National Infrastructure Data

A landscape analysis on data sharing with researchers

Prepared for
Data Analytics Facility for National Infrastructure (DAFNI)
Data Infrastructure for National Infrastructure (DINI)

Authors:
Ceri Stanaway, Emily Judson, Ewan Jarvis, Miranda Sharp, Lucy Stewart, Charlotte Horler, Gea Mikic, Caroline Fraser

Document: IB1-DAFNI-DINI-2024-REPORT
Version: v2024-12-19

Contents

Document title: IB1 DAFNI-DINI Landscape Report

Document code: IB1-DAFNI-DINI-2024-REPORT

URL: https://ib1.org/dafni

Version: v2024-12-18

Copyright and citations

IB1 DAFNI-DINI Landscape Report, v2024-12-18, https://ib1.org/dafni

This document is available at the URL above. If you would like any further information, would like to get involved or if you need a version of this document in a more accessible format, please contact research@ib1.org.

Audiences

This report details the data sharing landscaping conducted for DAFNI-DINI, its methodology and results.

Its primary audience are researchers and research institutions, research and innovation funders and infrastructure data providers who wish to understand how to **implement** improvements in the way data is shared between data providers and the research community, using multi-sector collaboration and a joined-up, systemic approach to data sharing.

Many of the principles and lessons herein are applicable beyond the scope described.

[1] Creative Commons, CC BY 4.0 Attribution 4.0 International https://creativecommons.org/licenses/by/4.0/

Executive summary

Icebreaker One, in collaboration with Arup, supported the DSIT-funded Data Infrastructure for National Infrastructure (DINI) project within the Data Analytics Facility for National Infrastructure (DAFNI) programme. They conducted a pilot study on the requirements and impact of supporting improved sharing of national infrastructure data with publicly funded researchers, focussing on energy, water and transportation sectors.

Key findings

Experience to date of infrastructure data sharing with researchers is diverse: some organisations consider this 'business as usual', while others never share data. While Open Data publishing using standardised portals and licenses is on the rise, the sharing of other types of data, subject to access controls or licence restrictions, is dominated by bespoke, individually negotiated agreements. Even where standardised portals exist, different portals may be used across (and sometimes even within) different sectors, causing potential issues when looking to combine datasets from different sectors. Infrastructure data providers demonstrate clear demand for standardised, interoperable, low cost, and scalable legal mechanisms to share data. They also indicate that better knowledge of the research community and researcher user needs could improve data sharing for all parties.

Benefits to sharing infrastructure data with researchers were analysed and mapped into the following categories: social, economic, environmental, and outcomes for data providers. Analysis generated the overarching finding that data provider understanding of benefits is often theoretical at present, with few organisations able to concretely evidence or measure such benefits either quantitatively or qualitatively. Where benefits are measured, this tends to prioritise direct monetary outcomes for data providers. There is a notable evidence gap regarding the translation of system-level benefits to their impacts on different types of infrastructure bodies.

This project also analysed and mapped current barriers to data sharing into legal, security, commercial, cultural, and technical categories. Analysis indicated that barriers varied amongst organisation types; however, of all barrier types, technical barriers presented the least substantial concern among participants. Security barriers to data sharing were identified as an area of growing concern, and in some cases were seen to be in conflict with approaches in academia (for example, cross-border sharing represents scholarly norm, but doing so poses additional security risks), but there is variable understanding of risk. Cultural barriers, including reputational and trust based

factors, were identified as strongly influential in decisions and practices governing infrastructure data sharing both with researchers and other audiences. Cultural barriers can be complex to pinpoint as they are often diffuse and influenced by relationships and events taking place across wider sectoral networks. However, the power of such concerns to block or restrict data sharing should not be underestimated.

Recommendations:

1. Data providers and those who facilitate data sharing (e.g. regulators, aggregators) **must** invest in co-designing appropriate and open governance for data sharing to foster data accessibility and interoperability within and across sectors. Good data governance **must** be based on the principles of transparency, accountability, engagement, and responsiveness in order to address legal, policy, security, and communications needs in addition to technical matters
2. Data sharing initiatives between infrastructure data providers and the research community **must** build on existing initiatives in order to promote interoperability, prepare for many-to-many data sharing, and reduce the burden on data providers and data users.
3. National data sharing initiatives **must** be designed with cohesion (with other initiatives), flexibility, extensibility and capacity-building in mind. This requires the ability to guide less digitally mature organisations through a gradual on-ramp, such as supporting the transition from manual to API data sharing.
4. Research and innovation funders **should** support research that makes clear and advocates for the practical and strategic benefits of sharing national infrastructure data with researchers. This should explicitly account for nuance concerning the sharing of different data types (e.g. business, operational, and geospatial) or sensitivities (e.g. personal, commercial, security) and provide frameworks or guidance on the translation of research outcomes and benefits back to data providing organisations.
5. Research funders **should** invest in documenting research-based best-practice case studies and use cases for commercial organisations to better understand what data researchers are looking for, and in what formats, to enable data providers to prioritise service development.
6. The research community **should** actively engage with commercial data providers to embed researchers within organisations to build a collaborative approach, which will ease the administrative burden on data providers.

Introduction

Globally, across all industries, vast amounts of data are generated on a daily basis. It is estimated that for every person on earth, 1.7 MB of data is created every second[2]. Data, used effectively, has the potential for huge socio-economic and environmental benefits. To maximise its potential, data has to be accessible – it has to be findable, shareable, usable and, where appropriate, reusable (in line with the FAIR principles[3]).

Sharing data for research is an important component of the data sharing landscape, which can offer distinct advantages both to data sharing organisations and wider society. Researchers that are external to an organisation in which data originated can offer fresh perspectives to maximise data value, enable new forms of analysis through dataset linking, and offer important insights into the quality of data. Data held and published by public bodies is valuable, but data held by private companies is equally crucial, potentially offering 'powerful insights into the behaviors of individuals, communities, organisations, systems, and the physical environment—as well as into the interactions among these levels' (The Royal Society; National Academy of Sciences, 2024, p6). However, accessing such data can be challenging, with researchers – and other users – facing issues around legalities, privacy, commercial resistance and cultural reticences, as well as technical challenges.

This research project took an engagement-led approach to developing a literature review and landscaping exercise to understand the current data sharing need, practice and barriers in the environment in the water, transportation, and energy sectors.

Project scope and methodology

Icebreaker One (IB1), in collaboration with Arup, supported the Data Infrastructure for National Infrastructure (DINI) project within the Data Analytics Facility for National Infrastructure (DAFNI) programme by conducting a pilot study on the requirements for and impact of supporting sharing and analysis of data across National Infrastructure systems in the UK focussing on energy, water and transportation.

Stakeholder mapping

Our methodology reflected the requirement to focus on effective engagement, and targeting the right questions for input to shape the report. Our initial phase of research

[2] Digital Decarbonisation (no date, accessed 2024.12) 'Digital Decarb - The Figures' https://digitaldecarb.org/the-figures/
[3] https://www.go-fair.org/fair-principles/

included a literature review and stakeholder and ecosystem mapping to ensure we were talking with and prioritising stakeholders to attend the workshops, participate in a survey, or interview.

The stakeholder list for this project was drawn up by IB1 in partnership with Arup, based on existing contact networks, with input provided by the DAFNI-DINI team.

When approaching potential research participants, IB1 aimed to include a representative sample of different types of stakeholders, including:

- Commercial companies (both energy/transport/water suppliers and consultancies)
- Government bodies (local and national)
- Regulators
- Arms-length public bodies
- Trade/professional bodies

IB1 aimed to include representatives from the devolved nations in the UK. Members of the DAFNI advisory group, academic and other publicly-funded research organisations were not specifically approached to take part in IB1's research, as such organisations are heavily involved in other strands of the DINI project. Some organisations participated in both the survey and either a workshop or an interview.

Literature review

The literature review was conducted between August and September 2024, and included literature published between 2017 and 2024 (inclusive). The review focused primarily on available 'grey literature', largely produced by industry, policy, and third-sector bodies, and on publications made available to the general public without paywalls or other access barriers.

Literature was reviewed to draw out evidence, analysis and good practice examples of:

1. The benefits of sharing infrastructure (energy, water, transport) data
2. The barriers to sharing infrastructure data.

The analysis focused on instances of data sharing with publicly funded researcher audiences (academic and publicly-funded adjacent research). However, significant literature gaps were identified during the research process. Accordingly, where appropriate, the analysis made inferences about data sharing practices which were assessed in the literature in terms of wider audiences (e.g. including data sharing with commercial or government bodies).

The full literature review, including more details on methodology, can be found here: RESEARCH: DAFNI-DINI Landscaping - Literature review - FINAL 2024-10-14.

Survey

In October and November 2024, IB1 carried out a structured survey, distributed to in-scope organisations across the energy, water and transport sectors, including academia, government bodies (including regulators), non-profits, and commercial enterprises (including trade bodies). The survey was designed to gather insights on data-sharing practices, perceived benefits, and barriers to sharing data with researchers external to the organisation. The survey received a total of 23 responses, representing diverse organisation types and roles across energy, water, and transport.

Survey structure

The survey consisted of a mix of closed and open-ended questions. Key areas of focus included:

- Types of data collected and shared
- Perceived benefits of data sharing (theoretical and tangible)
- Barriers to data sharing
- Measures implemented or needed to overcome barriers
- Sector-specific insights and case studies

A survey was chosen as a quantitative instrument with potential to collect data from a wide range of organisations and job roles. The survey underwent user testing, with representatives from IB1, Arup and DAFNI-DINI consulted prior to the main launch. The survey allowed respondents to rank barriers and benefits, as well as provide additional qualitative details where applicable.

Note that as not all organisations responded to all survey questions, the sample size for each response varies.

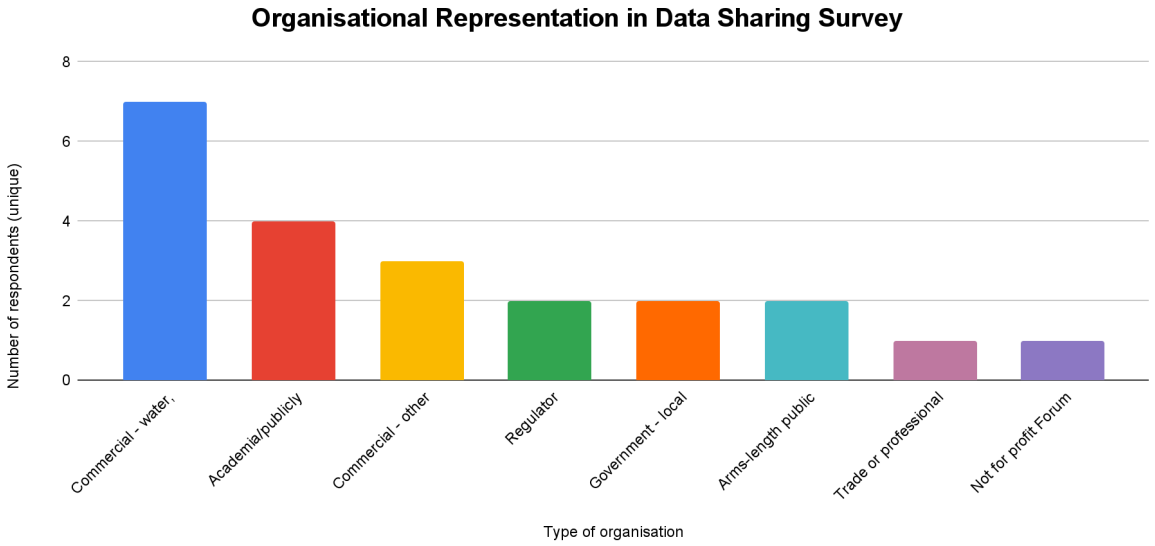

Chart 1: Organisational representation in the data sharing survey

Workshops

In November 2024, IB1 held two workshops to explore the benefits of and barriers to sharing infrastructure data with researchers – one focusing on transport (4 November 2024; six participants) and one focusing on water (5 November 2024; seven participants).

