considered some key technologies needed to build such a data cloud. This research data infrastructure could contribute to a future National Data Library, and ultimately the strategic missions[2] of the UK Government alongside other initiatives in this space.

The project was coordinated by the Data and Analytics Facility for National Infrastructure (DAFNI), working with the Energy Data Centre (EDC) and the JASMIN Facility, with contributions from Icebreaker One, the Digital Curation Centre (DCC) and the UK Collaboratorium for Research on Infrastructure and Cities (UKCRIC). Together, we undertook a landscaping exercise, with a literature review, survey and workshops and interviews from across the four nations of the UK and across the three sectors, with representatives from the research community, industry and the public sector. We sponsored a series of demonstrator use cases from real research studies to provide in depth studies in practice to gain insights into data sharing challenges and success in specific research projects; these are listed in *Table 1*. Additionally, design studies were undertaken by exploring specific technical aspects required for future implementation.

[2] https://www.gov.uk/missions

About this report

In this report, we summarise the outcomes of the Data Infrastructure for National Infrastructure pilot project.

We begin by presenting our vision for a DINI. This provides a statement of the major action which we propose from the work of the project, in response to its recommendations.

The report then gives a summary of the major results of the project, drawn from across the activities of the project. We discuss the background of research in infrastructure systems engineering, defining the project's scope. We characterise the use of data in this area, focussing on the use of operational data from data suppliers outside of the research community, which distinguishes this field from other areas of research.

We give the main results of the landscaping work of the pilot study. The major benefits which arise from data sharing are categorised into research benefits, benefits to non-research partners, and wider benefits to society. These are illustrated by examples and quotations drawn from the consultation activities and the demonstrator use cases.

The barriers are then discussed and categorised into five themes: legal, security, commercial, cultural and technical. We summarise the major barriers arising in each area, again illustrated by quotations and examples drawn from the consultation activities and the demonstrator use cases.

We briefly summarise the work of the technical design studies, which points the way for further technical development.

Finally, we present our sixteen recommendations, which together can be seen as recommending how to form a research data cloud to support infrastructure systems engineering. We use a typology of elements for research ecosystems to present recommendations. These synthesise the recommendations proposed by activities in the project. We believe that our vision for a future DINI provides a route to realising these recommendations.

There has been a large amount of underpinning work in this project. This report is accompanied by a collection of documents which describe these activities and summarise the results. These are listed at the end of this report and should be consulted to gain further insight on data sharing in information systems engineering and detailed recommendations from each activity.

Some highlights drawn from the range of the activities are presented in boxes through this report. These are representative of the quality of the work; any demonstrator, landscaping activity or design study could be chosen as a highlight of the project.

Table 1 Information on use cases supported by the DINI Pilot Study

Use cases	Project Goal
BRINES (Energy)	Explorations of the use of weather and climate data to highlight future resilience challenges to the UK power network from operational and asset management perspectives.
ClimaTracks Solutions (Rail Transport)	Computation of the risks of weather-related disruptions and asset failures in a railway network. It focussed on occurrence of uncertainty from weather conditions and asset failures and propagation of uncertainty through interdependent components in a network.
D-RES (Energy)	Modelled and assessed the level of distributed resilience and future energy security that can be provided by electric vehicles (EVs) especially during extreme weather events.
Enabling Urban Mobility Data Sharing (e-UMDS) (Urban Transport)	Identified the research questions which could be considered using ride data from urban electric mobility hiring schemes, and the issues which arise from sharing the data with researchers.
ForNET (Energy)	Addressed the use of state-of-the-art forecasting methodologies to enhance the accuracy and robustness of energy demand forecasts, especially in response to extreme weather events and human behaviour patterns.
IMPACT (Road Transport)	Developed an innovative people-centric digital twin (DT) to evaluate the dynamic congestion risks across multiple transportation modes during flooding events.
MARS (Air Transport)	Developed a computational model of the UK airport network and conduct simulations involving airport closures that result in mass diversions. It involved data collection and data pre-processing before the computational model was built.
SOFRAMODE (Water)	Developed and demonstrated a state-of-the-art platform for understanding and simulating urban drainage related to surface water flooding and high-profile storm overflow events, for any UK town or city.
STORMS (Water and Energy)	Developed a risk framework to address the vulnerability of buried infrastructure, like cables and pipes, to meteorological shocks or extreme weather events, such as floods and droughts**.**
The use, barriers and opportunities for Quasi-Real Time data in transport (Road Transport)	Understanding the barriers and opportunities that exist around the use of Quasi-Real-Time data (QRT) in transport infrastructure. QRT is useful for assisting controllers in making medium-term decisions to manage load on the system.
Water Systems Leakage (WSL) (Water)	Use of data to enhance leakage detection in water distribution systems. It identified barriers to data sharing, and proposed solutions to facilitate co-operation between stakeholders.

A Vision for a Data Infrastructure for National Infrastructure

A Data Infrastructure for National Infrastructure should coordinate digital research infrastructure and expertise to:

- **broker** with industry and government for the use of data within research;
- **provide** access to data via a federated infrastructure of trusted repositories;
- **support** a common data interoperability framework;
- **provide** access to trusted computing resources to perform analysis;
- **foster** a network of communities of practice through guidance, training and setting standards;
- and **collaborate** with cross-sector initiatives to build consensus.

Figure 1: A Vision for a Data Infrastructure for National Infrastructure

We have a vision for a future Data Infrastructure for National Infrastructure which would form part of a UK Research Data Cloud, while working with other initiatives in Infrastructure community to provide a bridge to the research community. Such a DINI would complement the emerging concept of the National Data Library.

Broker with industry and government for the use of data by research

A clear message from our pilot study is that the legal and contractual issues surrounding access, sharing and use of data within Infrastructure Systems is complicated and burdensome, and a DINI should work to help ease this barrier for the research sector.

We propose that DINI should work with suppliers in industry and government to negotiate access to reference data with clear data sharing agreements in place for use in research, noting the differences in sector cultural and technical readiness. DINI would work with the research community to identify reference data sets and negotiate access; data suppliers would only need to provide this data once, could handover responsibility for maintaining access to the data to DINI, and be assured that conditions of access will be enforced within a trusted environment.

DINI could provide model Data Sharing Agreements (DSA). However, we would not expect that DINI would provide DSAs for every request for data, as there are likely to be tailored legal requirements for a research project. DINI could provide guidance and advice.

Access to data from a federated infrastructure of trusted repositories.

Our pilot study has shown that access to data suitable for use in research is widely variable. While there are good public sources of baseline data, especially within the energy and water sectors, more detailed, and thus more restricted, data is hard to access, often dependent on personal connections. Negotiating access can then be time-consuming, and it can be burdensome for companies to provide the data in the form the researcher needs.

A DINI should establish a trusted data infrastructure to provide access to data sets from across the Infrastructure Systems sector for use in research. Access can be via means of reference to existing data sources, particularly of open data, or via its own repositories. These trusted repositories should manage and validate their users, where the content isn't open, as necessary to give data suppliers the assurance that the conditions of use are respected. This would give equal access to reference data to researchers in the sector.

A DINI would provide a domain specific place for researchers to publish FAIR data resulting from their research, enhancing impact and stimulating new research, while allowing credit to be attributed to researchers.

DINI should support Data Stewardship expertise for the ISE sector. It should provide Data Stewards to assist data suppliers and researchers in the good management of their data.

Support a common data interoperability framework

The complexity of interoperability between data was frequently raised as a barrier to research in infrastructure systems, especially where data and resources are used across domain and organisational boundaries. The sector is highly diverse with many data formats and standards within and across domains, and it is unrealistic to expect there to be a common approach to data representation. Nevertheless, there is a clear role for a "lingua franca", a set of common standards, formats, tools and processes which can promote interoperability, both by humans and machines, which DINI can contribute to, alongside colleagues working in this area both in industry, public sector and academia.

DINI should support a common data interoperability framework based on profiles of common, metadata standards which should work together to support FAIR data. A common data interoperability framework should not prescribe a rigid standard, but rather accommodate different standards, so that common elements can be drawn together in a harmonised manner, with the ability to draw mappings from one standard to another in the framework.

The Framework should include broad metadata concepts which are applicable across the domain, together with application profiles for infrastructure sectors, APIs for machine access and domain ontologies for semantic annotation. The framework should have a

strong geospatial foundation, and for example reference the UK Geospatial Data Standards Register.

Data quality is also a key issue; the quality of data can be hard to assess to understand its constraints and limitations. This can also be a barrier to releasing data as organisations can be reluctant to release data which they consider having shortcomings. The framework should include an indication of the quality of data.

Providing quality data which is comprehensively annotated and conformant to common interoperability standards can be a time-consuming task which required specialised expertise. Thus, DINI should provide data stewardship expertise to work with researchers and other data suppliers to provide long-lived research products.

Provide access to trusted computing resources to perform analysis.

A DINI should provide access to trusted computing resources. In many situations, DINI will provide access to reference data that researchers will download and use locally. However, DSAs for sensitive or secure data sets may require controlled or limited access and thus require processing using trusted computing resources. Further, DINI should be extended to cover other research assets, particular FAIR access to reference computational models. The Energy modelling community for example recognises the need to form a collection of reference energy models. Modern containerisation techniques can ease the porting of code onto common platforms so that they can be executed in a trusted environment.

We would not expect a single Digital Research Infrastructure to accommodate all the data and computational infrastructure needs of the infrastructure systems research community. Rather it should be provided by a federated network of digital infrastructure within the research ecosystem, potentially including access to commercial cloud. The Climate Resilience Demonstrator (CReDo[3]) programme is demonstrating how this can be managed across organisations, with a foundation in trusted research infrastructure.

A DINI would require the use of different platforms, and it would be highly beneficial to the user community to make co-use of these systems as straightforward as possible.

Foster a network of communities of practice through guidance, training and setting standards.

The importance of community in shaping and supporting changes to practice and standards came through clearly, and new communities of practice need to be created in data management, sharing and publication. Raising awareness of the value of data sharing, the development of data stewardship skills and the sharing of good practice should be a core function of DINI.

[3] Climate Resilience Demonstrator: https://digitaltwinhub.co.uk/climate-resilience-demonstrator-credo/

DINI should include a programme of advocacy and outreach, including workshops, webinars and other activities. DINI should provide guidance on community practices and build a knowledge base of training and guidance materials.

However, the value of data sharing can made clear by examples from practice. The DINI should have relationships with research funders to advocate research data sharing with researchers and their partners. It should work with funders to support research-based case studies with researchers, public sector bodies and commercial organisations to develop best-practices in data sharing and to demonstrate the tangible benefits of re-use of data for research. DINI also should engage to embed best practices with data providers. A possible route would be to embed researchers within organisations to build a collaborative approach, providing data stewardship expertise, which will ease the administrative burden on data providers.

Collaborate with cross-sector initiatives to build consensus.

We recognise that there are many participants in the Infrastructure Systems sector and other initiatives on data sharing. We see DINI as complementary to other initiatives. DINI would form a focus of consensus building, particularly forming a bridge between research, industry and government. It should work closely with other initiatives to foster data accessibility and interoperability within and across sectors. Thus, the governance of DINI should include representation from research and the wider ISE sector, including regulators and industrial aggregators, as well as infrastructure supplier companies.

Our vision acknowledges that there are significant differences between the three main sectors considered in this study. The Energy and Water industries are highly regulated with a relatively small number of supplier companies. Transport is much more diverse, and road transport in particular has fewer common approaches to data. However, as many of the practical problems to be solved in infrastructure are cross-sector, DINI should work across sectors to encourage the transfer of best practice and the co-design of solutions.

Highlight: Workshops in Wales, Scotland and Northern Ireland

DINI engaged with national infrastructure experts in academia, government, and industry at workshops held in Swansea, Glasgow and Belfast, facilitated by the Digital Curation Centre. Over the course of these workshops 50 participants shared their views and experiences on current barriers, challenges, opportunities, and solutions around data sharing in their field.