A workshop for each of the water and transport sectors was chosen as a method due to its strength of producing qualitative data and facilitating the exploration of more complex issue areas (e.g. political or ethical matters), with the acknowledged limitation of honesty amongst organisations with conflicts of interest or competitive concerns.

Attendees in each workshop were split into two groups and, after an initial poll to gauge the extent of data sharing with researchers among participants, discussions were facilitated across three sessions. The first session aimed to identify instances when the effective sharing of data with researchers resulted in positive impact. The second session identified barriers to sharing data with researchers and the third session aimed to identify possible solutions to the barriers identified.

Interviews

In November and December 2024, IB1 held one-to-one interviews with stakeholders from the energy (five stakeholders), transport (five stakeholders) and water (four stakeholders) sectors, prioritising stakeholders that had been unable to attend workshops.

Interview was chosen as a method due to its flexibility as an alternative to workshops, and the fact that it can be targeted to add depth or follow-up nuance to other methods. Interview questions focused on similar themes to the workshops, exploring stakeholders' experience of sharing data with researchers (with examples, where possible), the benefits and challenges of doing so, and any solutions they recommend to address barriers.

Synthesis and reporting process

Given the short timescale, data analysis has taken place iteratively across the project, with a period of synthesis in which findings can be integrated and written up and conclusions noted. This is the summary report to capture project process, collate inputs, and present analysis to ensure that the DAFNI-DINI team have a clear foundation for future work. In this report, we have synthesised the workshop and interview outputs as they covered similar ground and produced similar types of semi-structured free text data. Survey data has been analysed separately throughout due to its more structured format. This research analysis does not include the energy workshop run by the DINI team through the same time period. Data analysis notes can be found in Appendix 3.

Results Structure

The following sections of the report discuss the results from survey, workshop, and interview data. Firstly, results regarding current and recent data sharing practices between infrastructure and research are presented. Secondly, benefits to data sharing are discussed thematically, categorised in accordance with: unique benefits of sharing infrastructure data with researchers, societal benefits, economic benefits, environmental benefits, and benefits to data-providing organisations. Thirdly, barriers to data sharing are discussed thematically, categorised in accordance with: legal barriers, security barriers, commercial barriers, cultural barriers, and technical barriers. The solutions to barriers are then discussed where available in the data, focusing on the results of a ranking exercise conducted among transport and water workshop participants. The report closes by presenting a short set of recommendations and conclusions.

Current and recent data sharing practices

Survey

The survey represents the view of 23 organisations, comprising diverse organisation types and roles across energy, water, and transport. The sample for our research is likely to be biased towards organisations that are relatively mature in their data sharing, due to their readiness to engage with the research.

For example, the vast majority of respondents (19) said that they actively make data available to external researchers in some format.[4]

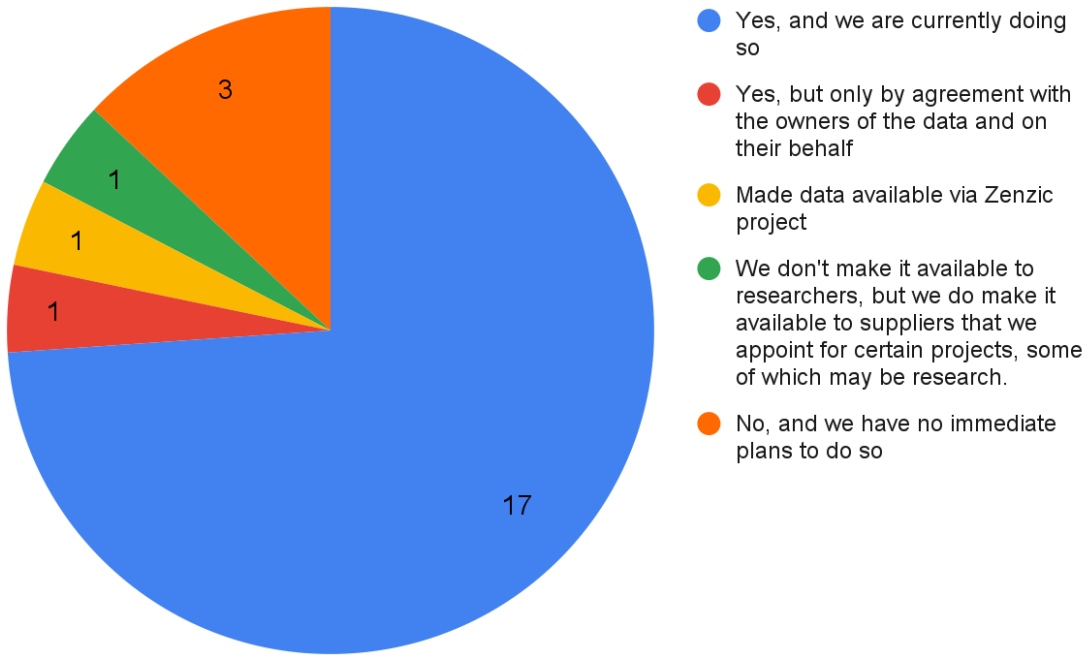

Chart 2: Responses to: Has your organisation/team made data available to external researchers?

How do organisations make data available to researchers?

Accessibility

Of the 20 organisations that 'make data available' to researchers in some way:

- three only make or have made data available as Open Data
- seven only make data available on a restricted basis[5]

[4] In one case this was with the caveat that this would require the consent of data owners, and another organisation noted that it makes data available to project suppliers, who may be researchers.
[5] This may variably take the form of Shared or Closed data arrangements.

- ten use both methods to share data.

Where data[6], beyond Open Data, is shared with researchers, this is typically on the basis of bespoke requests, though some organisations also use 'off the shelf' standard data products. Sharing data with researchers is most often through making data available to specific researchers or research consortia. Only five organisations said that they have made datasets available as a standardised product (e.g. via a data-sharing platform such as Open Data Soft). These included two research organisations, two commercial consultancies, and one government organisation.

Frequency

In terms of frequency, only two survey respondents (one arms-length public body and one research organisation) said that sharing data with researchers is part of their 'business as usual'. All other respondents share data either 'several times a year' or 'occasionally/on an ad hoc basis'.

Cost

Thirteen of the organisations do not typically charge researchers for access to data. Only one organisation said it charges for its data as standard, four said that they sometimes charge.

Why do organisations share data with researchers?

Organisations can share data with researchers for many reasons, for innovation programmes, policy making, and gaining new insights from datasets. The output from academic expertise can improve how they run their operations. Working together with researchers can also help them keep up with changes in their field and enhance their reputation as transparent, forward-thinking entities within their sector.

The table below summarises the type of organisation and their reasoning for sharing from the survey.

[6] As standard, Icebreaker One adopts Open, Shared and Closed data terminology to describe data which is shared with external actors as defined in: Icebreaker One (no date, accessed 2024.12) https://ib1.org/open-shared-closed/. While Open Data is considered, this report primarily addresses the sharing of Shared (data shared with a pre-emptive licence) or Closed (requiring, if shared at all, a user-specific custom licence or contract for use) data types with publicly funded researchers.

Type of organisation	Reason for sharing
Academia/publicly funded research body	For research, innovation, and policy making. However we vet organisations receiving data to avoid double or unintended use.
Academia/publicly funded research body	Evaluation and development of modelling tools that underpin design, siting and consenting of offshore infrastructure.
Academia/publicly funded research body	Usually for development / academic research purposes.
Regulator	Press articles
Arms-length public body	To assess impact on investment
Arms-length public body	Their own research
Commercial - other (e.g. data consultancy)	Collaborative research and PhDs etc. we fund or support, or consortiums of which we are a part.
Commercial - water, energy or transport service supplier	National Interest, The health of region, Contributing to the good of our communities, Environmental & Efficiency
Commercial - water, energy or transport service supplier	Usually academia for support with dissertations etc.
Commercial - other (e.g. data consultancy)	For the purposes of progressing a client's project or programme
Commercial - water, energy or transport service supplier	Innovation projects
Commercial - water, energy or transport service supplier	Usually students collecting data for dissertations
Commercial - water, energy or transport service supplier	Most commonly environmental impact studies, but the range of possibilities is broad
Regulator	Asset data openly available and widely used. Generally used for flood risk knowledge and modelling
Government - local	We have shared with suppliers appointed by the GLA Infrastructure team to deliver certain projects, like a report looking into the impact of decarbonisation on London's roads.
Government - local	Support for PhD studies and Masters studies, where we benefit from the insight and analysis undertaken. We help to set the topic with the students.
Arms-length public body (not-for-profit forum)	Research and advisory
Commercial - other (e.g. data consultancy)	To allow research and evaluation for potential new use cases

Table 1: Reasons for organisations sharing data with researchers

Interviews and workshops

Interview participants indicated varied experiences of research data sharing. While some organisations had a long history of sharing data with researchers (e.g. one Water

Network reported receiving 3-4 requests each week), others had more limited experience. Organisations also used a variety of different methods to share the data. They most commonly used bespoke sharing agreements (e.g. via partnerships or specific project consortia), which in some cases were associated with building partnerships in a tailored manner to benefit the data sharing organisation.

Interview participants also contributed limited discussion of sharing mechanisms consisting of 'off the shelf' data products, as well as hosting students or researchers within the organisation (e.g. one Water Network hosted a PhD student for a year). Participants also noted that some research collaborations go beyond simply supplying raw datasets. For instance, a Transport Service Provider has seen researchers request access to its proprietary models, and another stressed the importance of modelling and prediction to understand transport systems and advance connected vehicle technology. In these cases, sharing algorithms or other analytical tools can yield deeper insights than data alone.

During the transport and water workshops, the participants were polled specifically on the practicalities of sharing data with researchers, starting with the question ‘do you/have you shared data with researchers?’. Two participants had shared data via a data-sharing platform accessible to multiple researchers, five had shared through a number of ways (individual researchers, via a consortium or through a platform being the options listed), two had not shared with researchers, with the final two participants’ organisations not holding any relevant data, instead playing a facilitating role in the sharing of data.

Approaches to data access

Participants described a range of approaches to data provision, from portals only serving Open Data to more ad hoc, bespoke arrangements and evolving service models. While several organisations from each sector regularly published Open Data, neither publishing Open Data nor tracking its use (e.g. tracking researcher downloads or API connections) were universal practices. Many organisations currently do not maintain off-the-shelf products and instead prioritise case-by-case, relationship-driven collaborations tailored to mutual benefit (e.g. Transport Service Provider, Geospatial Data Provider). One participant discussed how in some cases, standard terms exist but still involve careful negotiation, underlining that even structured agreements require scrutiny to maintain data integrity and compliance (Anonymous).

Licensing approaches

Licensing approaches discussed by interview participants varied widely, ranging from standardised Open Data licences to tightly controlled, bespoke arrangements. Some

organisations indicated gradual adoption of more standardised licensing agreements particularly in the space of Open Data. For example the CC-BY-4.0 is used by an Energy Network as their standard Open Data licence. However, organisations from all sectors highlighted the current dominance of bespoke, negotiated data sharing agreements for Shared and Closed data exchange. These can take a variety of forms including: non-disclosure agreements (NDAs) (Water Network, Energy Industry Trade Association), 'light' contractual terms limiting raw data publication or onward sharing (Electric Vehicle Charging Provider, Transport Service Provider), a structured but flexible set of standard terms designed for researchers (Energy Sector, Anonymous), or specialist research licences (Geospatial Data Provider). Future demand for clearer, more uniform licensing terms was expressed by a variety of participants, largely in theoretical terms.