Benefits of data sharing fell broadly into eight themes:

- Quality assurance and transparency
- Robustness of analyses
- Facilitating innovation
- Increasing efficiency
- Improving decision making and emergency response
- Facilitating systems thinking
- Raising public awareness
- Helping researchers' careers

Commercial and cultural barriers were the most discussed during the workshops in particular the **economics of data sharing** and the **inefficiency in the process of finalising data sharing agreements** e.g. *"taking months", "taking at least 9 months"* the lengthy process is particularly difficult for short-term projects.

Differences were observed reflecting the different regulatory environments in the devolved nations. For example, Wales has a framework for public and private service providers sharing personal information (the Wales Accord on the Sharing of Personal Information, WASPI) that provides disclosure agreements or information sharing agreements guidance and templates which may account for the relatively low occurrence of data sharing agreements as a barrier in the Welsh workshop compared to the Scotland and Northern Ireland workshops. Another example of specific barriers was the availability and coverage of data in Northern Ireland as it is not always covered in UK national statistics or datasets, while cross-border interactions with the Irish Republic can also impact data coverage.

Figure 2: Highlight: Workshops in Wales, Scotland and Northern Ireland

Background

Defining the Infrastructure Systems Engineering Domain

Infrastructure can be defined as "*the physical components of interrelated systems providing commodities and services essential to enable, sustain, or enhance societal living conditions*" (Fulmer 2009). The UK National Infrastructure Commission's remit covers "*economic infrastructure, defined as: energy, transport, water and wastewater (drainage and sewerage), waste, flood risk management and digital communications.*" This is a very broad area, and DINI has focussed on **energy supply and distribution, water supply and disposal** and **transport networks,** and aimed to be representative rather than comprehensive.

The branch of Engineering Science which covers these domains can be called Infrastructure Systems Engineering (ISE) which has some different characteristics from other branches of traditional Engineering Science. In ISE research we are primarily interested in questions of how components interact with each other and with the natural and human environment. These questions are key to assessing the societal impacts of infrastructure, and therefore of prime interest of government to achieve policy goals.

Using Data in Infrastructure Systems Engineering Research

We observed some features in the use of data that are characteristic of the ISE research community which have influenced our approach.

Highly interdisciplinary. ISE research is inherently interdisciplinary, and many research questions consider the interactions between different infrastructure systems. Further, the interaction with other domains, such as environmental sciences, sociology and economics, and health science must also be considered to get a complete picture of the impact of infrastructure systems. This need to accommodate a wide range of domains means that data integration is complex.

Wide variety of data sources. The interdisciplinary nature of ISE research means there is a high diversity of potential data sources. Even within one infrastructure sector there are many data suppliers.

Operational data. Much of the source data used in research comes from industry or government rather than the results of previous research. Much data is operational in nature, describing the state and performance of infrastructure. Consequently, access to this data is dependent on whether government and industry stakeholders are willing or able to release data.

Sharing with Partners. Much of the impact arising from ISE research arises by sharing results with impact partners in government and industry, as well as with other researchers. The need to share data and results with other partners may influence the conditions that data suppliers impose on the use of the data.

Sensitive data. Some data is personal sensitive, particularly, when considering usage of infrastructure. Road traffic data may include routes of individuals, or energy consumption may be given at a household level. Further, data which originates from industry may contain business sensitive information kept confidential for commercial advantage. Some data can come under the category of Critical National Infrastructure[4] and could be subject to National Security. Thus, ISE data can have complex restrictions that need to be respected and when data is made available it needs to be within a trust relationship with the researcher.

Wide diversity in data representation. Data in the ISE sector comes in a variety of data and metadata formats and semantics, with few universal standards, although there are many standardisation initiatives within domains. Data has an underpinning Geospatial layer, as the distribution, interconnections and impacts of infrastructure assets across regions and communities of central concern.

The ISE sector long been an aware of the issues of data sharing, and there have been studies prior to the DINI pilot. Few have focussed on the research sector, but many of the issues raised are in common. A report commissioned by the NIC in 2017 (National Infrastructure Commission, 2017) forms a useful historical baseline for the DINI study. This report concluded by identifying five types of barriers to infrastructure data sharing which we have used as framing for this report.

In parallel with the DINI study, the Digital Twin Hub Data Sharing Working Group (2024) has surveyed UK Data Sharing Infrastructure Data Sharing Initiatives (High-level Landscape Snapshot, October 2024). We feel that our work is strongly complementary to this Working Group. The publication of FAIR data has been advocated by funders at national[5] and international level[6] . There have been surveys to assess adoption of FAIR data (Hahnel et al. 2024), however, evidence in ISE research is sparse and anecdotal. DINI did not systematically survey the adoption of FAIR data practice in ISE.

[4] National Protective Security Authority, Critical National Infrastructure https://www.npsa.gov.uk/critical-national-infrastructure-0

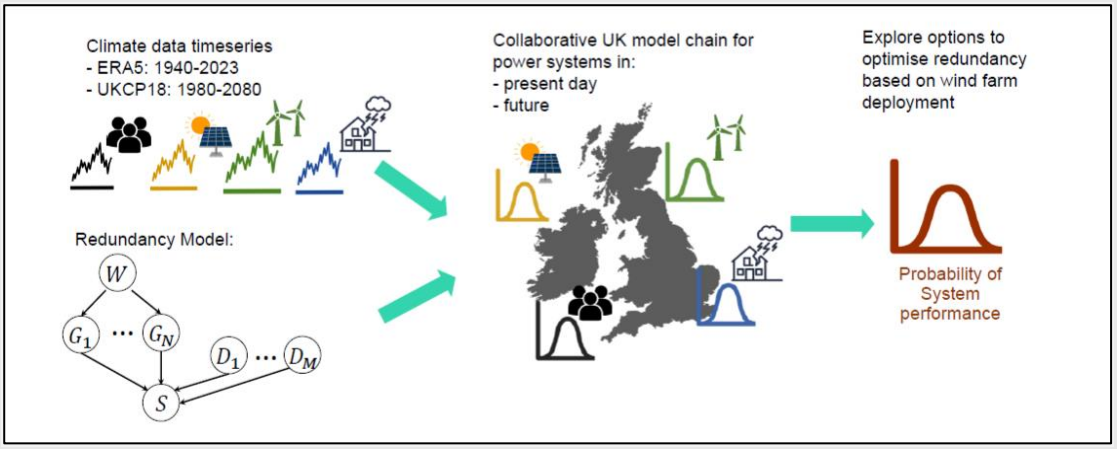
[5] See for example https://www.ukri.org/who-we-are/epsrc/our-policies-and-standards/policy-framework-on-research-data/

[6] e.g. in the European Commission research programmes: https://research-and-innovation.ec.europa.eu/strategy/strategy-research-and-innovation/our-digital-future/open-science/open-access_en

Highlight: Building Risk Informed redundancy for Net-zero Energy Systems (BRINES)

Dr Hannah Bloomfield, Dr Colin Manning, Dr Ji-Eun Byun, Prof Sean Wilkinson

The project explores the use of weather and climate data to highlight future resilience challenges to the UK power network from both an operational perspective (maintaining the balance of supply and demand) and from an asset management perspective (making sure assets are not damaged by extreme weather). It intends to harness the collated meteorological data for probabilistic modelling, thereby assessing the optimal design of system redundancy (i.e. construction of generation assets more than required to be armoured against high consequence events) from a risk-based perspective.



- **Barriers of Data Sharing**
  The report discusses various barriers around open access, licensing (privacy, legal and commercial barriers), data download limitations (security) and lack of appropriate skills (cultural barriers).

- **Benefits of Data Sharing**
  The main benefits identified were the use of DAFNI platform for removal of methodological challenges, stakeholder collaborations, stress testing and the use of data sharing as an educational resource.

BRINES has made state of the art weather and climate data more accessible to energy sector stakeholders for detailed system analysis

Figure 3: Highlight: BRINES Energy Use case

Benefits of Data Sharing

Data sharing in the ISE sector has the potential for significant socio-economic and environmental benefits. As part of this study, we captured some the most important benefits from data sharing that the community perceived.

Figure 4: Benefits of Data Sharing

We have observed that three categories of benefits emerged from the landscape analyses, as illustrated in *Figure 4*; immediate benefits of enabling research and engagement; downstream effects on stakeholders using the research results; and broader impacts on policy and planning for the UK.

Research Benefits

The immediate benefits of data sharing are seen as being on the increased potential for research, collaboration and engagement.

Enable research and innovation

Access to real-world operational data from across the ISE sector is seen as critical to meaningful research time and again by the research demonstrators, where access to these datasets is crucial to their success. This supports our view of the ISE research community's use of operational data, with access to data sets driving the research effort to advance knowledge and develop innovative solutions. Availability of data can drive new research questions and collaboration, as emphasised in the Urban Observatory Workshop; e-UMDS derived a list of possible research projects that might derive benefit and insight from the data with more than 30 examples to show the range of potential research questioned which could be explored by unlocking a data set.

"The input datasets of the D-RES Project include operation of the chargers by the EV users (i.e. energy timeseries), location of the chargers. The collection of the aforementioned data would require many years, hence, without public datasets, D-RES would not be able to deliver its outputs" D-RES

"The new datasets are crucial for us to extend our model to new networks, i.e., Cadent Gas network and validate our model against historical records, which is essential for ensuring the credibility of the climate risk assessment model" STORMS

Enhance Collaboration and interoperability

A further benefit is in enhanced research collaboration among government, industry, and academic stakeholders, to promote innovation. Data resources form a focus for research, with groups with different skills coming together to form new insights on the shared data. This can unlock the potential of data which may not have been seen before, especially when the data supplier does not necessarily have the time and resources to exploit the data. e-UMDS works with local authorities who may not have the expertise and resources to explore the potential insights arising from the data they collect. In these situations, sharing with research institutions can be an effective way of exploiting the data.

This is particularly pertinent when the data is combined from different source and across sectors. e-UMDS goes on to observe the extra value to be gained from combining different data sources, and in IMPACT highlights the opportunities of data sharing across sectors. This is borne out in the Climate Research Demonstrator (CReDo) project, which is pioneering the development of cross-sector digital twins.

"Coming together in a geographic space such as an urban observatory can lead to new partnerships. Bringing data from multiple sources together into a single computational structure enables multidisciplinary research" UKCRIC, Urban Observatory Workshop

"extra value can often be added through the merging of data, which can be challenging for a regional authority to achieve" e-UMDS

Foster Transparency and trust

Companies can see that working with researchers aligns with their company values and fosters research community engagement. Further, releasing data for scrutiny to outside bodies and can increase transparency and accountability. This was widely recognised across the study and a motivating factor in the actions of regulators to release data; it is

noticeable that the more strongly regulated sectors of water and energy made data available for sharing more readily via platforms like Stream.[7] than in the more diverse transport sector. For example, in the water industry trust can be gained by being open about water leakage, as noted in the WSL study. This can be extended to the use of data in research in generating new models and results; if they can also be made open, with suitable auditing and provenance, the overall trust in result can be enhanced, and this can be reflected into future planning decisions which take those results into account.

"open data sharing increases accountability and demonstrates utilities' efforts to address water loss" WSL

Effects on Project Stakeholders

The benefits of enabling better research collaboration and engagement can lead to positive effects on research partners and stakeholders which can consume the output of the enhanced research.

Support policy development and evidence-based planning decisions

---

The potential benefits identified by the Urban ride Research project have mostly been in form of policy insights" e-UMDS

"Transparent access to leakage data informs evidence-based decision-making and contributes to sustainable development goals such as reducing carbon emissions" WSL

Researchers also see the potential impacts and emphasise the predictive power of their research to provide insights into infrastructure performance, vulnerability and impact on people and the environment. This can support evidence-based decisions by industry and government, and the impact on policy is emphasised. For example, the Water Systems Leakage (WSL) project identified the potential for better data contributing to carbon-reduction targets, IMPACT noted the potential impact on the better planning of road infrastructure, while SOFRAMODE is contributing to Newcastle City Council's Blue-Green Initiative to enable better urban planning.

[7] Stream. water data portal https://www.streamwaterdata.co.uk

"Shared datasets enable better modelling of flood impacts on infrastructure systems, leading to more informed decision-making for resilience planning" IMPACT

"More robust decision making e.g. awareness of new policies and plans for urban areas so that interventions complement upcoming developments" SOFRAMODE

It is also noted that data generated within research projects can also have impacts, as in BRINES. Thus, data sharing, and in particular data sharing for research purposes, enhances public safety and mobility by improving infrastructure resilience, benefiting society at large.