In the transport and water workshops, participants were asked a structured question: 'How does licensing for sharing data with researchers work?' For all of the participants that hold relevant data or share data with researchers, their response was 'something in between: one size fits all licence and a bespoke licence created every time'. This 'use case by use case' approach to licensing when data isn't approved for open release was corroborated by further discussion at a group level. Additionally, one participant discussed use of the Educational Services Provider Contract (ESPC) under the Framework Partner Contract in order to share data with universities which has one partner signed up (EDINA and its Digimap service[7]).

Charging and business models

Participants described a range of payment and business models tied to data sharing. While general sentiment supported free or low cost data access for research uses, this view was not universal. Some organisations charged fees under specific circumstances, as seen with Transport Service Providers, where access, support, and data might incur costs—albeit discounted for academic partners. Similarly, Water Networks indicated that while researchers could sometimes access data free of charge, commercial or development-focused use-cases might involve paid services. In other cases, payment was not the norm. An Energy Network, Electric Vehicle Charging Provider, Energy Industry Trade Association and Water Infrastructure Provider all stated they do not typically charge researchers, though some impose conditions such as restricting raw data distribution. A Regulator also advocated for free access in research contexts, and another Transport Service Provider's exploration of a data-as-a-service model pointed to the potential evolution of business terms.

[7] Digimap is an online map and data delivery service, available by subscription to UK Higher and Further Education establishments. Operated by EDINA at the University of Edinburgh, Digimap offers a number of data collections, including Ordnance Survey, historical, geological, LiDAR and marine maps and spatial data.

The influence of sensitivity factors

Participants highlighted various mechanisms and considerations tied to data sensitivity which have informed research data sharing to date. Several examples were given pertaining to legal compliance concerning sensitive or personal data. For example, one Transport Service Provider discussed how its organisation imposes strict conditions to safeguard personal data, including mandating how it is stored on secure university systems. One Water Network discussed its imperative to follow Environmental Information Regulations (EIR) protocols and ensure no infringement of regulated business areas, while another Water Network identified the internal use of a risk framework as a tool used to assess data prior to sharing in order to guard proactively against legal issues and time-consuming redactions. Organisations from the energy and transport sectors cited commercial sensitivity as a priority informing sharing practices to date. Intellectual property was a particularly important consideration for commercial organisations. These sensitivity factors are discussed in more detail in the section on barriers to sharing data for research.

Regulatory compliance and sharing initiatives

Emphasis was placed on the influence of regulatory and policy frameworks on data sharing. In some cases, compliance with Environmental Information Regulations (EIR) or licence conditions set by Ofgem guided what could be shared and how. For example, an Energy Network said that it operates under Ofgem licensing conditions, with another noting they follow Ofgem licence condition including the principle that data assets must be treated as 'presumed open'[8]. A Transport Service Provider participated in projects funded by bodies such as the Department for Transport and Innovate UK, indicating alignment with government-supported research initiatives. An Energy Industry Trade Association noted the value of Ofgem's efforts to standardise practice around publicly owned data, suggesting that clearer regulatory guidance could streamline the sharing of information

[8] Presumed open means that data must be made available for all people to use, unless the organisation responsible for handling the data provides evidence of a specific reason for needing to reduce its availability. Ofgem (2021) Data Best Practice Guidance https://www.ofgem.gov.uk/sites/default/files/docs/2021/05/data_best_practice_guidance_v0.3_0.pdf

Benefits of sharing infrastructure data

Data sharing has the potential for huge socio-economic and environmental benefits. These benefits can be explored by external researchers who offer different insights, and can lead to innovation or knowledge which maximises data value and enables new forms of analysis.

Survey

The majority of organisations who responded to our survey recognised the benefits of sharing infrastructure data with researchers, with 19 out of the 23 organisations we surveyed acknowledging at least one benefit, as shown in the chart below. Benefits to the organisation were most widely recognised, followed by benefits to the environment and society, and finally the UK economy as a whole. However, it is worth noting that, across all categories, most benefits were highlighted as being theoretical, rather than organisations having tangible evidence of the benefits.

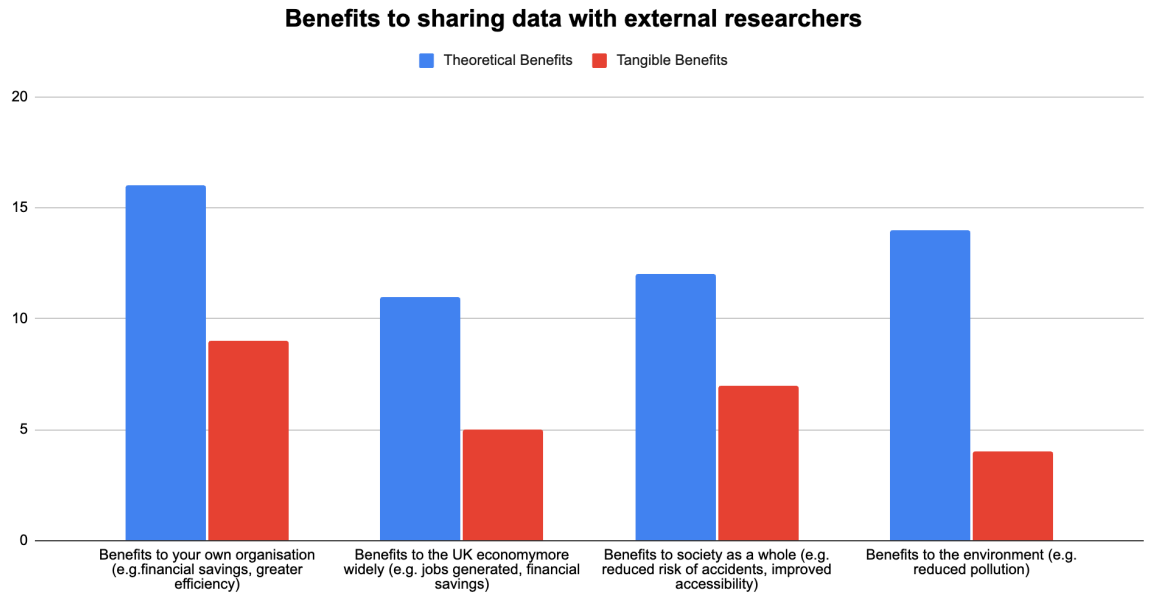

Chart 3: Benefits to sharing data with external researchers

We asked survey respondents to expand on the benefits they had identified, breaking the benefits down into societal benefits, economical benefits, environmental benefits, and benefits to their own organisation.

Benefits to society

Noted benefits to society range from advancing and building upon robust research, helping make the value case of data sharing, and protecting vulnerable customers.

Survey respondents highlighted that data sharing has advanced the field of research through: supporting and sponsoring geospatial research in Great Britain (GB); contributing significantly to making vehicles cleaner and safer; and upskilling students in research areas, which provides broader societal benefits.

Data sharing has proven the value case for organisations to continue to share data, including: enabling the detailed tracking of downloads, API access, and intended usage for river flow data (e.g. flooding, water resources); contributing to financial savings, efficiencies, and reduced incidents through analysis and data science models (although evidence is often intangible); and reducing disruption on London's roads through collaborative streetworks or district heating initiatives. Sharing data with research organisations can also produce benefits to vulnerable customers through initiatives such as the water industry initiative, Support for All[9].

Benefits to the UK economy

Cross-sector initiatives generate broader economic benefits. Sharing data with researchers can benefit the UK economy through reducing damages from environmental events, enabling efficiencies, and advancing innovation and opportunities in the UK.

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

Time, resource, and cost efficiencies can be demonstrated through projects such as the National Underground Asset Registry (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[10]. companies in its scope, and some data science models result in financial savings, efficiencies, or reduced incidents, though evidence may not always be tangible.

Aiding innovation and opportunities 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

[9] Ofwat (2023) Support for All https://waterinnovation.challenges.org/winners/support-for-all/
[10] Department for Science, Innovation and Technology and Geospatial Commission (2024) National Underground Asset Register https://www.gov.uk/guidance/national-underground-asset-register-nuar

increases revenue and tax income and contributes to the economy; and upskilling development among students, which increases the workforce in specialist field and supports long-term economic growth.

Benefits to data providers

The data provider benefits of sharing data with researchers include: increased collaborative projects, cost savings, and innovation development. Participants noted benefits of enabling collaboration including: supporting student recruitment, third-sector engagement, co-authored research, collaborations, and research income. One organisation noted that working with researchers aligns with their company values and fosters research community engagement.

One organisation discussed the cost savings associated with working with researchers. As an example, relevant research applied to predict traffic and air quality saved them the costs of engaging with a consultancy, which had estimated £20k–£50k.

Innovation development with researchers can enable: advancements in design standards, practice, and spin-out companies for data commercialisation; income generation and efficiency savings through data science models and research integration; enhanced project outcomes and performance commitments while reducing resource needs for data management; and the potential for new products and services development.

Environmental benefits

While there is an absence of concrete tangible evidence of environmental benefits of sharing data with researchers, 6 of 23 respondents noted possible applications of research to improving air quality, energy planning, and developing holistic approaches to climate adaptation.

Research 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 (though the practical impact may sometimes be overstated), and some data science models have contributed to reduced pollution incidents and financial savings, though tangible evidence is limited.

Research can also contribute to Local Area Energy Plans focusing on energy demand and capacity, and to the development of holistic models for climate resilience, prediction, and catchment-based operation.

Interviews and workshops

Benefits of data sharing specifically with academics or publicly funded researchers

The ability of academics, and other publicly funded researchers, to bring fresh perspectives and delve deeper into issues was cited as a cross-cutting benefit to sharing data with this community by organisations representing energy, water, and transport sectors. Participants particularly raised the value of researchers' – and research bodies' – independence and freedom in being able to select and analyse topics of inquiry and to present related results. Workshop participants also particularly identified the ability of researchers to pursue different forms of (more open) collaboration to commercial bodies as a distinct advantage.

The above attributes were highlighted to offer particular utility when exploring issues that are complex (e.g. multi-domain), contentious (e.g. ethical issues), require cutting edge skills and knowledge (e.g. modelling or data science expertise), or disincentivised by current market and/or value structures (e.g. short market cycles, perceptions of risk, or sunk costs). While not always explicitly discussed, these conversations indicated a perception among participants that researchers often hold positions which are more resilient to certain types of challenge or critique (e.g. findings from unsuccessful trials) than industry or the public sector. This was framed by one Water Sector Organisation as a key difference in primary motivation and capacity of research bodies; to pursue research as a goal itself rather than an activity performed in service of other goals.

The capacity for researchers and research bodies to pursue non-commercial goals was also repeated by participants across energy, water and transport sectors. Benefits of this broader focus were cited to include: creating room for different types of innovation; creating robust and multi-faceted test environments; gaining new perspectives on knotty challenges; and experimenting in environments with higher risk tolerances.

Despite such benefits being highlighted broadly by participants from different organisation types and sectors, it was notable that participants had difficulties conceptualising how such advantages could bring specific benefits back to their own organisations. While not explicitly discussed, in many cases this appeared to reflect a narrowing of how participants conceptualised benefits to their own organisation by focusing primarily on short term, direct, commercial returns. For example, this was highlighted by one interview participant from an Energy Network who cited limited benefits of data sharing for their organisation despite simultaneously identifying a range of social, environmental, and economic benefits of data sharing with the research community. Reflecting this juxtaposition in understanding, there is utility in further

research and communications activity to improve commercial organisations' understanding of how systemic benefits translate back to their particular organisations.

Society

Research participants highlighted a range of benefits to society stemming from sharing infrastructure data with publicly funded researchers. This section examines three main themes: public engagement, emergencies, and customer or service user benefits.

Firstly, participants from energy and transport sectors highlighted the value of data in improving public engagement with changes to infrastructures, citing the specific example of the net zero energy transition. A representative from an Energy Sector Trade Association discussed the capacity for better use of data in research to open new avenues for public engagement, particularly on a more personalised or granular level, for example using smart applications. Several participants (transport sector) highlighted perceptions that the publishing of Open Data holds the potential to improve transparency and thus build public trust, which can be particularly hard to foster during times of change or disruption to established practices.

Secondly, participants from the water and transport sectors discussed the particular value of research data sharing to address the area of emergency planning and response. This was highlighted both in terms of regional emergency planning concerning climate resilience and in terms of responding to health emergencies such as the COVID pandemic.

> The Climate Resilience Demonstrator (CReDO)[11] is a climate change adaptation digital twin project connecting data to improve climate adaptation and resilience. CReDO was cited as a particular example of good practice for data sharing in this area, with infrastructure data used to contribute to regional emergency response models. As the effects of climate change progress, for example in areas such as extreme weather or consequences to human health, it is likely that both the demands for, and benefits of, data sharing to address further emergency use cases is likely to grow.

Case Study 1: Climate Resilience Demonstrator

Additionally, water workshop participants highlighted how effective sharing of infrastructure data fuels the development of new services, tools and industry insights. It enables innovation in flood warning systems which can be integrated into various internal and external platforms to support decision making.

[11] Digital Twin Hub (2024) Climate Resilience Demonstrator (CReDO) https://digitaltwinhub.co.uk/climate-resilience-demonstrator-credo/

Finally, participants from energy and transport sectors highlighted both known and potential benefits to customers or service users stemming from sharing data with researchers. For example, the use of data to facilitate systemic efficiencies and thus pass on cost savings to consumers was raised as one example by an Energy Sector Trade Association. In the transport sector, participants from National Government and Transport Service Providers highlighted examples of where data sharing had informed innovations that produced co-benefits to customers and Transport Service Providers. This included examples of customer benefits in unexpected areas of innovation such as the crossover with information provision concerning public car parking facilities and electric vehicle charger availability.

Transportation workshop participants highlighted benefits to customers including transport link efficiency, safety, and equity. This includes data informing an investment plan to mitigate for underserviced areas, benefitting from proactive planning of transportation connection points, and increased insight as to how people use infrastructure and if they are using it in the intended way (e.g. does outdoor lighting encourage people to walk in the winter).

Benefits to the UK economy

Research participants representing all target sectors, and a range of organisation types, highlighted economic benefits of research data sharing focused on two main themes: innovation and efficiency.

Innovation

Participants identified a broad range of benefits from research data sharing concerning the stimulation of the UK's innovation landscape, both within target sectors and more broadly across the economy. Benefits applicable to all sectors were cited to include factors such as the stimulation of new business models based on technological development (e.g. machine learning or AI), the 'crowd sourcing' or generation of new ideas for system or service improvements by releasing data to research audiences, and the potential for research to contribute innovative models to knotty problems within industry, such as the development of new pricing models (e.g. time of use pricing of energy and water).

One Water Network further highlighted how sharing data with researchers can act as a catalyst for relationship development preceding broader collaboration with research bodies, inviting opportunities for further co-benefits to come to fruition. The Forth-ERA initiative was noted:

Forth Environmental Resilience Array (Forth-ERA)[12] is a digital observatory of the Firth of Forth's entire water catchment run by the University of Stirling. This provides environmental data and analytics with the aim to promote efficient environmental management and regulation, and stimulate net zero business innovation.

Case Study 2: Forth-ERA

Relatedly, but separately, participants from one Energy Network and one Energy Trade Association highlighted that research data sharing could also contribute to the broader innovation environment by supporting appropriate and robust market development. For example, one Energy Network highlighted that data sharing mechanisms, particularly for Open Data, provide one way in which infrastructure networks can support data and related market development rather than adopting a monopoly position concerning data 'ownership'. Separately, an Energy Trade Sector Association highlighted the role of research data sharing in supporting developing markets and value streams in active innovation areas, thus contributing to the development of a more certain and stable environment in which commercial bodies can make investment decisions.

While not stated directly, such conversations highlighted the importance of regulatory monitoring and development in innovation spaces to support infrastructure bodies navigate change to their roles, responsibilities and business models. Without such development, the infrastructure data landscape risks the emergence of areas of market failure, with potential consequences for the wider economy and society. Currently, Ofgem in the energy sector was seen to have taken a lead in regulatory development concerning Open Data provisions for its licensees. Ofwat does not have a license condition for Open Data, but has called on all water companies to unlock the benefits of open data.[13]

Cost savings and efficiencies

In addition to changes within the wider innovation environment highlighted above, participants in the the transport sector particularly highlighted how research data sharing can create benefits to infrastructure bodies (both commercial and otherwise) and wider systems in the form of research and innovations producing efficiencies or other cost saving mechanisms.

For example, research data sharing was suggested to have been used by LNER in the process of developing AI models to manage catering stock on their services, which have

[12] University of Stirling (2024) About Forth Environmental Resilience Array (Forth-ERA) https://stir.ac.uk/about/scotlands-international-environment-centre/forth-environmental-resilience-array/about-forth-era

[13] Ofwat (2023) Ofwat calls on water companies to act now on open data https://www.ofwat.gov.uk/ofwat-calls-on-water-companies-to-act-now-on-open-data

since created over £1 million in cost savings as well as a significant environmental co-benefit concerning the reduction of food waste. A representative from national government also gave the example of one particular area of potential for research data sharing to produce system and operational efficiencies, with environmental co-benefits, which could be supported by expanding research data sharing.

In this case, the participant highlighted the capacity for data sharing to support integrated transport solutions, whereby data sharing could reduce the need for infrastructure buildout. One energy sector participant identified similar research challenges, and associated data sharing opportunities, whereby data sharing supporting system flexibility could significantly reduce demand for network infrastructure buildout while still supporting the systemic changes required to decarbonise the UK's electricity grid[14].

Findings from the transport workshop highlighted provider franchise agreements in places including Manchester and London, which enable data sharing across public transport services (buses, trains, and trams), optimise travel modes and schedules. The franchise business model means routes are secure and without a competitive tendering process, and therefore, data is more openly shared.

Benefits to the organisation involved in data provision

As illustrated in the section above, sharing data with researchers potentially offers infrastructure bodies the opportunity to generate economic benefits within their own organisations by using research findings to operationalise organisation-level efficiency or cost saving measures. Participants also highlighted several other benefits to the data providing organisation generated by research data sharing.

One Transport Service Provider highlighted the opportunity to form research partnerships with data recipients, which can enable both organisations to work together to address industry or sector challenges. For example, this was cited to take place via the formation of project consortia, including funding co-bidding, which may open new income streams for both research and sectoral organisations (e.g. innovation funds). The same Transport Service Provider highlighted opportunities to work with research data recipients to co-work on technology solutions for particular business needs. In the energy sector, one participant representing an Energy Network highlighted how research data provision, and broader research participation, can test proofs of concept and/or pilot initiatives in lower risk environments.

[14] Shakoor, A., Davies, G, and Strbac, G. (2017) 'A roadmap for flexibility services to 2030' https://www.theccc.org.uk/wp-content/uploads/2017/06/Roadmap-for-flexibility-services-to-2030-Poyry-and-Imperial-College-London.pdf

Environmental benefits

Participants from all target sectors – primarily representing a mixture of commercial, regulated monopoly, and government bodies – identified the provision of environmental benefits from research data sharing. Environmental benefits were discussed in four main areas: uncovering new sources of environment-adjacent data, the environment-health nexus, the environment-infrastructure nexus, and environmental compliance.

Participants from water and transport sectors identified how research data sharing in their sectors offered environmental benefits by providing new sources of data to researchers through which environmental information can be inferred. For instance, one participant gave the example of how Transport for London (TfL) Oyster card data on ticketing can be used to reveal patterns about transport usage, human behaviour, and associated patterns in carbon emissions. In the case of the TfL example, this could provide insights into forms of behaviour change which may be incentivised to reduce overall carbon emissions from city transport (e.g. changes to transport mode, type or time of service use).

In the water sector, one non-profit initiative also raised the example of an increase in requests for data concerning particular kinds of pollutant, such as per- and polyfluoroalkyl substances (PFAS). In one example discussed, it was implied that water quality metrics concerning PFAS may be used to infer data about bioaccumulation of these chemicals, for example in animals such as fish.

Relatedly, transport and water sector participants highlighted how infrastructure research data sharing can generate environmental co-benefits that impact human health and wellbeing. For example, the use of water data to inform research on hazardous chemicals in the wider ecosystem can also be used in human health research and potentially related policymaking (e.g. the consideration of reductions or preventions to certain substances used in manufacturing). An example was highlighted in the water workshop of research in drinking water quality supporting evidence-based policy making and tightening lead standards. Additionally, one transport sector participant also discussed the use of their (open) data on congestion in research, conducted by a local university, regarding air quality and transport emissions. The visibility and accessibility of open data was highlighted to be particularly valuable in this case study as it is transparently available for scrutiny by anyone.

The co-benefits of infrastructure research data sharing were highlighted at the meeting point of infrastructure transitions and ecological impact. One key example of this was provided by a water sector participant representing a non-profit network, which

highlighted how water and/or maritime data can be a useful input to assessing the environmental impacts of energy infrastructure such as offshore windfarms. This contribution highlights the capacity of research data sharing to contribute to assessing complex problems with nuance, for example concerning the wider ecological impacts of carbon reduction technologies. Additionally, the water workshop participants highlighted that the effective sharing of infrastructure data has resulted in positive impact when it is able to integrate variables such as population growth, climate change scenarios, industrial development, and green transition for long-term planning.

Barriers to data being shared for research

The above sections address benefits to data sharing for research. This section explores barriers that may prevent data being shared with researchers and research organisations.