"The datasets created through BRINES will be useful for energy sector stakeholders wishing to understand the importance of climate variability and climate change on both security of supply … or times where traditional extreme weather may cause weather-induced failures" BRINES

Improve accuracy and validation

The availability of data for research also facilitates the validation of research outputs. These resulting models can act as reference benchmarks as for example in FORNET who expect that their results can be used to compare and evaluate other approaches to modelling energy consumer behaviour. As WSL noted, sharing data enables researchers to validate models using real-world scenarios.

"[used] data from Newcastle City Council and crowd-sourced data from the general public to validate model outputs" SOFRAMODE

"Cadent Gas dataset shared in way has helped us to extend the model's use cases and validate the model" STORMS

"academics and practitioners can evaluate the performance of their forecasting approaches versus the FORNET machine learning method" FORNET

Improve operational efficiency and monitoring

A further impact of research is improving the economic efficiency and competitiveness of organisations, by evaluating different planning scenarios and by developing novel approaches to systems operation and maintenance. Benefits can be of a direct

operational nature, as in improved water leakage detection and or the development of near real-time monitoring tools for flood events.

These can be general effects such as reducing duplication and better resource allocation; stakeholders need to evaluate the results carefully to assess the real impact. But it can lead direct cost savings, with one company saying that relevant research applied to predict traffic and air quality saved them the costs of engaging with a consultancy, which had estimated value of £20k–£50k (IceBreaker One).

Policy Impacts

Data sharing in ISE supports key public policy goals of a secure and resilient society, economic growth, and environmentally sustainable development.

Impact on Society

The benefits of research to society are well understood, and this is recognised by participants in the Industry Consultation Report produced by IceBreaker One, where the benefits of sharing data with researchers were seen as an increase in collaborative projects for innovation development, which could feed into policy and operations, to deliver better outcomes for society. Benefits to society can include protecting vulnerable customers, while increasing the skills of the UK's workforce.

We highlight two impacts on society: increasing the security and resilience of society in the face of extreme events; and the increase in engagement by citizens in decisions.

"Survey respondents highlighted data sharing has advanced the field of research through: … contributing significantly to making vehicles cleaner and safer; and upskilling students in research areas, which provides broader societal benefits" IceBreakerOne

"Sharing data with research organisations can also produce benefits to vulnerable customers through initiatives such as the water industry initiative, Support for All" IceBreakerOne

---

Resilience in the face of short-term disruption, such as extreme weather, has been a major theme of the demonstrators (for example, ForNet, IMPACT, ClimaTracks, STORMS) in the project. The additional resilience their results can bring to the planning and preparation process for the emergency response to such events can mitigate their effects, potentially reducing harm to people and property, as well as reducing the disruption to people's lives. Longer term planning in the face of long-term threats such as climate change can pre-empt the disruption to society, as in BRINES. Data sharing

initiatives such as the National Underground Asset Register (NUAR)[8] have the potential to reduce disruption to citizens; by opening the NUAR to researchers, new insights can be gained, especially when combined with other data, such geological data as in STORMS.

"The datasets created through BRINES will be useful for energy sector stakeholders wishing to understand the importance of climate variability and climate change on both security of supply … or times where traditional extreme weather may cause weather-induced failures" BRINES

Data sharing can also increase the engagement of citizens. We have also noted the confidence that openness about leakage data could bring. We also see the engagement of local councils in Bristol and Newcastle who are willing to work with researchers as a way of pursuing their policy goals. The London Datastore[9] publishes a wide range of data on the city and in its blogs tells the story of how the data gives insights into the life of the city.

Impact on the UK Economy

Sharing data with researchers can generate broader economic benefits to the UK. We have already discussed the increased resilience in the face of extreme events; this clearly has a significant economic as well as societal benefit. Further economic benefits accrue by enabling efficiencies and advancing innovation and opportunities in the UK.

Time, resource, and cost efficiencies can be demonstrated through initiatives such as NUAR project, and the use of QRT to manage transport infrastructure more efficiently. The use of data science models can result in financial savings, efficiencies, or reduced incidents, though evidence may not always be tangible.

Aiding innovation in the UK includes improving numerical modelling tools, which aids sectors like offshore wind and tidal stream projects and contributes to the UK energy supply; advancing innovation programmes through Innovate UK, which increases revenue and tax income and contributes to the economy; and upskilling students in analytic skills, which increases the workforce in this specialist field and supports long-term economic growth.

"The economic benefits of flood estimation data include time savings (£0.3m–£1.3m per year) and reduced construction costs and flood damages (£8m–£30m per year at 2006 prices, or £11.2m–£42.2m at 2021 prices). The

---

[8] National Underground Asset Register (NUAR) https://www.gov.uk/guidance/national-underground-asset-register-nuar

[9] London Datastore, Greater London Authority https://data.london.gov.uk/

reduction in flood damages due to early warning systems is estimated at £86m–£145m annually (2021 prices)" IceBreaker One

"the National Underground Asset Register (NUAR) project which estimates £400m annual savings through increased efficiency of data sharing and excavations, fewer accidental strikes on underground pipes and cables, and reduced disruptions for the public and businesses for Northern Ireland, England and Wales" IceBreaker One

"A digital twin uses a mixture of real time and quasi real time data to simulate the status of a networked transport infrastructure, allowing targeted interventions (such as the use of variable speed limits, diversions and road closures) to manage congestion and improve outcomes for the infrastructure users" QRT

Impact on the Environment

Research into ISE can contribute to understanding the environmental impact of infrastructure developments, making them clear and providing mitigation pathways. Further, ISE research can analyse the pathways to Net Zero so that it can be achieved in a robust and efficient manner. Thus, the project BRINES and DRES are both considering different aspects of how a renewal-energy based electricity generation network might be resilient to extreme weather, including taking into account the increased use of electric vehicles and their charging infrastructure.

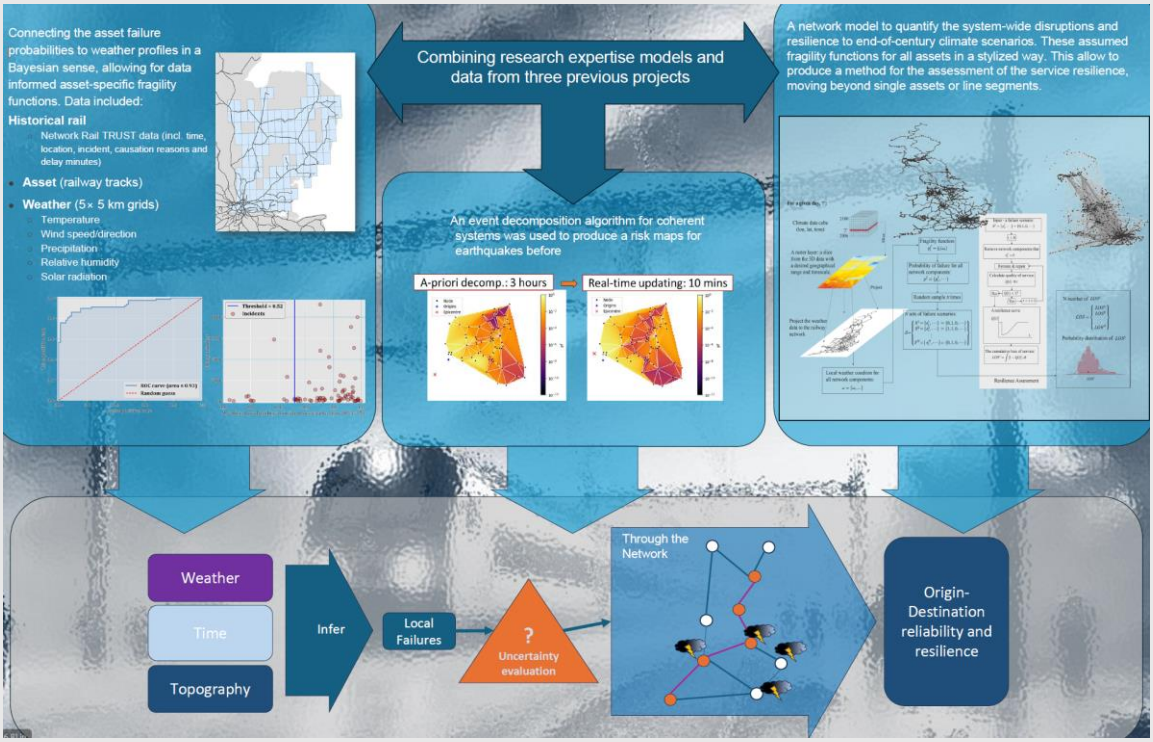
Other research discussed in project interviews and surveys can lead to advances in vehicle design and testing/modelling methodologies, including efforts to reduce transport emissions and investigate brake and tyre particulates. Examples include monitoring of impacts from initiatives such as the Birmingham Clean Air Zone, and some data science models have contributed to reduced pollution incidents and financial savings, though tangible evidence is limited.

6 of 23 respondents noted possible applications of research to improving air quality, energy planning, and developing holistic approaches to climate.

Highlight: ClimaTracks - Forecasting resilience of railway network under propagating uncertainty

Giuliano Punzo, Ji-Eun Byun, Qian Fu, Iryna Yevseyeva, Tohid Erfani, Konstantinos Nikolopoulos

The project aims to compute risks of weather-related disruptions and asset failures in a railway network. It focuses on occurrence of uncertainty from weather conditions and asset failures and propagation of uncertainty through interdependent components in a network. Through this the project produces resilience maps for the UK railway network.



- **Barriers of Data Sharing**

  The report identified barriers in legacy data sharing agreements (commercial barrier), computational architecture (technical barrier) and data literacy (cultural barrier).

- **Benefits of Data Sharing**

  The report discusses the benefit towards project development provided there are data sharing agreements in place. It also discusses the benefit to model training through data availability and standardisation and reuse of data results by industrial stakeholders.

Figure 5: Highlight: Use case CLIMATRACKS Transport

Barriers to Data Sharing

In the DINI pilot study, a key question was to identify what are seen as the current barriers to sharing data, especially sharing for the purpose of research. In the evidence collected, many individual barriers were identified, and details of these can be found in the activity reports given as supporting information. We summarise the NIC barrier areas as:

- **Privacy and legal barriers.** The legal, regulatory and contractual concerns which arise around the sharing of infrastructure data. This applies to data containing personal information, commercially sensitive or national security critical data.
- **Security barriers**: The risk that sharing data may lead to security breaches, data losses and, in extreme instances, high-impact cyber-attacks. This may lead to damage to systems and sensitive data being accessed by unauthorised actors.
- **Commercial barriers**: The costs of sharing are perceived to be greater than the expected benefits, and so the sharing data may be seen as a loss of competitive advantage.
- **Cultural barriers**: The habits, policies and attitudes within organisations that hinder or oppose data sharing.
- **Technical barriers**: The challenges in sharing the data from a technical perspective, such as non-standard formats, legacy systems, poor documentation.

The literature review concluded that the nature of the barriers has changed little from 2017, indicating that there is still much work to be done to address these challenges. Future research could potentially benefit from further exploration of the range of sociotechnical barriers to achieving change. We highlight the key barriers affecting the use of research data in each area.

Figure 6 illustrates the types of major barriers highlighted in the reports, categorised into the five areas outlined above.