Survey

Key barriers to data sharing across all survey participants

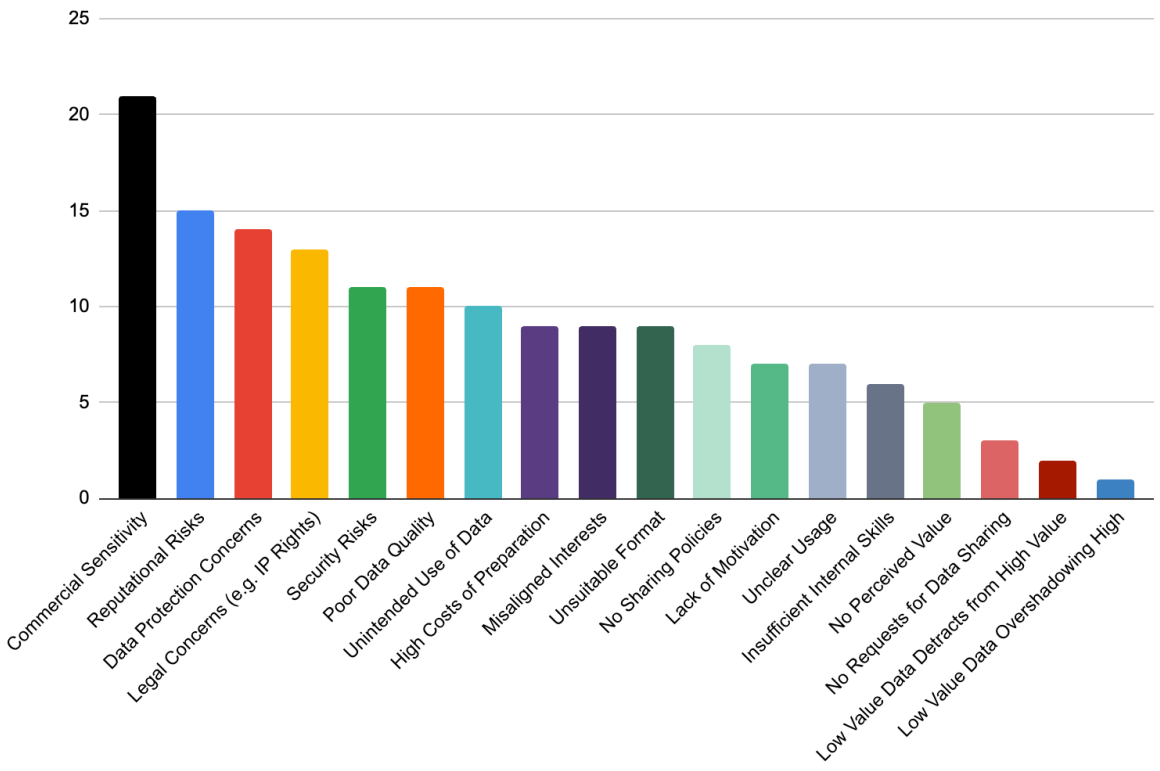

Chart 4: Barriers in relation to organisation sharing data for research purposes

As part of the survey, participants were asked to select all barriers relevant to their organisation. As the chart above shows, the most commonly chosen barriers emerged as: commercial sensitivity, reputational risk and data protection risks.

The most commonly cited barrier was that of data being commercially sensitive – this was selected 21 times. This showed concern within organisations about protecting data that is perceived to hold strategic or financial value. The second most commonly cited barrier was data will be used in a way that risks reputational damage to the organisation, which appeared 15 times. Many organisations expressed concern that sharing data could harm the organisation's reputation if it is misused or misinterpreted. Concerns about data protection was chosen 14 times, in particular highlighting that compliance with data protection laws and a focus on privacy and regulatory adherence is a significant worry for organisations.

Legal concerns (e.g. intellectual property rights) was selected 13 times, security risks was chosen 11 times, and concerns about data quality was chosen 11 times, as participants expressed hesitancy to share data due to fears it may not meet the required standards.

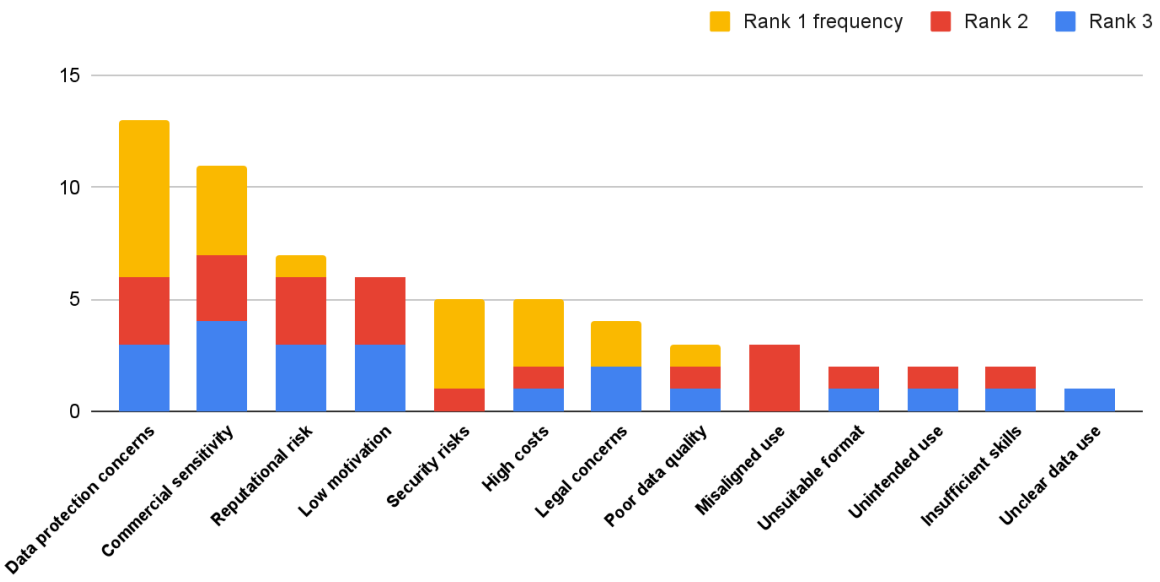

Chart 5: Barriers to data sharing frequency

When participants prioritised their top three key barriers, a similar ranking as above is evident. Data protection concerns, commercial sensitivity, and reputational risk, have the highest combined frequencies across Rank 1, Rank 2, and Rank 3. This reinforces the suggestion these are widely recognised challenges across different contexts.

Barriers including data protection concerns and commercial sensitivity were frequently ranked first, highlighting their perceived importance. Others, such as legal concerns and

high costs, appear more consistently across lower ranks, indicating they are secondary but still significant challenges. A wide range of barriers are represented, from technical (e.g., unsuitable format) to organisational (e.g., lack of motivation) and trust-related issues (e.g., reputational risk).

Barriers such as unclear data use and insufficient skills have lower frequencies, suggesting they may be less widespread but still critical in specific contexts or organisations.

Top barriers by organisation type

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

Table 2: Barriers by organisation type: How different types of organisations perceive barriers.

The survey revealed differences by type of organisation:

- Government bodies were more likely to emphasise legal, security, and reputational concerns. For these organisations, balancing transparency with operational security is likely to be a central challenge.
- Commercial organisations focus on competitive risks, proprietary nature of their data and data protection.
- Academia's top three barriers related to the cost in terms of time, money and other resources. This includes getting the data ready or sharing, lack of organisational motivation to share data and the potential risks of reputational damage to the organisation.

Top barriers by sector type

We also looked at differences in the top barriers between organisations working in the three sectors of interest (energy, transport, and water). Three organisations that responded to our survey focus on energy; two focus on transport; nine focus on water; and nine work across multiple sectors.

Overall, there were fewer clear-cut differences between sectors than between organisation types, though a few distinctions are worthy of note.

For organisations working within the energy and water sectors, and those that work across multiple sectors, **concern about data protectio**n was frequently cited as a top-three barrier. Interestingly neither of the two organisations that are focused on transport highlighted this as a top-three concern, perhaps suggesting that such organisations more commonly work with non-personal data (though the sample size is very small).

Data being commercially sensitive was frequently selected as one of the top three barriers by water organisations and those that work across multiple sectors; this was also mentioned by one energy company and one water company.

Water companies were more likely than those in other sectors to flag **data being used in a way that risks reputational damage to the organisation** as a substantial concern. This may reflect recent negative press coverage of water companies in relation to sewage discharges.

Meanwhile the **cost of getting data ready for sharing** (in terms of time, money and other resources) came out as a highly-prioritised barrier for energy organisations, being rated the number one barrier to data sharing by two of the three energy organisations.

It was also the number one barrier for one transport organisation. This barrier didn't appear in the top three for any organisation focused on the water sector.

Interviews and workshops

Feedback from interviews and workshops supports many of the key issues highlighted in the survey, with numerous and often cross-cutting barriers to data sharing identified.

> 'In any organisation, the organisation is dealing with a number of issues on a number of fronts at any given time. Taking time out to work out how to break down barriers to share data externally just isn't a priority. It's not at the bottom of the list. It's not even on the list.' *Energy Network*

Legal barriers

Research participants highlighted a range of barriers to data sharing with publicly funded researchers that stem from legal concerns. This section examines the following themes: challenges related to legal mechanisms enabling data sharing, privacy concerns, legislative or regulatory grey areas, and identifying who can make decisions to share data.

Legal mechanisms enabling data sharing

Firstly, participants from all sectors identified the legal mechanisms supporting data sharing – variously identified as agreements, contracts, and licences – as forming frequent and significant barriers to data sharing. The cost of establishing and interpreting legal arrangements was raised most frequently, with costs cited to stem most prominently from resourcing formal legal review and secondarily from internal non-legal resourcing required to conduct activities such as data protection compliance checks (often conducted on a bespoke basis therefore can be lengthy). Participants from the transport sector further identified a related cultural tendency for universities to request changes to standardised legal texts, even where data providers have attempted to ensure standardisation for particular datasets or data products. In both cases, change requests were cited to increase costs both for sharing and consuming data.

> 'Data sharing agreements are a huge resource drain, they are all point to point and it is impossible to scale'. *Anonymous*

Participants from the water and transport sectors additionally identified non-disclosure agreements (NDAs) as a form of data sharing agreement with particularly restrictive tendencies, which have been challenged on the basis of impeding research and/or research collaboration efforts. For example, participants in the transport workshop further identified instances where NDAs have prevented data being used in

combination, thus restricting systems-level research. Such clauses were identified to be particularly problematic when researchers join pre-existing agreements – notably those which may not have been designed for a research audience with associated knowledge of typical research activities – with no or minimal capacity for adjustment. This finding suggests that further research into the user needs of researchers who consume data could deliver value in supporting the creation of constructive legal mechanisms for data sharing.

Even in cases where the legal mechanisms were considered functional, participants from all sectors raised concerns related to the enforceability of such mechanisms. In all sectors, instances where data may be shared internationally, either in its raw format or as derivative data (e.g. research results data, particularly where published openly), were identified as particularly high risk. This is difficult to navigate in an academic environment where international collaboration is strongly encouraged by sectoral metrics used to assess research success (e.g. the Research Excellence Framework[15]).

Concerns about international data sharing and/or processing also relate to barriers discussed in the security section as related to human error and/or data governance knowledge gaps in a sector where researchers have a high tendency to be mobile. For example, one Transport Service Provider identified instances in which students were known to have downloaded data onto a personal laptop that was then transported and used overseas. Such findings suggest that knowledge and enforcement of data governance practices could be improved among the research community and must cover all types of researchers using infrastructure data (e.g. including students, who were identified as particularly high risk due to lower skill and knowledge levels).

Concerns regarding both traceability and enforceability of licence terms were also particularly prevalent, governing the onward sharing of either raw or derived data. Several organisations expressed explicit reluctance to share data in a manner permitting any form of onward sharing due to perceptions that this can result in a loss of control over their data assets. This finding is particularly significant in an academic environment in which the publishing of both input and results data alongside academic papers is increasingly encouraged, or may even be mandated, by funders (e.g. Open Data publishing commitments by the European Commission[16]). Participants from the water workshop further highlighted that some data breaches can go 'unnoticed' and lead to data misuse that is unintentional. For example, particularly where NDAs are complex, it may be possible for a researcher to unknowingly share protected data with

[15] Research Excellence Framework (no date) 'about' https://2029.ref.ac.uk/about/
[16] European Commission (no date) 'Open Access' https://research-and-innovation.ec.europa.eu/strategy/strategy-research-and-innovation/our-digital-future/open-science/open-access_en#:~:text=Open%20Research%20Europe%20is%20a,everyone%20to%20access%20the%20results.

non-authorised parties (e.g. individuals within a university research group who are not part of a specific consortium agreement).