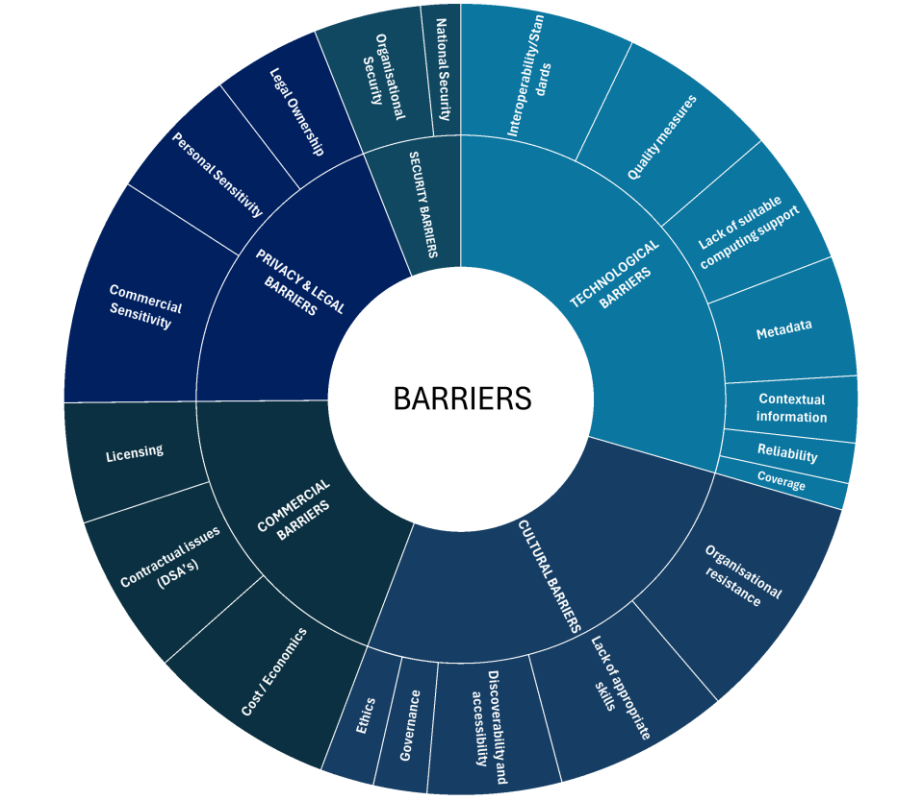


Figure 6: The major barriers to sharing data

Privacy and Legal Barriers

Legal and contractual concerns have a significant impact on data sharing and have been widely raised from across the DINI landscaping activities. These can manifest in many forms, as concerns about the legislative and regulatory framework, especially in the more regulated sectors, the overheads and complexities of negotiating Data Sharing Agreements, and the requirement to respect privacy of individuals when handling personally sensitive data. These concerns dominate the top barriers identified by organisation type within the IceBreaker One Landscape report, reproduced in *Table 2* and they go on to say that these concerns are shared across the water and energy sectors in particular.

The exception appears to be the academic research community itself; however, the DINI case studies demonstrate that the overhead of establishing legal agreements is a major impediment to research projects.

Table 2 Barriers by organisation type (from IceBreaker One Industry Consultation Report)

Organisation type	Barrier 1	Barrier 2	Barrier 3
Academia/publicly funded research body	The cost (in terms of time, money and other resources) of getting the data ready for sharing	Lack of organisational motivation to share data	Data will be used in a way that risks reputational damage to the organisation
Arms-length public body	Concerns about data protection	Data being commercially sensitive	Data will be used in a way that risks reputational damage to the organisation
Commercial - other (e.g. data consultancy)	Data being commercially sensitive	Other legal concerns (e.g. over intellectual property rights)	Concerns about data protection
Commercial - water, energy or transport service supplier	Concerns about data protection	Data being commercially sensitive	Data will be used in a way that risks reputational damage to the organisation
Government - central/national (including devolved government)	Sharing the data could pose security risks	Data being commercially sensitive	Concerns about data protection
Government - local	Data will be used in a way that risks reputational damage to the organisation	Data being commercially sensitive	Data will be used in a way that does not align with the interests of the organisation
Regulator	Data is not good enough quality to share	Sharing the data could pose security risks	Concerns about data protection
Trade or professional body	Data being commercially sensitive	Data will be used in a way that does not align with the interests of the organisation	Lack of organisational motivation to share data

"Aviation is a highly competitive sector, where data sharing and achieving better shared situational awareness have consistently been key focus areas … These goals, however, have historically faced significant challenges due to the commercial sensitivity of the data and legal complexities surrounding ownership." MARS

"large amounts of time to go into setting up agreements to access data … [contractual issues have] delayed our project significantly. … The biggest barrier that we have faced is the delay of the signing of the contract." e-UMDS

Security Barriers

Security barriers were less frequently raised overall, as seen in *Figure 6*. Nevertheless, there is clearly an underlying recognition that there is a need for safe and secure handling of data, again clearly reflected in the prioritisation table above. In addition to the direct security barriers raised of commercial confidentiality, cybersecurity protection and national security, we can also infer that many of the concerns couched as legal barriers are driven by the security risks of sharing data.

Security is a complex concept that depends on the risks being considered. Two key areas of threat which actively inform practices of research data sharing in infrastructures: cyber threats, by unauthorised actors, which might damage the data supplier in the former and society in general in the latter case. Thus, there is understandable nervousness in releasing data into the relatively uncontrolled research environment.

There is a tension here. There is a recognition that openness in data is valuable in raising public trust through transparency, with regulators in particular understanding that not sharing is a barrier to a positive public reaction to change. And openness also drives the benefits in use in research, particularly in interdisciplinary scenarios. The mantra of the FAIR data principles should be applied here: as open as possible, as closed as necessary.

"Utility companies, understandably, are cautious about sharing their infrastructure data due to concerns over security, misuse, and operational sensitivity. For example, they are reluctant to expose the accurate location of assets and their safety-related operational records. This protectiveness creates a barrier to open collaboration." STORMS

Commercial Barriers

Many of the commercial barriers highlighted were also closely associated with legal and contractual barriers, sharing data always has a cost to the supplier organisation, particularly the labour costs involved. Some organisations may have the skills and people to make standard data products available, which may be useful internally as well as externally, and may also be required to satisfy regulatory requirements. But other organisations – particularly small companies and local authorities - may find the costs involved excessively burdensome and may lack the skilled staff altogether to provide well-described data.

"Data sharing places a burden on those sharing the data that includes time and financial costs. Simply handing over data is not sufficient for data sharing. The format, quality, context, provenance, limitations and appropriateness of the data must be put into forms that can be received and understood by the user, and the user must take the time to understand them." UKCRIC Data Sharing Workshops

A further issue raised several times was the data requests from researchers were costly to fulfil due to being unclear, being in large number or requesting complex or non-standard views of data. Fulfilling data requests may not be seen as a priority immediate business need justifying the time, effort and cost involved, even if there is a willingness in principle to support research.

"The costs associated with acquiring, processing, and sharing traffic data may outweigh perceived benefits for some stakeholders" IMPACT

Cultural Barriers

Cultural barriers form a diverse set of issues. These include responsibilities, policies, attitudes and skills in both the research and the partner organisations. There may be low awareness of the benefits of data sharing, and thus the costs may be seen as being too great.

Risk aversion may also lead to a reluctance to release data. Releasing poor quality or partial data may be seen as reflecting on poor record keeping practices of the organisation, while data which indicates poor organisational performance may lead to significant reputational damage. Organisations may not always have clear policies on how to assess these risks and who has the responsibility for taking the decision to share data.

"Reuse of the data could expose problems or errors in the data (and in the methods, including data collection, data processing, etc.) - and this could cause reputational damage to the researcher or provider of the data as well as to their organisation" DAFNI Devolved Nations Workshop

In our study we have met keen champions of data sharing in commercial and governmental organisations, who are seeking to drive change. However, those attitudes may not be universally shared in their organisation. Researchers that interact with such champions may be able to negotiate privileged access to data, not available to other researchers, creating an unfairness in the research landscape.

Technical Barriers

Overall, the non-technical issues discussed above form the major barriers to data sharing, with the technical barriers seen as being of less as an absolute blocker to progress. Data integration can be achieved with sufficient effort involved. Nevertheless, technical barriers can be extremely challenging: the conceptual and practical incompatibility of data can be complex and requires considerable effort, time and resource to resolve. Technical barriers formed nearly a third of the total raised, and there remain significant challenges to overcome if data sharing and integration is to become seamless and routine.

"People collect and think about data in different ways, so the same kind of data may be recorded differently." UKERC/EDC -DINI Workshop

---

Data discoverability and good metadata provision were identified as two key areas in which research data sharing is not always working well. Researchers are not always aware of the data that is available, and once it has been discovered how it can be accessed and reused. Poor metadata provision and data description make it hard to understand and reuse data.

Several project activities highlighted a lack of standardisation, including data formats and terminologies, can further impede dataset integration and interoperability, and this issue is multiplied when projects require data from across infrastructure sectors.

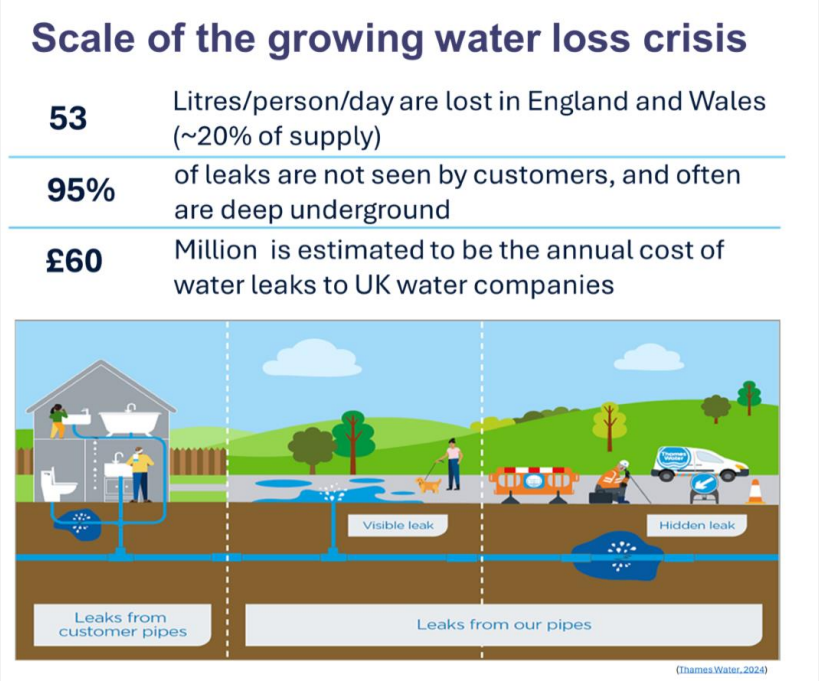
Reproducing a model is not only laborious; it is unlikely to yield fully comparable results when carried out by two different parties, with the most probable cause due to differing data sources/versions. This creates a potential for lack of trust, which may be inferred as reliability issues. In general, many energy systems models, which result in rich and diverse forms of output to support widespread research innovation, lack standardisation leading to interoperability issues. These barriers affect Researchers, Data Users, Data Aggregators, Innovation / Research Agencies. D-RES

A further significant barrier which was raised several times is the lack of indication of the quality of data. This is of particular concern to researchers who need to be able to assess the quality data to assess the value of their subsequent analysis. Common indicators of data quality should be provided with the data.

Highlight: Water Systems Leakage (WSL)

Ruoqing Yin, Haonan Xu, Jiaqian Wei, Liz Varga

The project aims to use data to enhance leakage detection in water distribution systems, to identify barriers to data sharing, and to propose solutions to facilitate co-operation between stakeholders. The research focuses on the importance of consistent data collection and standardisation, creation of Trusted Research Environments and collaborative platforms for data confidentiality and access to infrastructure data for faster detection and repair of leaks.



- **Barriers of Data Sharing**
  The project identified Cultural barriers such as discoverability, resistance and ethics and technical barriers such as Reliability, Interoperability standards, metadata as some of the main barriers to data sharing.

- **Benefits of Data Sharing**
  Building public trust, enabling policy support and research collaboration and improving operational efficiency and scalability were some of the main benefits identified by the research.

UCL Civil, Environmental and Geomatic Engineering

Figure 7: Highlight: Use case WSL Water

Technical Design Studies

The DINI pilot has undertaken some initial design studies to explore how we might realise components of the DINI vision. These studies were chosen as they represent key technological challenges to the development of an interoperable data landscape, as summarised in *Figure 8*.

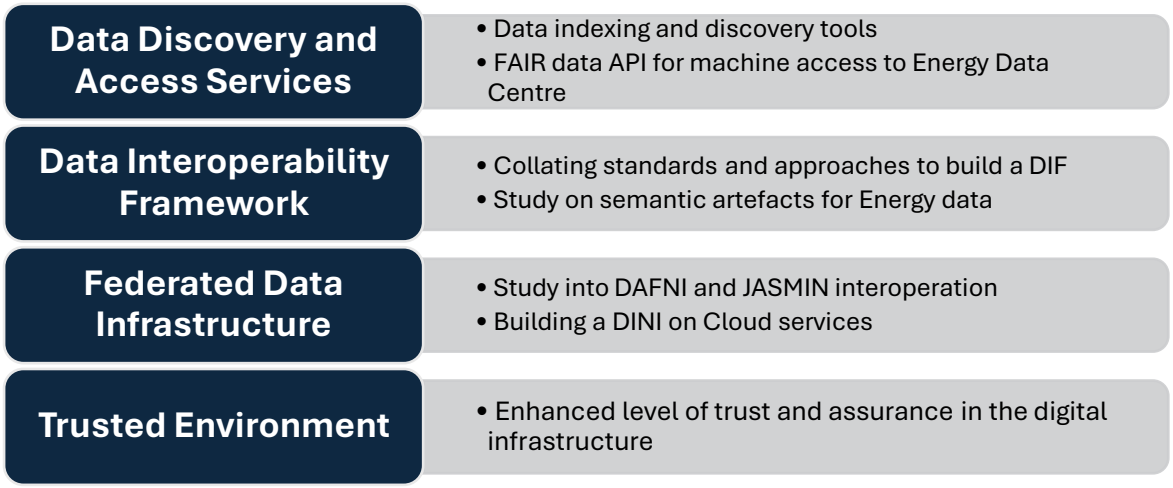


Figure 8: How the technical studies map to the DINI components

The reports from the individual technical studies are available as outputs from this study; we summarise the main points of each here.

Data Discovery and Access Services

Data discovery and access were identified as key barriers to data sharing, and we would anticipate that a future DINI would provide services to enable researchers to find data which might be valuable to their research, allows them to explore that data in further detail and access the data with suitable access conditions.

Data is distributed across many suppliers and stored in multiple places designed for different purposes. It would be inappropriate to attempt to centralize such data within a single data store. However, a service directing researchers to data would be of great value. Data indexing services, which catalogue data sets across the sector and provide redirection and data access services, would form the basis for such discovery and access services. The Data Indexing study considers the requirements for data indexing in the DINI. It outlines the approach which the DAFNI platform uses to index data, including search and discovery services and a common metadata standard based on the DCAT2 standard[10]. To index data sets within the community, this service would need to be

[10] Data Catalog Vocabulary (DCAT) - Version 2, 4 February 2020, W3C Recommendation, https://www.w3.org/TR/vocab-dcat-2/

extended with richer metadata and subject classifications, and APIs to external data repositories, which are not standard within the ISE community.

In recognition of the need to promote standard APIs to repositories, we undertook a case study on APIs within the Energy Data Centre. The API development focussed on what metadata should be provided through an API to establish the FAIR-ness of data and the trustworthiness of the repository. Recommendations for future development include creating a FAIR Implementation Profile for this domain and examining how policy and terms and conditions can be made more machine actionable.

Data Interoperability Framework

Common semantic meaning plays a central role in data interoperability, both with and across domains, sharing understanding of concepts by people and machines. We reviewed the landscape in the energy sector looking at semantic artefacts, discussed the benefits and challenges of applying subject descriptions in a production environment within the Energy Data Centre portal. This work recommended that DINI could act as a catalyst to bring together interested parties on the topic of interoperability of semantic meaning in the ISE and related domains.

Federated Data Infrastructure

We do not anticipate that in a future DINI there will be a single data and compute platform to hold the data required for research. Rather, the DINI would be formed from a distributed set of data services and computational platforms, with different content and different capabilities. Existing and new platforms should be federated to work together, to allow users to access data resources and compute across the federation.

In the DINI pilot study, we have explored how a federated infrastructure might work in practice. The existing platforms, DAFNI and JASMIN have explored how they could interoperate technically. Key to this is having a suitable authentication, authorisation infrastructure to make access seamless for the users of both DAFNI and JASMIN.

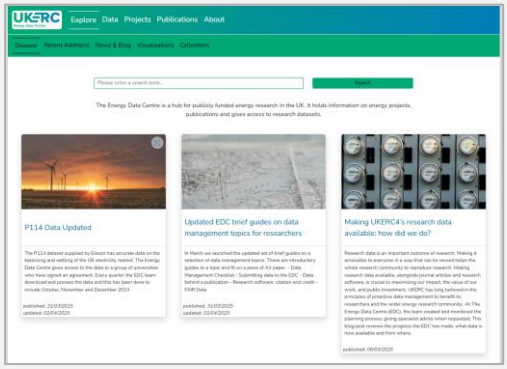
Trusted Environments

The need for Trusted Environments to support digital research in ISE has emerged during the course of the DINI study and has been identified by several of our use cases. As there is a significant body of work particularly in the health and social sciences on supporting trusted research environments (TREs), including their federation, notably in the DARE-UK programme and other Research Data Cloud Pilots (e.g. SOURSD and EDS-Boost), DINI has not addressed this topic. We note that a future DINI programme should work with these initiatives to determine how to most effectively support TREs for the ISE community.

Highlight: Raising the FAIR-ness of Data

Catherine Jones, Peter Holt, Karen VanHaltren, Teagan Zoldoske (STFC)

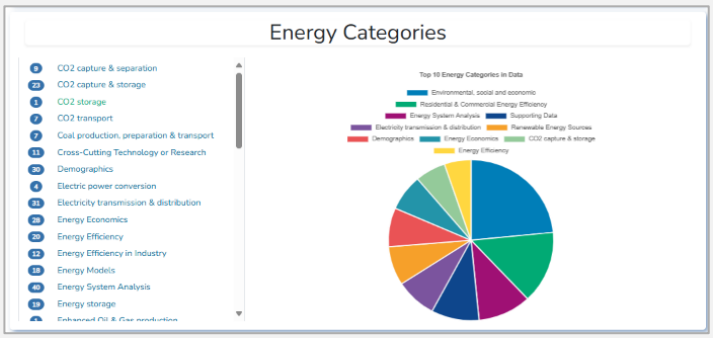
The DINI vision proposes a network of machine-accessible trusted, FAIR repositories with a common understanding of meaning. The Energy Data Centre is a hub for publicly funded energy research in the UK, giving access to projects, publications and research datasets. The EDC explored how to make its resources more accessible, and to support the long-term application of richer annotation of its resources to encourage reuse.



Energy Data Centre Portal

A FAIR-enabled API. The EDC specified a FAIR API using OPENAPI to enable machine access to the content, demonstrating trustworthiness of the repository, and sharing FAIR metrics for specific datasets, and investigating their benefits and challenges. Enabling users to evaluate the FAIR-ness of data may reduce the barriers to data sharing.

Semantic artefacts: The EDC reviewed the challenges of applying semantic annotation within a data repository. It showed that there is a need for consistent, agreed concepts to aid discovery and interoperability of energy research data content and noted the issues of consistent application. There is much work on ontologies in energy research and development, but there is no overall co-ordination at the top-level of the terminology. Being able to map terms between well-established and well used schemes will become more important in a future DINI and will need to consider the granularity and precision required to provide meaningful annotation based on the underpinning repositories.



Use of subjects for data discovery in EDC portal

UKERC Energy Data Centre

UKRI Science and Technology Facilities Council

Figure 9: Highlight: EDC's Technical work on raising the FAIRness of data

Recommendations

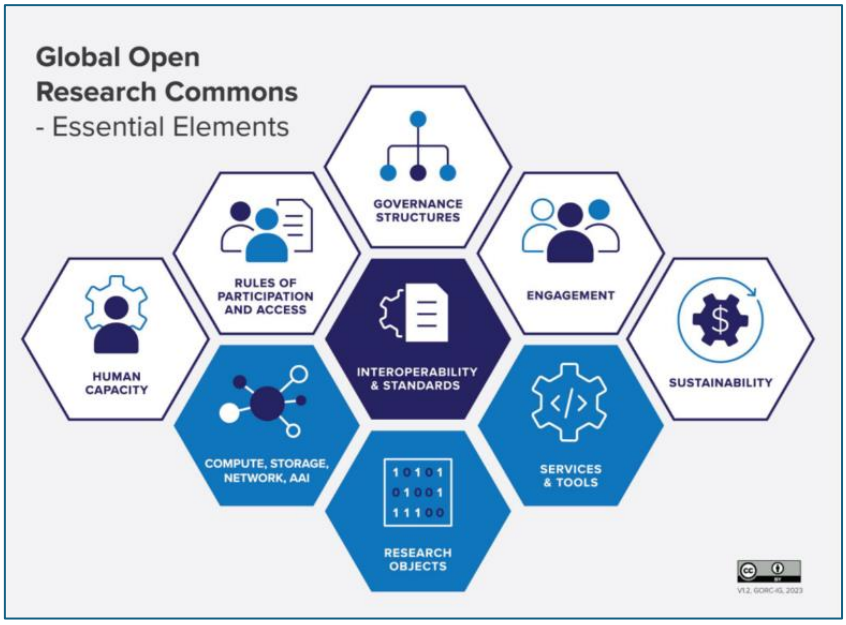


Figure 10: GORC: Typology of the essential elements in a Data Commons (Jones et. al. 2023)

The DINI pilot study has investigated the benefits and barriers for ISE researchers using and sharing research data in the context of a Research Data Cloud. This has led to our recommendations to support data sharing for research in ISE and our vision of a future DINI. To organise our recommendations, we have used the framework developed by the Research Data Alliance Interest Group on a Global Open Research Commons which defined the essential elements for a research data common (analogous to a Research Data Cloud), as illustrated in *Figure 10*. We present recommendations to support each of these elements.

Governance structures

Good data governance is at the heart of providing a principled data infrastructure and must be based on the principles of transparency, accountability, engagement, and responsiveness to address the legal, policy, security, and communications needs in addition to technical matters.

Together, DINI could form a focus of consensus building within the ISE domain, particularly forming a bridge between research, industry and government, and work closely with other wider initiatives in the ISE domain to foster data accessibility and interoperability within and across sectors. To support this, the governance of DINI itself should include wide representation from across the research sector and across the wider ISE sector, including regulators and industrial aggregators, as well as infrastructure supplier companies.

1. **Data providers and policy makers must invest in co-designing appropriate and open governance for data sharing within and across sectors.**

Rules of participation and access

A clear message from our pilot study was that the legal and contractual issues surrounding access, sharing and use of data within ISE is complicated and burdensome, and a future DINI should seek to help ease this barrier to data reuse in the research sector working with partners across the sectors to broker access to data sources.

2. **A DINI should manage trusted access and broker data sharing agreements between data suppliers and researchers.**
3. **A DINI should broker access to industry and government data for use within research.**

An approach to mitigating legal barriers could include the provision of reference data products to ensure the supply of suitable data with clear usage conditions for use in research in a timely manner. A DINI should work with data suppliers to create and manage the provision of standard data products with predefined terms and conditions. This could include generating synthetic data sets to act as open proxies to the real data,

This approach would also reduce the costs associated with data sharing. Data suppliers need only work with one representative to reach a wide research community, accessing data curation expertise to assist suppliers to enhance the quality of the data should be provided. DINI could act as a conduit for the research community to work with and interact with commerce and government. It would also build on the work in various sectors on regulatory open data requirements.

DINI could provide model DSAs. However, we would not expect that DINI would provide DSAs for every request for data outside the reference data, as there are likely to be specific legal requirements for a specific research project. However, a DINI should provide guidance and advice on accessing data.

We note that it is also important to provide infrastructure that supports research users to respect the conditions of use for data, captured in the Accessibility conditions of the FAIR data principles.

4. **A DINI should support the identification and creation of reference data with clear access conditions and agreed data sharing arrangements.**

Human Capacity

The skills and knowledge of all stakeholders to enable data sharing is key to success. A future DINI can support this through outreach activities.