Privacy

Privacy or personal data sensitivity was identified in the literature review as a prominent type of sensitivity that is well known to infrastructure data providers. This finding was replicated by interview and workshop data, indicating broad participant knowledge of this sensitivity type as an active consideration when assessing opportunities for research data sharing. However, participant responses also indicated that handling privacy and personal data sensitivities remained an evolving challenge as the data and technical landscape of infrastructures continues to change. For example, participants from the transport workshop indicated some lack of clarity regarding how certain data on private vehicles should be treated. As demand for this data increases, for example as related to electric vehicle charging and mobile charging demand, it is likely that further challenges to the interpretation and application of General Data Protection Regulation (GDPR) will be identified.

Legislative and regulatory interpretation

As discussed in the literature review, data governance mechanisms continue to evolve at both national and sectoral levels, with further activity anticipated in 2025 (e.g. passage of the Data Bill[17] and any secondary legislation). While sentiment towards landscape evolution was broadly encouraging, participants across all sectors identified a growing risk associated with differential interpretation of data legislation and regulation, leading to divergent approaches to data sharing. For example, participants from the energy sector gave the example of how Section 105 of the Utilities Act 2000[18] has been used in the past by licensed bodies as a reason to restrict data sharing. One participant discussed how this may come into some conflict with new regulatory obligations[19] surrounding Open Data. Participants from the transport sector also identified divergent interpretations of GDPR and related concepts, such as consent, as barriers to sharing personal data which may cause infrastructure bodies to take a more precautionary approach.

Decision-making on sharing

[17] UK Parliament (2024) 'Data (Use and Access) Bill https://bills.parliament.uk/bills/3825

[18] UK Government (2000) 'Utilities Act' Section 105 https://www.legislation.gov.uk/ukpga/2000/27/section/105

[19] Ofgem (2023) 'Decision on updates to Data Best Practice Guidance and Digitalisation Strategy and Action Plan Guidance' https://ofgem.gov.uk/decision/decision-updates-data-best-practice-guidance-and-digitalisation-strategy-and-action-plan-guidance

Finally, participants (particularly from the transport sector) identified frequent difficulties in identifying who in their organisation holds responsibility for deciding what data can be shared and how. Duplication or gaps in responsibilities were inferred in this space, as well as an overall concern regarding the potential impacts that getting a decision 'wrong' could bring on their organisation or team. For example, one Local Government body with responsibility for local transport services identified significant differences across projects in which some have enabled easy sharing of data – supported by clear governance and sign off processes – while some have not. In some cases, this issue was present even for comparable datasets, suggesting that internal process and governance sometimes presented more of a barrier to sharing than any other aspect of dataset sensitivity or compliance.

One organisation from the energy sector explicitly linked this type of issues to cultural factors (further discussed in a subsequent section) as much as legal or governance factors, as illustrated through the quote below:

> 'I feel like our corporate DNA is very much about boots on the ground and people dealing with physical assets in the real world. We've never been a technology sort of leader nor have most utilities. Our management culture is generally people who come through that background primarily closer to where the value is created rather than the back office. There's very much a tendency to try to deprioritize spending on kind of back office work and focus investment in actual tangible assets' (Energy Network)

Security barriers

Participants from all sectors, particularly representing large commercial organisations (e.g. utilities networks, transport providers) and public sector bodies, identified security as a significant and multi-faceted consideration that can present barriers to research data sharing. However, data collected for this project implicitly revealed a highly divergent understanding of how the concept of security is interpreted across infrastructures and organisation types, as well as variable prioritisation of different threat types, and mitigation areas.

Across the project dataset, participants broadly drew attention to a perception that national security in the domain of infrastructure data was becoming a growing policy priority. However, participants also clearly indicated a perception that this may not be entirely compatible with recent changes pertaining to Open Data publishing or the

concept of 'presumed open'[20]. One participant representing a regulator discussed this as a point of active internal debate at present, highlighting that not sharing data due to security concerns is not a 'neutral' option, but rather one which may reduce societal capacity to create positive change, for example regarding the environmental impacts of key infrastructures. Demand for improved policy and regulatory join-up, including across different infrastructure sectors, was accordingly identified.

The following sub-sections discuss themes related to priority threats and sensitivity reduction mechanisms within the landscape of security barriers.

Priority threats

Security is a complex concept which can be interpreted differently according to the particular risks or needs associated with different dimensions of threat. Data collected for this project identified two key areas of threat which actively inform practices of research data sharing in infrastructures: cyber threats and national security threats.

Cybersecurity threats were identified mainly by energy and water sector participants as a growing area of concern. While not discussed in depth, participants inferred that cyber threats were an active area of consideration as the data load of these infrastructures increases. For some participants, assessment of cybersecurity threats and mitigations was considered to be an element of technical and process development within the organisation. For example, a participant representing a Water Network discussed how previous data-sharing practices of sending data as an email attachment are now replaceable with more 'modern' systems reducing the potential for error, misuse, and other breaches (e.g. sharing via insecure internet connections). However, other participants highlighted inconsistent data maturity among data users as a sometimes unexpected barrier to sharing data in more secure ways. For example, one participant gave the example of a large data request having to be delivered to clients via an Excel file as the client was unable to use the company's API format (Electric Vehicle Charging Provider).

Further research into this area was broadly inferred to be required in order to ensure continued security of data held within infrastructure organisations, held by researchers, and data passing through sharing mechanisms between organisations. While such research is developing in generalised terms, there would be benefit to conducting specific research into the socio-technical context of data sharing between infrastructure

[20] Ofgem (2023) 'Data Best Practice Guidance' Section 11 https://www.ofgem.gov.uk/sites/default/files/2023-08/Track%20Changes%20Data%20Best%20Practice%20Guidance%20v1.2%20for%20consultation.pdf

bodies and publicly funded researchers, to better pinpoint areas of priority consideration.

National security concerns were also discussed frequently by participants from all sectors. Human error and dataset combination were two threat types that were notably identified. When discussing human error, organisations indicated that technology and governance can both play a role in ensuring security. For example, one Water Network discussed how sharing data through APIs with designated identities was less vulnerable to human error than emailing spreadsheets (a common practice to date). Participants also drew attention to how data governance skills, knowledge, and organisational processes can reduce instances of incorrect sharing (also discussed in the legal barriers section).

As infrastructure data is increasingly gathered and shared, participants discussed how dataset combination could lead to heightened security risks, particularly where this is done in novel ways that may expose unintended information (e.g. about adjacent assets or sectors), is shared beyond the UK jurisdiction, or is shared with actors who may not be adequately screened (e.g. inadequate background checks, inconsistencies in background checks outside the UK, hostile actors posing as academics). Research was identified as a particular site of dataset combination risks as researchers tend to use data in novel ways and cross-border sharing represents scholarly norm. While this was previously identified as a benefit to research data sharing, it also potentially indicates emergent responsibilities for the research community to consider the security implications of findings where data is combined in new ways. Further research is necessary to understand this area in more depth and to understand how challenges affect academic freedoms as well as national security concerns.

Sensitivity reduction mechanisms

Participants from the water sector workshop actively discussed where sensitivity reduction mechanisms could potentially be introduced to support data sharing while preserving national security. The development of secure research environments, possibly similar to those applied in the health sector, was raised as a particular point of interest, albeit one which may invoke significant time and financial cost to operationalise.

> ‘Processes to get access to more secure, sensitive data can be very lengthy and time-consuming (we found this in health data)’ (Regulator)

However, participants also identified that some common sensitivity reduction mechanisms – in particular data redaction – can in themselves form a barrier to research data sharing by reducing the utility of the underlying dataset. This was identified by one public sector body as a particular challenge when handling free text fields. Another participant also associated the publishing of redacted data as a reputational risk, citing where such datasets had resulted in an overall increase in requests covered by freedom of information (FOI) or environmental information regulations (EIR). The potential reputational risks of sharing redacted data can therefore lead to an overall increased reticence to share any data, for fear of creating false perceptions that organisations are withholding potentially significant or damaging information.

Factors such as the above are linked to a finding related to an overall culture of risk aversion within infrastructure sectors, as discussed in the cultural barriers section.

Commercial barriers

Commercial barriers to research data sharing were identified by a diverse range of participants representing all target infrastructure sectors and a spread of organisation types, with a slight focus on commercial (profit making, including regulated) organisations. Internal costs, commercial sensitivity, and skills/knowledge shortages were identified as three core sub-themes discussed below.

Internal costs

Participants broadly highlighted that sharing data is always associated with some amount of internal cost. While certain innovations (e.g. standardised data products or services) can reduce these costs, it is never a 'free' service and costs must be met in order to incentivise data sharing. Labour costs were identified as the most significant resourcing barrier which can involve several different areas of an organisation (e.g. technical, operational, sales, legal, customer service) and should not be interpreted as a solely 'technical' cost. Participant responses highlighted that a significant portion of this internal cost may be front-loaded (e.g. dataset quality assurance, maintenance, risk assessment, compliance assessment, technical investment, conversion from manual processes, etc) and thus occur prior to the benefits of data sharing being obtained. This can act as an exclusionary factor, particularly for smaller, time-poor, or cash-poor organisations. It can also be difficult to establish the business value for organisations to share data at the early stages of sharing preparation, particularly where market needs are not clearly established.

Participants, particularly from the water and energy sectors, also identified several common practices by researchers that were cited to make data requests harder to

service. These include: submitting poorly worded or ambiguous requests[21], submitting a large volume of different requests (particularly common by students), and requesting complex or non-standardised data (which may require complex negotiations and time to make shareable). While not discussed in depth, it was discussed that a research cloud environment has the potential to filter or reduce some of these challenges by providing education to the research community on good practices for data requests, potentially triaging requests (although this generated disagreement), and supporting data users to articulate their needs.

Commercial sensitivity

Commercial sensitivity of infrastructure data itself was cited as a barrier to sharing data with researchers by organisations in all target sectors. However, there was significant inconsistency in how this barrier was expressed in relation to a parallel benefit noted previously; that researchers working in non-commercial environments may be able to use sensitive data in ways which supports access to other benefits (e.g. social, environmental) without jeopardising the commercial entity.

Several organisations from the transport sector raised concerns that data requests potentially endanger core intellectual property, for example data about more profitable routes or service frequencies, which underpins their business models. Unintentional revealing of this information, either to researchers or by proxy via research publications, was deemed to present undesirable risks to the competitive landscape or forms of 'exploitation' (Transport Service Provider) that do not match the intentions or defined purpose of sharing data. In some cases, participants added that this kind of activity would be hard to trace or prevent, due to the limitations of legal protections previously discussed. While unable to provide a specific example, one participant further raised that such competitive threats could potentially disincentivise investors, leading to a negative cycle of commercial impacts on the future business.

Water sector organisations were also cited to produce barriers to data sharing 'by default' (Water Sector Non-profit Organisation) for fear of exposing information to current or future potential competitors. It was noted that this can create a culture of risk aversion, preventing even low risk data from being shared. The same organisation drew attention to perceptions within the sector that academics themselves could potentially act as competitors, however the basis of this perception was not explored in depth. Further research into how data is used within academia, as well as the establishment of clear purpose-based barriers (e.g. data use for non-traditional

[21] Noting that this may also be the product of immature technical landscapes in which comparable datasets are not consistently labelled or terminology remains non-standardised.

activities such as consulting or spin-off innovations), may help reduce sectoral risk perceptions in this area.

Skills and knowledge shortages

Participants drew attention to two key barriers preventing organisations from sharing data with researchers on a wider basis. Cost and resourcing issues highlighted above can be exacerbated by a lack of access to adequate skills in house. In addition to technical skills required to build and maintain datasets or data services, skills shortages were also identified to impact other roles, for example those required to conduct compliance tasks, risk assessments, or conduct market research to understand infrastructure data use. One organisation from the water sector highlighted that obtaining such skills requires investment as they are often new for the business, and this can be difficult for regulated entities to justify if the benefits of data sharing are not clearly articulated in service of the company's regulatory imperative (e.g. customer service).

Participants from the transport and water sectors discussed sectoral knowledge shortages concerning 'market demand' for infrastructure data by publicly funded researchers as well as their particular user needs as a barrier. This was cited to add an additional barrier to effectively and efficiently fulfilling data requests. For example, one participant representing a public body discussed how their organisation had failed to identify particular types of customers who had become frequent data users. Without such knowledge, data sharing systems can easily be built in ways that fail to serve potentially valuable customer bases and can therefore exclude users who may otherwise provide revenue streams. As identified above, a lack of knowledge about market demand can more generally make it difficult to determine the value of proactively publishing data. This uncertainty can lead to additional hesitancy in data sharing and suggests that further research into the academic market for infrastructure data would be valuable at a cross-sectoral level.

> 'Not everyone sees the full benefit of end-to-end. Compared with [the] NHS [and health data sharing], where there's been a real weight of pressure to make data available because it is so valuable, there hasn't been the same pressure from academia to release water data.' (Regulator)

Participants from the transport sector workshop indicated that improved transparency at a sectoral level regarding how researchers use infrastructure data could potentially address some of the challenges outlined above. While not explicitly discussed, it was implied that there may be an imperative for research institutions, and the research

community, to take a more proactive approach in this space in terms of clearly communicating back to data providers. This sentiment was echoed by an Energy Sector Trade Association representative, who called attention to a lack of skills (particularly communication) and wider cultural practices supporting feedback loops between academia and industry. This participant also drew attention to a new area of good practice perceived to be supported by requirements for academic-industry knowledge exchange being built into innovation funding pots. Further research into the skills and institutional supports required to support improved communications around data access and use is likely to be necessary in order to support and share areas of best practice going forwards.

Cultural barriers

Cultural barriers to research data sharing were noted by participants from all sectors as presenting notable – but often less visible – barriers to sharing. Participants consistently associated cultural barriers to data sharing with the concept of 'trust', or the current lack thereof, in research data sharing relationships.

> 'The failure to take a partnership approach discussed stems from a lack of trust, knowledge and early engagement.' (Water Infrastructure Provider)

The concept of trust can be difficult to explore as it is contextually specific, complex to quantify, and highly relational[22]. The need for systemic mechanisms to address trust in data sharing at the sectoral or use case level was identified by several participants from energy and water sectors who advocated for the building of Trust Frameworks in infrastructure sectors to expedite processes and cultural acceptance of more standardised, trusted data sharing. The application of Trust Frameworks[23] in the domain of research data sharing has not yet been widely explored but may be worth further consideration in relation to the building of a research data cloud.

> 'There are a range of social factors that inhibit sharing, which stem from not trust[ing] people and technology.' (Anonymous)

[22] Trust is not built alone but through relationships and networks. Accordingly, perceptions of trust can change significantly in response to events or changes in other parts of a network.

[23] Trust Frameworks operate at the sector level to establish and maintain a light layer of collaboratively developed and governed interventions which create the foundations of a trusted data sharing ecosystem. This includes, but is not limited to: identity management, assured Open Data publishing (baseline), and maintenance of a library of common principles, definitions, and Open Standards. Schemes co-develop and maintain the additional specific rules, governance, and enabling technology necessary to facilitate the exchange of data for a particular use case (or set of related use cases) among a defined set of participating actors. Icebreaker One (2024) 'Definition: Trust Framework' https://ib1.org/definitions/trust-framework/ and Icebreaker One (2024) 'Definition: Scheme' https://ib1.org/definitions/scheme

Discussion of cultural barriers to research data sharing focused on three main themes, which are explored in more depth below: reputational concerns, risk aversion, and data maturity.

Reputational concerns

Participants from all sectors highlighted perceived reputational barriers to sharing data, however the rationales given for such concerns - and the degree to which they present a barrier to research data sharing – varied. Concerns around the quality of data were cited as a significant barrier to sharing within the transport workshop. In other interviews, participants associated potential for misinterpretation with the sharing of poor quality data, which could also heighten related risks such as the reduction of data user and public trust in the organisation as a result. Damage to public trust as a result of data sharing more generally emerged as a notable theme among water sector workshop and interview participants in particular. One participant representing a Water Network further identified the potential for poor quality data to create unintended negative consequences for public health, giving the example of how quality variations in data used to calculate the dosage of chemicals for water quality and sanity could lead to over or under dosing, thus exposing water users to unsafe levels of chemical or biological pollutants.

Even in cases where data shared is of good quality, a variety of organisations raised concerns about unintended public and/or press attention, and associated potential for interpretation of the data without adequate context. One participant from the energy sector identified the example of water sector data releases, particularly concerning spills, becoming 'front page news' (Energy Network) and inferred that similar situations may occur unexpectedly for standard data releases in other infrastructure sectors. One participant from the transport sector highlighted that it can be difficult to adequately brief researchers on some of the context surrounding their data due to the time and resourcing costs associated with this activity (stated to take at least an hour of largely one to one conversation conducted without charge and without any guarantee that nuance would be acknowledged in research outputs).

> '[Organisations] have tried to be more open, then they have been penalised as a result of that. The risk is that people take data and information out of context that makes them look bad. Context must be provided.' (Energy Network)

Other participants also highlighted that such issues can become more complex if adjacent context or adjacent data is too sensitive to share, leading to researchers building an incomplete picture of infrastructure ecosystems.

Risk aversion

A culture of risk aversion to data sharing was noted by participants representing the full sample of organisations participating in this research project. Discussions revealed two related components to risk aversion: a lack of knowledge and expertise to appropriately assess risks, and the (recent) shift from asset-based to asset-and-data organisations.

Firstly, participants clearly identified that risk assessments concerning data sharing were performed in an inconsistent manner. This issue was identified to occur across sectors and even within single (particularly large) organisations. The key reason identified was a generalised culture of risk aversion which has been particularly fueled by a lack of expertise and knowledge regarding data-specific risks.

Summary of barrier identified by Water Network at the Water sector workshop:

> Different organisations, and even different individuals within the same organisation, may not have the same level of expertise in understanding data-sharing regulations to assess risk.

A lack of expertise at the senior management level was noted as a particular concern by several participants, who highlighted the capacity of high level actors to block data sharing practices without full knowledge of their context, and/or to create precedents against future instances of data sharing. In two cases, participants also highlighted the influence of investors on organisational priorities, discussing that investors can have a more 'traditional' view of infrastructure business priorities which can inadvertently deprioritise spend related to digitalisation and data strategy.

> 'Our investors don't care what [Energy Network] does around data – it's not even a consideration. They take it for granted that we will do the right thing, but that's just kind of a corporate hygiene team as far as they're concerned. They don't see investment in data as something [Energy Network] should be doing – they see [Energy Network] needs to be investing in physical assets.' (Energy Network)

These remarks highlight the ongoing nature of infrastructure sectors' digital transitions, with organisations shifting away from seeing themselves as purely asset-based bodies and towards an asset-and-data hybrid operational model. While in certain sectors, such as energy, governance expectations of such changes have been reinforced by the

introduction of new data and digital licence conditions, the pace and uniformity of the change still remains highly uneven.

Technical barriers

While participants from all sectors and a variety of organisation types discussed technical barriers to research data sharing, data collected for this project broadly inferred that technical barriers were of less concern than other barrier types discussed in this report. Discussions focused on two main areas of technical challenge: accessibility and interoperability.

Accessibility

Data discoverability and good metadata provision were identified as two key areas in which research data sharing is currently impeded by accessibility concerns. Participants from the water sector workshop particularly identified how there can be a disconnect between data that is collected and available to share, versus data which researchers are aware is available. Such gaps were identified to lead to missed opportunities for sharing useful data and/or duplication of requests for data types which already exist.

Relatedly, participants identified that poor metadata provision can prompt scenarios in which researchers are unable to actually use data, even when they are aware it exists, as metadata is insufficiently descriptive for researchers to ascertain whether the data is useful. Intervention by governance bodies to promote consistent approaches to discoverability and metadata provision at the sector level may act as useful catalysts to addressing such barriers and improving researcher access to data which is already in existence and shareable.

Interoperability

Participants further highlighted a lack of interoperability as a barrier to research data use that is particularly significant for projects where data must be combined (e.g. to produce system level or cross-geography insights). One Regulator highlighted how the influence of different technology and data services suppliers across infrastructure sectors can have a negative impact on data interoperability due to divergent supplier practices, some of which are protected due to commercial value.

Participants from the transport and water sectors highlighted how a lack of standardisation, including but not limited to data formats and terminologies, can further impede interoperability and dataset integration. This was anecdotally identified as a particular barrier to dataset combination across different geographic areas (e.g. utility networks or transport provider areas) and across sectors (e.g. for cross-sector use cases such as emergency planning). Participants in the water sector workshop

particularly encouraged further alignment between governance bodies across infrastructure in order to improve data interoperability. In this sector it was noted that water companies are being asked, for example, to submit data to multiple different platforms that use different technologies (such as Stream and NUAR), risking duplication of effort. The future value of such work to machine readability and automated data ingestion was also highlighted, though not discussed in depth.

Ranked solutions: workshops

The following section discusses potential solutions to addressing barriers to research data sharing. Although data on this item was collected from across workshop and interview methods, data from interviews was sparse. Where possible, discussion of potential solutions generated through interviews have been woven into the discussion of barriers in the preceding section. The below table identifies the top five solutions to research data sharing challenges identified through workshops in the water and transport sectors.

To produce these results, attendees in the workshop were split into two groups and discussions were facilitated across three sessions. The third session aimed to identify possible solutions to the identified barriers. After solutions were discussed and refined by the group, they were placed in a ranked voting survey, and sent to participants to rank their top solutions from the workshop.

	Water workshop	**Transport workshop**
1	Development of regulatory landscape supporting data sharing across sectors	Researchers being more clear on how data should be used
2	Central repository for research requirements and required datasets	A framework that delivers confidence to share data (e.g. Feasibility, Relevance, Effectiveness, Sustainability, and Harm FRESH model[24])
3	Data sharing agreement templates / modular licensing	Providing a broad schema for all data to be shared (such as Model for Underground Data Definition and Integration - MUDDI[25])
4	Agreement on cross sector standards	Make sure data owners understand the value to them

[24] Rail Data Marketplace (2024) A FRESH perspective: the new era in data-driven decision making https://raildata.org.uk/blog/blogDetails/fresh-perspective

[25] Open Geospatial Consortium (no date) MUDDI https://www.ogc.org/publications/standard/muddi/

5	Access control - secure environment for data sharing to ensure security issues are addressed	Understanding what concerns of those whose data is being shared are and showcasing possible mitigations/involving stakeholders/specific groups of stakeholders

Table 3: top ranked solutions generated by water and transport workshop participants

Recommendations

Drawing from analysis conducted across this research, the following recommendations have been generated with the goal of improving research data sharing between infrastructure data providers and publicly funded researchers.