We propose that raising awareness of the value of data sharing and publication, the embedding of data stewardship skills and the sharing of good practice should be a core function of DINI. It should include a programme of advocacy and outreach, including workshops, webinars and other support activities. DINI should provide guidance on community practices and build a knowledge base of training and guidance materials, in conjunction with others in the wider data management and sharing landscape.

DINI should provide a cadre of data stewards to support data providers and researchers to embed community practices.

5. **A DINI should provide guidance for data suppliers and researchers on best data publication practices.**
6. **A DINI should build a network of communities of practice through guidance, training and setting standards.**

Engagement

DINI should include programme of advocacy and outreach, working across the sector. The provision of data discovery services and reference data sets may help level the playing field to a wider research audience. The development of further case studies and example projects with clear impacts would also encourage the wider adoption of data sharing practices.

7. **Research funders should advocate for the practical and strategic benefits of sharing national infrastructure data with researchers.**

Sustainability

We have seen that the data of value to research is diverse in nature and widely spread across the sector. The quantity of data and the costs of creating and maintaining a national data cloud should not be underestimated. If a DINI is to provide a trustworthy service that researchers and their partners can depend on into the future, it needs to have a consistent and long-term existence, and therefore it needs to have a sustainable level of investment from research funders.

8. **A DINI should have a long-term plan and sustained investment on its maintenance.**

The investment in a DINI should be justified by the value it returns in more and better research, and greater impact to wider stakeholders. The value of data sharing can made clear by examples from practice. This DINI should engage in a programme of research-based best-practice case studies with researchers and commercial organisations to better understand what researchers are looking for, and what value commercial and public sector organisations gain from working together.

9. **Funders of research and other data sharing programmes should invest in best-practice case studies with commercial organisations to demonstrate the tangible benefits of academic re-use of data**.

Compute, storage, network and AAI

Our pilot study has shown that access to data suitable for use in research is widely variable. Providing a federated infrastructure, building on resources already available, such as DAFNI, JASMIN, the UK Data Service and the EDC, together with community sources such as Stream, and identifying and filling gaps in the repository landscape will support researchers to find and use the data needed for their research. These repositories should be within a network of trusted repositories.

ISE is a large and complicated research area that ranges across a wide number of disciplines, and we would not expect a single Digital Research Infrastructure to accommodate all the computational infrastructure needs of the community. A future DINI would require the use of different platforms, and it would be highly beneficial to the user community to make co-use of these systems as straightforward as possible.

There is also a need for DINI to provide trusted access to computing resources. DSAs for sensitive or secure data sets may require controlled or limited access, and thus processing within a trusted research environment, moving the code to the data. The infrastructure should provide a trusted data space with suitable controls to allow trusted users to access, explore data and use it for further analysis securely.

10. **A DINI should provide access to data from a federated infrastructure of trusted repositories.**
11. **A DINI should provide access to federated trusted computing resources to perform analysis**.

Interoperability and standards

Consistent adoption of standards is key to interoperability. DINI should work with sector governance and community bodies to help set standards to form a data interoperability framework of standards for data and metadata representation and semantic annotation,

starting within energy, water and transport domains. This should have the aim of working towards an inter-disciplinary data framework, while acknowledging this is a complex interdisciplinary space.

A DINI should support a common data interoperability framework based on profiles of common, metadata standards which should work together to support FAIR data. The Framework should include broad metadata concepts which are applicable across the domain, together with ISE specific application profiles, be supported by APIs for machine access and allow annotation using domain ontologies.

A DINI would build on recent approaches to developing a common framework such as the Common Data Interoperability Framework (Gregory et. al. 2024), and the National Digital Twin Information Management Framework (IMF) (Hetherington and West, 2020) together with geospatial data standards. We note that (Varga et. al 2023) identifies that for infrastructure and city ontologies there is a "need for a small number of top-level ontologies and translation mechanisms between them". Thus, the framework should also build on existing work on Ontologies such as the UK Information Exchange Standard (IEF 2024).

It can be difficult to assess the quality of data, understanding its constraints and limitations. This can also be a barrier to releasing data as organisations can be reluctant to release data which they consider having shortcomings. Putting in place some quality indicators, or setting expectations for the information that should be provided for others to make data quality judgements would move this area forward. This is not something that a single body can do but is a community activity; DINI could act as a convenor in this space, starting with the suggested reference data sets.

12. **A DINI should support a common data interoperability framework in conjunction with wider community efforts to set data sharing standards.**
13. **Data suppliers and users should work to establish and apply, domain expectations and standards on data quality.**

Services and tools

Discovering data to be able to use it is an acknowledged barrier and so a future DINI needs to provide tools, such as a data registry and data indexing, to allow the discovery and access to data in different sources and data aggregators. We do not propose to take a copy of everything but rather index discovery via resource catalogues and data orchestration. These tools could be extended to cover other research outputs such as registries of energy models as recognised by the energy modelling community.

DINI should support the publishing and curation for data resulting from research, within any restrictions. This would provide a domain specific place for researchers to publish

FAIR data resulting from their research, enhancing impact and stimulating new research, while allowing credit to be attributed to researchers.

14. **A DINI should provide services to discover and access research data, and other associated research outputs.**
15. **A DINI should provide data stewardship expertise to work with researchers and other data suppliers to support long-lived research products**

Research Objects

The digital artefacts which we have been primarily concerned with in this study have been data and data collections which are used as the input to and the output of research activity. Datasets are the primary driver of research, and our analysis has focussed on the issues around sharing data. However, the research process also involves other research artefacts, such as formal publications and informal documents, training material, video etc. A class of artefact highlighted was software code and executables, particularly for computational models, which embody a considerable investment in effort and expertise; the Energy Modelling workshop proposed that there should be reference collections of models for use and reuse in research. We recommend that while the DINI should focus on data, it should also take into account other research objects and encourage and support their management for FAIR use and record the provenance of research outputs for traceability. Researchers should seek to apply the same standards and principles to the data they produce as they would like to see for the data they consume.

16. **Researchers and their research organisations should be encouraged to manage and describe their research outputs, such as data, models, research software and protocols, to ensure application of FAIR principles and to support interoperability, transparency and reproducibility.**

Concluding Remarks

Research into the evolution of infrastructure systems can have a real impact on the delivery of the strategic goals of government. Our work in use cases in this project has demonstrated that there can be multiple impacts of research; we highlight the potential contribution to the strategic goals of building a secure and resilient society, of boosting economic investment and innovation, and of ensuring the health of our environment.

However, we have also seen the extent that problems in accessing and sharing data have impeded the interaction between researchers and their partners. The participants in our consultations have highlighted barriers, particularly the caution that data suppliers can have on sharing data, and how researchers can respond to that challenge.

We have made sixteen recommendations in response which address these challenges; these synthesise the larger number of recommendations which can be found in the activity reports. Together, these characterise a research data cloud, and we believe that our vision for a future Data Infrastructure for National Infrastructure would deliver such a research data cloud.

The value which can return from data is recognised in the AI Opportunities Action Plan[11], which calls to *"build public sector data collection infrastructure and finance the creation of new high-value datasets that meet public sector, academia and startup needs"* and *"incentivise and reward researchers and industry to curate and unlock private datasets*. The National Data Library initiative is considering how to bring together existing research programmes and help deliver data-driven public services. A DINI would complement this goal, forming a bridge to the ISE research sector.

Infrastructure systems are inherently complex and multi-disciplinary, so the solutions require partnership, systems thinking and interoperability. We propose that a DINI should continue to build the community connections to meet the complexities faced by organizations across industry, government and academia.

[11] AI Opportunities Action Plan, Independent Report, Department for Science Innovation and Technology, 13 January 2025, https://www.gov.uk/government/publications/ai-opportunities-action-plan

Abbreviations Glossary

Abbreviations and acronyms used in this report

BRINES	Building Risk-Informed redundancy for Net-zero Energy Systems
ClimaTracks	Forecasting resilience of railway network under propagating uncertainty
CReDo	Climate Resilience Demonstrator
D-RES	Provision of distributed grid resilience using EVs during extreme weather events
DAFNI	Data and Analytics Facility for National Infrastructure
DCC	Digital Curation Centre
DCAT2	Data Catalog Vocabulary (DCAT) - Version 2
DINI	Data Infrastructure for National Infrastructure
DSA	Data Sharing Agreement
DSIT	Department for Science Innovation and Technology
e-UMDS	Enabling Urban Mobility Data Sharing
EDC	Energy Data Centre
FAIR	Findable, Accessible, Interoperable and Reusable
ForNet	FORecasting Services for Energy NETworks
IMPACT	IMproving flood disruPted road networks with a dynAmic people-Centric digital Twins
ISE	Infrastructure Systems Engineering
JASMIN	A data analysis facility within the UK providing storage and compute facilities to enable data-intensive environmental science.
MARS	Modelling Aviation Resilience Scenarios
NIC	National Infrastructure Commission
NUAR	National Underground Asset Register
SOURSD	Safe Organisation and User Registry for Sensitive Data
SOFRAMODE	Sewer overflow flood risk analysis model DAFNI enabled
STFC	Science and Technology Facilities Council
STORMS	Strategies and tools for resilience of buried infrastructure to meteorological shocks

TRE	Trusted Research Environment
UKCRIC	UK Collaboratorium for Research on Infrastructures and Cities
UKERC	UK Energy Research Centre
UKRI	UK Research and Innovation
WSL	Water Systems Leakage

Contributors to the DINI Project

Many people have contributed to this pilot project.

DAFNI (STFC): Brian Matthews (Project Lead), Katie Cartmell (Project Manager), Jens Jensen, Bethan Perkins, Lewis Sampson, Tom Kirkham, Kyle Stevenson, Server Kasap, Sarah Byrne, Alison Oliver, Catherine Dhanjal

Open Science (STFC): Elizabeth Newbold (Final Report Editor), Teagan Zoldoske, Karen VanHaltren, Katie Yates

UKERC Energy Data Centre (STFC): Catherine Jones (Final Report Editor), Peter Holt

JASMIN: Adrian Hines, Sam Pepler, Matthew Pritchard and Alexander Manning

Digital Curation Centre: Kevin Ashley, Joy Davidson, Clara Lines Diaz, Agnes Jasinska and Daniel Turner

Icebreaker One: Charlotte Horler, Caroline Fraser, Emily Judson, Zoe Swiderski, Ceri Stanaway, Gea Mikic, Miranda Sharp, Lucy Stewart, Arup, Ewan Jarvis

DAFNI Strategy Board Champions:

Water Systems Leakage (WSL): Prof Liz Varga, Ruoqing Yin, Haonan Xu, Jiaqian Wei, Univ. College London

The use, barriers and opportunities for Quasi-Real Time data in transport: Dr Giuliano Punzo and Toby Willis, Univ. of Sheffield; Prof. Daniel Coca, Newcastle Univ.

Enabling Urban Mobility Data Sharing (e-UMDS): Dr Sam Gunner, Prof. Theo Tryfonas, (Univ. of Bristol)

User project teams:

BRINES: Dr Hannah Bloomfield, Dr Colin Manning, Prof Sean Wilkinson (Newcastle Univ.) and Dr Ji-Eun Byun (Glasgow Univ.)

D-RES: Dr Desen Kirli, Univ. Of Edinburgh;, Peter McCallum, Univ. Of Edinburgh;, Laiz Souto, Univ. Of Bristol; and Killua Qin, Univ. Of Edinburgh.

ClimaTracks: Giuliano Punzo, Univ. Of Sheffield; Ji-Eun Byun, Univ. Of Glasgow;, Tohid Erfani, UCL; Qian Fu, Univ. Of Birmingham; Konstantinos Nikolopoulos, Durham Univ., and Iryna Yevseyeva, De Montfort Univ.

ForNET: Prof Kostas Nikolopoulos and Dr Vasileios Bougioukos, Durham Univ.

IMPACT: Tao Cheng, UCL;, Tohid Erfani, UCL;, Xuhui Lin, UCL;, Qiuchen Lu, UCL; Trung Hieu Tran, Cranfield Univ.; and Xianghui Zhang, UCL.

MARS: Desmond Bisandu, Irene Moulitsas and Fabian Steinmann (Cranfield Univ.)