1. Data providers and those who facilitate data sharing (e.g. regulators, aggregators) **must** invest in co-designing appropriate and open governance for data sharing to foster data accessibility and interoperability within and across sectors. Good data governance **must** be based on the principles of transparency, accountability, engagement and responsiveness in order to address legal, policy, security, and communications needs in addition to technical matters.
2. Data sharing initiatives between infrastructure data providers and the research community **must** build on existing initiatives in order to promote interoperability, prepare for many-to-many data sharing, and reduce the burden on data providers and data users.
3. National data sharing initiatives **must** be designed with cohesion (with other initiatives) flexibility, extensibility and capacity-building in mind. This requires the ability to guide less digitally mature organisations through a gradual on-ramp, such as supporting the transition from manual to API data sharing.
4. Research and innovation funders **should** support research that makes clear and advocates for the practical and strategic benefits of sharing national infrastructure data with researchers. This should explicitly account for nuance concerning the sharing of different data types (e.g. business, operational, and geospatial) or sensitivities (e.g. personal, commercial, security) and provide frameworks or guidance on the translation of research outcomes and benefits back to data providing organisations.
5. Research funders **should** invest in documenting research-based best-practice case studies and use cases for commercial organisations to better understand what data researchers are looking for, and in what formats, to enable data providers to prioritise service development.

6. The research community **should** actively engage with commercial data providers to embed researchers within organisations to build a collaborative approach, which will ease the administrative burden on data providers

Conclusion

This pilot study has looked at the requirements and impact of improved sharing of national infrastructure data with publicly funded researchers, focusing on energy, water and transportation sectors. The report has been shaped by findings collected from a literature review, survey, workshops and interviews conducted between August and December 2024. These findings have been presented in four main sections: current data sharing practices, benefits of sharing infrastructure data, barriers to sharing infrastructure data, and a series of recommendations. Based on the analysis conducted across the research, these recommendations have been designed to enhance data sharing between infrastructure data providers and publicly funded researchers.

Our findings across the three sectors suggest that data sharing with researchers is diverse. There are varying levels of data maturity across organisations, some of which never share data with researchers. Despite the rise of standardised portals and licences as a result of Open Data publishing, there is a clear demand from infrastructure data providers for more standardised, interoperable, low cost and scalable legal mechanisms to share data.

Our analysis revealed that data providers understand a range of benefits of data sharing but in many cases this understanding is hypothetical. Few organisations were able to demonstrate these benefits through concrete quantitative or qualitative evidence. In situations where benefits have been measured, direct financial returns for data providers have been a priority. A significant gap exists in understanding how system-wide benefits translate into tangible impacts for different types of infrastructure bodies.

There were significant barriers to sharing infrastructure data highlighted across all areas of the research, which were mapped into five categories: legal, security, commercial, cultural and technical. However, of the barriers presented, technical issues were not deemed a substantial area of concern for participants. The key barriers that were highlighted varied between organisation types, but the cultural appetite for data sharing was overall identified to be hesitant, with communications, reputation and trust based concerns playing an influential and prohibitive role. In some instances, security barriers were seen to be in conflict to publishing approaches in academia. Moving forward, collaborative efforts by researchers, funders, data providers, and governance bodies will be necessary to support data sharing that is backed by robust and open governance mechanisms supporting appropriate data use.

Appendices

Appendix 1: Glossary

Term	Acronym	Definition
Data Provider	–	Organisation which provides data to a researcher / research facility. This includes both those who hold the data and those that facilitate the sharing of data, eg regulators
Data Infrastructure for National Infrastructure	DINI	Project within DAFNI exploring the challenges and opportunities in data sharing within the domain of national infrastructure systems research.
Data Analytics Facility for National Infrastructure	DAFNI	Programme located within the Scientific Computing Department, within the Science and Technology Facilities Council (STFC)
Open Data[26]		Can be used by anyone for anything for free [any-to-any]
Shared Data[27]		Data with a preemptive licence [many-to-many, some-to-many] (e.g. 'data as a service' that can be used with certain restrictions.)
Closed Data[28]		Requires, if shared, a user-specific custom licence or contract for use [some-to-some or none]
Restricted Data		Term adopted by the DAFN-DINI team to refer to a combination of Shared and Closed data types (as defined above)
Presumed open		Data which must be made available for all people to use, unless the organisation responsible for handling the data provides evidence of a specific reason for needing to reduce its availability (defined in Ofgem (2021) Data Best Practice Guidance)
Ofgem		Regulator for the energy sector in Great Britain (England, Scotland, Wales)
Ofwat		Regulator for the water sector in England and Wales
Trust Framework		Trust Frameworks operate at the sector level to establish and maintain a light layer of collaboratively developed and governed

[26] https://ib1.org/open-shared-closed/
[27] ibid
[28] ibid

		interventions which create the foundations of a trusted data sharing ecosystem. This includes, but is not limited to: identity management, assured Open Data publishing (baseline), and maintenance of a library of common principles, definitions, and Open Standards. Schemes co-develop and maintain the additional specific rules, governance, and enabling technology necessary to facilitate the exchange of data for a particular use case (or set of related use cases) among a defined set of participating actors.

Appendix 2: Case studies

The following case studies were highlighted by interview participants as examples of good practice supporting research data sharing. Case studies span research programmes which have involved data sharing between infrastructure bodies and publicly funded researchers, as well as examples of specific practices and technologies which could support the expansion of research data sharing in this space.

#	Case Study	Further Information
1	The Climate Resilience Demonstrator (CReDO) is a climate change adaptation digital twin project connecting data to improve climate adaptation and resilience. CReDO was cited as a particular example of good practice for data sharing in this area, with infrastructure data used to contribute to regional emergency response models.	Digital Twin Hub (no date) Climate Resilience Demonstrator (CReDO) https://digitaltwinhub.co.uk/climate-resilience-demonstrator-credo/
2	Forth Environmental Resilience Array (ERA) project with Stirling University: case study of successful data sharing between the water sector and academia, with the aim to set up a digital twin of the Firth of Forth catchment area.	https://www.stir.ac.uk/about/scotlands-international-environment-centre/forth-environmental-resilience-array/about-forth-era/
3	Research into heat on the underground. Now at impact stage with embedded change to ways of working.	No links to specific reports shared - information found via search: TfL Info on Adaptation to climate change: https://tfl.gov.uk/corporate/about-tfl/adapting-to-climate-change Paper on Impact of heat on London underground: https://rmets.onlinelibrary.wiley.com/doi/10.1002/wea.4421 Likely to be the study they are speaking about.
4	LNER (train operator) has used AI to manage	No Links Shared, no public material found via

#	Case Study	Further Information
	catering stock on services. It claims to have created a £1m saving (including food waste).	search.
5	Example of data requests to the water sector and research project use: Exeter University has operated a collaborative project to create a model to reduce spills. Data requested includes water recycling data, catchment areas and sewer network data. From solutions section: Depends on the individual data shared - BBC ran analysis on spills occurring during dry weather and wet weather. All based on one point of time in the 24 hour period. No real background info to situate the analysis in. Context would have been more helpful for real understanding.	Identified via researcher search: BBC Article: https://www.bbc.co.uk/news/articles/c4nn46rjej6o Anglian Response: https://www.anglianwater.co.uk/news/de-mystifying-dry-day-spills/
6	Rail Data Marketplace (RDM): RDM aims to deliver better service to customers through the sharing of data across government, Regional Development Offices and combined authorities. It was set up after the disruption caused by the May 2018 timetable changes, which has been partially blamed on lack of access to network data.	Database available via: https://raildata.org.uk/ Data on sharing with researchers specifically is not currently available, however research uses were cited anecdotally.
7	Manchester Metropolitan University project establishing a large-scale city air pollution monitoring network and using the data, in combination with transport data, to examine the relationship between emissions, congestion and air quality.	https://www.mmu.ac.uk/research/projects/air-pollution-monitoring-modelling
8	Future Electric Vehicle Energy Networks supporting Renewables (FEVER) project, University of Sheffield: given as a good example of sharing data and it being used effectively by academics in the field of operations and decision science.	https://www.charginginfrastructuresymposium.com/programme-2024 , 15:00-15.30 slot. https://www.sheffield.ac.uk/omds/logistics-and-supply-chain-management/research-projects/future-electric-vehicle-energy-networks-supporting-renewables
9	Smarter regulation programme - supporting data and digital governance in infrastructure sectors via new approaches and responsiveness.	Smarter Regulation Programme: https://www.gov.uk/government/collections/smarter-regulation
10	Privacy enhancing technology has the potential to offer differential privacy for users and uses.	Example given by participant: https://es.catapult.org.uk/report/digitalising-licensing-in-energy/

Appendix 3: Data analysis notes

The following factors informed survey data analysis:

1. Respondent categorisation was largely informed by a self-classification question answered by participants. To ensure more accurate representation, the organisation type for four respondents was reclassified to better align with the actual characteristics of these organisations (e.g. recategorisation of one respondent from arms-length body to regulator in reflection of official definitions provided by the UK government).
2. While designed to handle a larger sample size, survey responses were limited in practice at least in part due to a short data collection window. Accordingly, survey results have limited statistical power and have been analysed sensitively to account for this limitation.
3. Limitations of the survey software led to a small number of instances whereby respondents interpreted, and therefore answered, questions in a manner not intended during the research design. These instances were not identified through the initial user testing phase and were only identified during analysis following survey closure. Where relevant, these responses were removed from the analysis and/or analytical methods adjusted to accommodate differences in response type.
4. The limited sample size means the results may not provide a definitive hierarchy of barriers across all participants. Analysis refrains from numerically ranking barriers in certain sections of the results.

The following factors informed interview and workshop data analysis:

1. Categorisation of benefits and barriers was informed by findings from the literature review as a starting point. However, thematic coding was not restricted to these categories only – researchers evaluated the data *de novo* to examine where categories remained prescient, where new categories emerged, and where these were discussed as standalone items versus sub-categories within existing ones. Where relevant, the results sections indicate where new themes have emerged which challenge or go beyond data identified through the literature review.
2. Category delineation is not an ‘exact science’ as discussions often touched on more than one topic and/or identify intrinsic relationships (e.g. environmental and societal co-benefits). Where these items have occurred, researchers have categorised them according to their primary presentation in the raw data and/or weight of attention paid by participants.

3. Discussion of benefits and challenges was often interwoven in natural language/speech (interviews) and group discussions (workshops). Accordingly, data informing the benefits section was sourced from across different question areas in data collection activities.

Appendix 4: Mapping of original barrier descriptions to shortened versions suitable for charts and tables

Original Barriers	**Shortened Barrier Names for Chart**
Data being commercially sensitive	Commercial Sensitivity
Data will be used in a way that risks reputational damage to the organisation	Reputational Risks
Concerns about data protection	Data Protection Concerns
Legal concerns (e.g. over intellectual property rights)	Legal Concerns (e.g. IP Rights)
Sharing the data could pose security risks	Security Risks
Data is not good enough quality to share	Poor Data Quality
Data will be used for reasons other than those initially stated (eg partly for commercial purposes)	Unintended Use of Data
The cost (in terms of time money and other resources) of getting the data ready for sharing	High Costs of Preparation
Data will be used in a way that does not align with the interests of the organisation	Misaligned Interests
Data will be used in a way that does not align with the interests of the organisation	Unsuitable Format
Data not being in a suitable format for sharing	No Sharing Policies
No organisational policies or procedures for sharing data for research purposes	No Policies for data sharing (for Research)
Lack of organisational motivation to share data	Lack of Motivation
Lack of clarity about how data will be used	Unclear Usage
Insufficient skills internally to support data sharing with researchers	Insufficient Internal Skills
Unable to see the value in sharing data with researchers	No Perceived Value
We haven't been asked to share data for research purposes	No Requests for Data Sharing
Proliferation of low value data samples detracts from potential higher value of datasets of curated data	Low Value Data Detracts from High Value