SOFRAMODE: Dr Vassilis Glenis and Dr Claire Walsh (Univ. of Newcastle)

STORMS: Xilin Xia, Nikolaos Reppas, Ali Mashhadi, Phatharaphong Yensri, Qian Li, Soroosh Sharifi, Asaad Faramarzi, Nicole Metje, David Hannah, Univ. of Birmingham; Steven Cole, Robert Moore, Adam Griffin, Alison Kay, Steven Wells, UK Centre for Ecology & Hydrology; Andrew Hughes and Javid Yousaf, British Geological Survey

DAFNI Strategy Board:

Prof. Michael Batty, Univ. College London

Dr Ruchi Choudhary, Univ. of Cambridge

Prof. Daniel Coca, Newcastle Univ.

Prof. Asaad Faramarzi, Univ. of Birmingham

Prof. Jim Hall, Univ. of Oxford

Prof. Stephen Hallet, Cranfield Univ.

Prof. Julien Harou, Univ. of Manchester

Prof. Phil James, Newcastle Univ.

Prof. Nik Lomax, Univ. of Leeds

Dr Giuliano Punzo, Univ. of Sheffield

Prof. Theodoras Tryfonas, Univ. of Bristol

Prof. Liz Varga, Univ. College London

DINI Expert Advisory Board:

Prof. Jim Hall, University of Oxford

George Economides, Dept. of Transport

Prof. Mark Enzer, Mott McDonald

Prof. Deborah Greaves, Univ. Of Plymouth

Sarah Hayes, Connected Places Catapult

Holger Kessler, Atkins Realis

Prof. Jason Lowe, Meteorological Office

Prof. Jeremy Morley, Ordnance Survey

Prof. Katherine Royse, STFC Hartree Centre

Prof. Jennifer Schooling, Anglia Ruskin University

Steven Steer, Zuhlke Group

Martine Wauben, Greater London Authority

References

Digital Twin Hub Data Sharing Working Group (2024). Cross-sector UK Data Sharing Infrastructure Data Sharing Initiatives – High-level Landscape Snapshot, October 2024. https://digitaltwinhub.co.uk/download/cross-sector-uk-data-sharing-infrastructure/

Fulmer, Jeffrey (2009). "What in the world is infrastructure?". PEI Infrastructure Investor (July/August): 30–32.

Gregory, A., Bell, D., Brickley, D., Buttigieg, P. L., Cox, S., Edwards, M., Doug, F., Gonzalez Morales, L. G., Heus, P., Hodson, S., Kanjala, C., Le Franc, Y., Maxwell, L., Molloy, L., Richard, S., Rizzolo, F., Winstanley, P., Wyborn, L., & Burton, A. (2024). WorldFAIR (D2.3) Cross-Domain Interoperability Framework (CDIF) (Report Synthesising Recommendations for Disciplines and Cross-Disciplinary Research Areas) (Version 1). https://doi.org/10.5281/zenodo.11236871

Hahnel, Mark; Smith, Graham; Campbell, Ann (2024). The State of Open Data 2024: Special Report Bridging policy and practice in data sharing. Digital Science. Report. https://doi.org/10.6084/m9.figshare.27337476.v2

Hetherington, J., & West, M. (2020). The pathway towards an Information Management Framework - A 'Commons' for Digital Built Britain. https://doi.org/10.17863/CAM.52659

Jones, S., Leggott, M., Lopez Albacete, J., Pascu, C., Payne, K., Schouppe, M., Treloar, A., & Global Open Research Commons IG. (2023). GORC IG: Typology and Definitions (1.01). Zenodo. https://doi.org/10.15497/RDA00087

National Infrastructure Commission (2017). Data sharing in infrastructure, a New Technologies Case Study. Prepared by Deloitte for the National Infrastructure Commission, 2017. https://nic.org.uk/studies-reports/new-technologies/data-sharing-in-infrastructure-2/

UK Geospatial Data Standards Register, 15 April 2024 https://www.gov.uk/government/publications/uk-geospatial-data-standards-register/national-geospatial-data-standards-register

UK Information Exchange Standard, Release 4.3.0, 16 December 2024, https://github.com/dstl/IES4/blob/master/IES%20Specification%20Docs/Information%20Exchange%20Standard%20r4.3.0%202024-12-16.pdf

Varga L, McMillan L, Hallett S et al. (2023). Infrastructure and city ontologies. Proceedings of the Institution of Civil Engineers – Smart Infrastructure and Construction 176 (2): 43–52, https://doi.org/10.1680/jsmic.22.00005

Appendix: Methodology

The DINI project was coordinated by the DAFNI team at Scientific Computing, STFC Open Science, the UK Energy Data Centre (EDC) and the Centre for Environmental Data Analysis (CEDA), alongside project partners; DAFNI Strategy Board members conducting 'Champions' projects, Digital Curation Centre (DCC), Icebreaker One (IB1), academics conducting 'Use case projects' and UK Collaboratorium for Research in Infrastructure and Cities (UKCRIC).

The project completed a preparatory phase, refining the project plan, completing a stakeholder mapping exercise, stakeholder and community engagement, reviewed all project work packages and conducted and complete monitoring and evaluation reports. Following this, the project completed a Situational analysis, funded academic use cases, conducted technical demonstrations and synthesis phase to review the project evidence and conclude findings and recommendations. All documents referenced throughout the methodology document and the projects final report can be found in the supporting documents section.

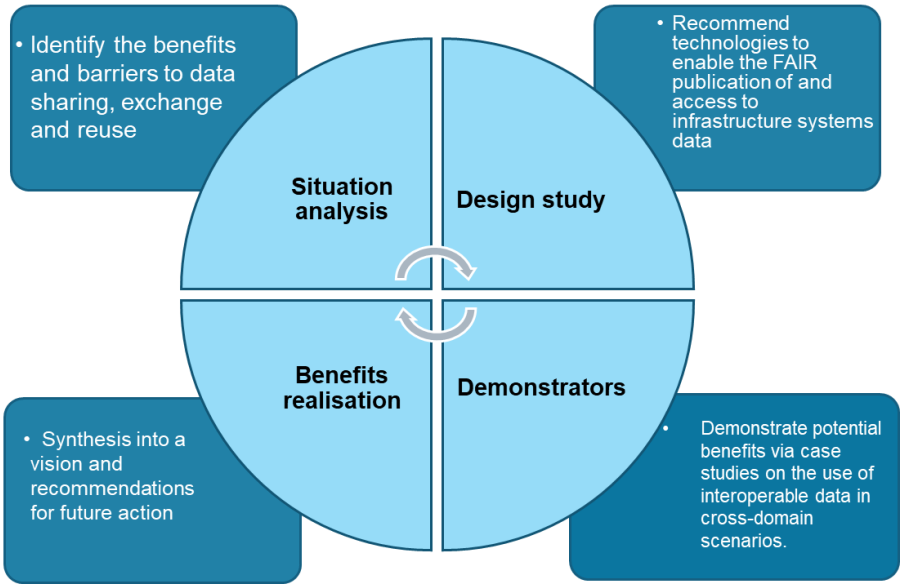


Figure 11: Pictorial representation of the project

Situational analysis

The situational analysis identified the benefits and barriers to data sharing, exchange and reuse for infrastructure systems data and related domains. The project conducted a literature review, surveys, interviews, gathered evidence from the community, created an expert advisory group and a summary landscape report.

Literature review

The purpose of the literature review was to identify and review of information on relevant pre-existing studies of data availability, sharing and integration in the relevant domain.

This part of the project was completed by Icebreaker One, the document can be found in the supporting documents 'RESEARCH: DAFNI-DINI Landscaping – Literature review'.

The community engagement and information gathering are an extensive part of the projects work, this included surveying the community, organising and hosting devolved nations workshops and gaining strategic direction from an expert advisory board.

Surveys

Project partners Icebreaker One (IB1) conducted surveys on behalf of the project, IB1 conducted a stakeholder mapping exercise and worked with the project to finalise the stakeholder survey list. IB1 and the DINI project team shared the invitation via email and used online social media platforms such as LinkedIn to survey the communities and stakeholders. The target audience was academia, government and industry, across energy, water and transport sectors.

Interviews

IB1 conducted the interviews on behalf of the project. IB1 completed the following steps; stakeholder identification, review stakeholder list with the project team, prepare interview questions, finalise questions with the project team, schedule interviews, hold interviews and analyse outputs. The interviews were held across October and November 2024. The target audience was academia, government and industry, across energy, water and transport sectors.

Workshops

Workshops in devolved nations

The project conducted a large-scale stakeholder mapping exercise within academia, government and industry across Northern Ireland, Scotland and Wales, specifically looking at stakeholders devolved nations in national infrastructure systems, in energy, water and transport. The project team invited stakeholders to attend via email and online meetings prior to the event. The workshops were held in-person on the 17th of September for Wales workshop, 23rd for Scotland workshop and the 29th for the Northern Ireland. The workshops were led and facilitated by the Digital Curation Centre. The work report can be found in the supporting documents 'Devolved Nations Workshops'.

Sector Workshops

Energy workshop

The UK Energy Data Centre conducted an energy modelling data workshop which focused on exploring the benefits and barriers to sharing data used and generated in modelling and to identify concrete actions to progress data sharing. The event was held virtually on the 2nd of October 2024. The UK Energy Data Centre targeted academia, government and industry participation. The workshop was led and facilitated by Cultivate Innovation. Workshop report available from https://doi.org/10.5286/UKERC.EDC.000985

Energy Small-Medium Enterprise (SME) Workshop

The project conducted an energy sector workshop which focused on current data sharing culture, approaches and benefits and data sharing challenges and gaps in support. The event was held in-person on the 19th of November. The project team and the Digital Curation Centre worked in collaboration with Innovate UK and their competition winners from the funding opportunity 'UK Digital Twins Energy Grids', the workshop audience range from academia, SME and industry. The workshop was led and facilitated by Digital Curation Centre. The workshop report can be found in the supporting documents 'Innovate UK Workshop'.

Transport and Water workshop

IB1 conducted a Transport and Water workshop which focused on the benefits and barriers of data sharing with researchers and solutions to barriers. The events were held virtually online, the transport workshop was held on the 4th of November and the water workshop held on the 5th of November. IB1 targeted private and government sectors. The workshops were led and facilitated by IB1.

UKCRIC 's workshops

UKCRIC conducted two workshops, the first workshop with their advisory board which was comprised of government and industry stakeholders. The focus of this workshop to explore the challenges and opportunities of data sharing and data gathering across industry-academia interface. This event was held on the 20th of May 2024. The second work conducted was research led workshop, focusing on academics who lead urban observatory facilities. The focus of this workshop was to explore challenges and opportunities of data sharing and data gathering, focusing on urban observatories perspective. This event was held on the 6th of September 2024. Both workshops were led and facilitated by UKCRIC. The workshop report can be found in the supporting documents 'UKCRIC Workshops'.

Expert Advisory Board

The project established an advisory board to gain strategic advice, guidance and steer throughout the project direction and opportunity to receive feedback on project direction and outputs, alongside engagement and outreach support from the board members. The project defined the purpose of the board, defined the terms of reference and board requirements (board size, meeting frequency, meeting preferences) and conducted a stakeholder mapping exercise. The DAFNI programme lead Dr Brian Matthews invited all board members, followed by a series of meetings scheduled throughout FY24/25 (July 2024, November 2024 and March 2025). The DINI Expert Advisory board members details can be found in the appendix of the 'Data Infrastructure for National Infrastructure A UK Research Data Cloud Pilot, Final Report'.

Design studies

The project team conducted technical design studies throughout the project to support the realisation of the DINI vision. The following studies were undertaken: DAFNI technical team researched DAFNI's current data indexing service and detailed how this could be developed in the future; DAFNI and JASMIN technical teams investigated how the two platforms could be linked (DAFNI platform and JASMIN) and the EDC explored semantic artefacts to add additional meaning and API specification focussed on FAIR data retrieval. Trusted Environments also identified as a key topic, however, were investigated in other projects in the Research Data Cloud programme.

Data Discovery and Access Services

Data Indexing

The technical team explored the DAFNI platforms current data indexing service, its capabilities and highlighted planned development for the future and requirements for a data indexing service for a research data cloud. A report was created and can be found in the supporting documents 'Data Infrastructure for National Infrastructure: Data Indexing Service Review'.

API's for FAIR data access

EDC researched, identified use cases, designed and defined an API specification for FAIR data retrieval for the EDC, this can be found here: Developing a specification for a FAIR-enabled API for the Energy Data Centre https://doi.org/10.5286/UKERC.EDC.000986

Data Interoperability Framework

Semantic artefacts

The UK Energy Data Centre explored semantic artefacts such as ontologies and subject classifications across the energy sector. The EDC reviewed the current ontologies used by the EDC and other services (UK and wider) and wrote a report: Reviewing the Energy Semantic Artefacts Landscape available from https://doi.org/10.5286/UKERC.EDC.000987

Federated Data Infrastructure

DAFNI Platform and JASMIN Integration

The DAFNI technical team collaborated with the Centre for Environmental Data Analysis (CEDA) and JASMIN team, to understand the architecture of both platforms, the user stories that drive the reason for the integration of the platforms, explored the technical feasibility and options for integration. A report has been created and can be found in the support documents 'Federating Research Infrastructures: DAFNI & JASMIN. The project team summarised this information and presented a technical poster at the projects

showcase event, this can be found in the supporting documents 'DAFNI DINI Technical Poster'.

Demonstrators

The project funded academic research to demonstrate the potential benefits via real world exemplars on the use of interoperable data in cross-domain scenarios. To demonstrate the benefits, the project organised three funding opportunities, the first opportunity was to the public via a sandpit process, the second and third opportunity was a closed funding call, this was due to the project requiring a niche expertise target group. All demonstrator project use case templates can be found in the supporting documents.

Demonstrator projects

The project team funded 8 demonstrator projects. The first open funding call, the project defined requirements for the funding opportunity, released an announcement of opportunity, organised an online sandpit process and arranged two expert panels to score and award the funding (transport and energy). Detailed information on the sandpit process and announcement of opportunity can be found in the supporting documents 'Infrastructure Resilience Sandpit Announcement of Opportunity'.

The second closed funding opportunity, the project defined the funding requirements, released an expression of interest to the community, review the expression of interest forms and scored within the DAFNI programme team and awarded two projects funding. The announcement of opportunity can be found in the supporting documents 'DAFNI Expression of Interest – Challenges and Barriers to Data Sharing'.

Champions

The second closed funding opportunity was via a DAFNI 'Champions' process, the project defined an announcement of opportunity, released the announcement to the targeted group, organised an expert panel to score and award the funding. Detailed information on the champions announcement of opportunity can be found in the supporting documents 'DINI Champions Announcement of Opportunity'.

Benefits Realisation

The DINI project collected vast amount of qualitative and quantitative data to understand the data sharing context of research within national infrastructure engineering, looking at the opportunities and barriers to data sharing, and considering appropriate solutions to address the barriers and enhance the potential for research impact from data sharing.

The work of gathering evidence is thus broad and undertaken by several project partners. The project created a template for reporting, which enabled the project to complete quality control checks, compare and analyse the results to draw common conclusions

and recommendations. The three key areas analysed from the reports were benefits, barriers and recommendations for future development.

The project team defined roles and responsibilities for delivering the report and the report sign off process. All project outputs, including evidence gathered in the situational analysis, use cases, technical demonstration and synthesis phase are presented as standalone reports in the appendices.

Supporting Documents

A: Literature Review: Ceri Stanaway; Contributors: Emily Judson, Charlotte Horler, Caroline Fraser, IceBreaker One

B: Industry Consultation Report: Ceri Stanaway; Contributors: Emily Judson, Charlotte Horler, Caroline Fraser, IceBreaker One, Lucy Stewart, Arup, Ewan Jarvis

C: Workshop reports

C1: UKCRIC workshops: Dr Joanne M Leach, UKCRIC

C2: Devolved Nations Workshops: Kevin Ashley, Joy Davidson, Clara Lines Diaz, Agnes Jasinska, Daniel Turner, DCC

C3: Innovate UK Workshop: Kevin Ashley, Joy Davidson, Clara Lines Diaz, Agnes Jasinska, Daniel Turner, DCC

C4: Energy Data Centre Workshop: Catherine Jones (UKERC, STFC); Elizabeth Newbold, Katie Yates, Teagan Zoldoske (Open Science, STFC), Andy Boston (RedVector), Mike Colechin, Keelan Colechin (Cultivate Ltd); Joy Davidson, Clara Lines and Agnes Jasinska (DCC)

D: Detailed Case Studies:

D1: Water Systems Leakage data (WSL): Ruoqing Yin, Haonan Xu, Jiaqian Wei, Liz Varga, Univ. College, London

D2: The Use of Quasi-Real-Time Data in Transport Infrastructure (QRT): Daniel Coca, Newcastle Univ.; Giuliano Punzo and Toby Willis, Sheffield Univ.

D3: Urban mobility data (e-UMDS): Sam Gurner and Theo Tryfonas, Univ.of Bristol

E: Use cases

E1: Building risk-informed redundancy in energy systems transitioning to Net Zero (BRINES): Dr Hannah Bloomfield, Dr Ji-Eun Byun, Dr Colin Manning, Professor Sean Wilkinson

E2: Forecasting resilience of railway network under propagating uncertainty (ClimaTracks): Giuliano Punzo, Univ. Of Sheffield; Ji-Eun Byun, Univ. Of Glasgow;, Tohid Erfani, UCL; Qian Fu, Univ. Of Birmingham; Konstantinos Nikolopoulos, Durham Univ., and Iryna Yevseyeva, De Montfort Univ.

E3: Provision of distributed grid resilience using EVs during extreme weather events (D-RES): Desen Kirli, Peter McCallum; Laiz Souto; Killua Qin

E4: Forecasting Services for Energy Networks (ForNet): Dr Vasileios Bougioukos and Professor Konstantinos Nikolopoulos

E5: Improving flood-disrupted road networks resilience with dynamic people centric digital twins (IMPACT): Tao Cheng, Tohid Erfani, Xuhui Lin, Qiuchen Lu, Trung Hieu Tran, Xianghui Zhang

E6: Modelling Aviation Resilience Scenarios (MARS): Desmond Bisandu, Irene Moulitsas, Fabian Steinmann

E7: Sewer Overflow Flood Risk Analysis Model DAFNI Enabled (SOFRAMODE): Vassilis Glenis and Claire Walsh

E8: Strategies and Tools for Resilience of Buried Infrastructure to Meteorological Shocks (STORMS): Xilin Xia, Nikolaos Reppas, Ali Mashhadi, Phatharaphong Yensri, Qian Li, Soroosh Sharifi, Asaad Faramarzi, Nicole Metje, David Hannah; Steven Cole, Robert Moore, Adam Griffin, Alison Kay, Steven Wells, Andrew Hughes, Javid Yousaf

F: Infrastructure Components

F1: Federating research Infrastructures: DAFNI and JASMIN: Katie Cartmell, Adrian Hines, Jens Jensen, Tom Kirkham, Lewis Sampson and Brian Matthews

F2: Data Infrastructure for National Infrastructure: Data Indexing Service Review: Lewis Sampson and Brian Matthews

F3: DAFNI DINI Technical Poster: Katie Cartmell, Jens Jensen, Brian Matthews & Lewis Sampson

F4: Reviewing the Energy Semantic Artefacts Landscape: Catherine Jones, Karen VanHaltren and Teagan Zoldoske

F5: Developing a specification for a FAIR-enabled API for the Energy Data Centre: Peter Holt and Catherine Jones

G: Infrastructure Resilience Sandpit Announcement of Opportunity: Katie Cartmell and Brian Matthews

H: DAFNI Expression of Interest – Challenges and Barriers to Data Sharing: Katie Cartmell and Brian Matthews

I: DINI Champions Announcement of Opportunity: Katie Cartmell and Brian Matthews

Editors

Brian Matthews leads the Data Engineering Theme in Scientific Computing at STFC and leads the DAFNI Programme. He has a background in research data management and metadata, working with STFC facilities, and with scientists in chemistry, materials science, environment, and infrastructure systems engineering. He has published extensively on FAIR data and open data sharing within scientific research practice.

Katie Cartmell manages the DAFNI programme within the Scientific Computing department at STFC and is responsible for the successful delivery of projects and ensures the team fulfil their deliverables. With a decade of experience in project management, she specialises in managing science-based projects. Holding a strong educational background in science, she also brings a unique blend of technical knowledge and project management expertise to every project.

Catherine Jones leads the Energy Data Centre within the Energy Research Unit, Technology Department, UKRI/STFC. She has a wide experience in providing information systems and services to the academic community, both within and external to STFC, using her software engineering and information management expertise to deliver effective services to user communities. Her work interests are the digital curation of software & data, demonstrating FAIR-ness of research data and linking research outputs (data, publications and software).

Elizabeth Newbold leads the Open Science Theme in Scientific Computing at STFC. She has a background in scientific information and libraries. Her work interests focus on supporting and facilitating open science practices.

Partners

The Data and Analytics Facility for National Infrastructure (DAFNI) (https://www.dafni.ac.uk) is a computing platform hosted by STFC which supports collaborative research into national infrastructure, including transport, water, and energy. DAFNI aims to improve the efficiency, reliability and sustainability of infrastructure through better sharing and use of data, and sharing and integration of models to explore the impacts of changes in infrastructure.

The UK Energy Research Centre (www.ukerc.ac.uk) undertakes "*Independent whole systems research for a sustainable energy future*". The Energy Data Centre (www.ukerc.rl.ac.uk) is a capability of UKERC, hosted at STFC and provides a discovery portal to research data, grey literature and funded projects content. The focus is on giving FAIR access to research data. It has been in existence since 2004.

JASMIN (https://jasmin.ac.uk/) is a data analysis facility providing storage and compute facilities to enable data-intensive environmental science. Over 1500 users are explore topics from climate change and oceanography to air pollution, earthquake deformation and analysis of wildlife populations. JASMIN is designed and operated by the Science & Technology Facilities Council on behalf of the Natural Environment Research Council.

The Digital Curation Centre (DCC) (https://www.dcc.ac.uk/) is a centre of expertise in digital information curation with a focus on building capacity, capability and skills for research data management. Based at the Universities of Edinburgh and Glasgow, the DCC provides expert advice and practical help on how to store, manage, protect and share digital research data.

IceBreaker One (https://ib1.org/) is a neutral non-profit that works on data sharing and sustainability. IB1 operates at the intersection of the public, private, and national interest with a focus on data infrastructure, governance, and strategic insight. Its work is designed to ensure that data is leveraged as a key asset to drive economic growth, policy effectiveness, and net zero innovation.

The UK Collaboratorium for Research on Infrastructure and Cities (UKCRIC) (https://www.ukcric.com/) is a network of UK universities connecting research with policy and practice in infrastructure and urban systems, UKCRIC provides the transdisciplinary, systems-based research for the transformation of infrastructure and urban systems, enabling safe, resilient and sustainable living, and generating economic opportunities for the UK.

Acknowledgements

The Data Infrastructure for National Infrastructure project was funded by the Department of Science, Innovation and Technology as part of the Research Data Cloud Pilot Programme (2023-25).

This DINI pilot project was made possible with the collaboration and research efforts of all our project partners, thank you for your contributions. We would like to thank the many participants who have made time to attend our workshop and events and contributed with their expertise and insight.

In addition, we would like to thank the DAFNI Strategy Board and DINI Advisory Board members for their advice, guidance and feedback throughout the project lifecycle.

Lastly, we would like to thank the Department for Science Innovation and Technology for the funding opportunity and their support throughout the project lifecycle.

Science and Technology Facilities Council

Enquiries concerning this report should be addressed to:

RAL Library
STFC Rutherford Appleton Laboratory
Harwell Oxford
Didcot
OX11 0QX

Tel: +44(0)1235 445577
email: library@stfc.ac.uk

Science and Technology Facilities Council reports are available online at:

https://epubs.stfc.ac.uk

Accessibility: a Microsoft Word version of this document (for use with assistive technology) may be available on request.

DOI: 10.5286/stfctr.2024001

ISSN 2753-5797