Exploring new approaches for sharing data in the built environment

Whitepaper

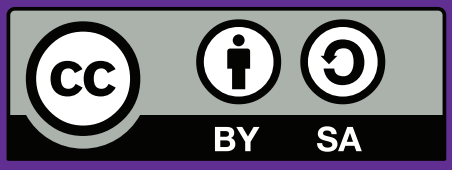
August 2021

Open Data Institute for Arup

ARUP

About

This report has been researched and produced by the Open Data Institute, for Arup. The authors are Leigh Dodds, Ben Snaith and Joe Massey. Jack Hardinges provided additional contributions. We would like to thank Francesco Anselmo, Simon Evans, Neil Rebeugeot and the wider Arup team who have supported our research.



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Contents

Foreword from Arup

At Arup, we decided to change how we work with data. We understood in 2018 that the status quo was no longer delivering on key strategies for sustainability, inclusion and digitisation. A rethink was required to maintain our competitive advantage and our ability to shape a better world.

In 2019 we reached out to clients, our members and partners such as the Open Data Institute (ODI); to listen and prototype new ways of working. We were guided by our first data strategy, the 'Data Target Operating Model'. Progress was made in working with asset management data in civil engineering, the use of natural language processing in our corporate services, the application of machine learning in structural engineering or agent-based models in urban planning and climate change work. Towards the end of this phase, we developed our research programme 'Data Supernova' to help identify and shape critical trends.

Throughout this work it became clear what we needed to do in the second phase of our data strategy 'Creativity with Data at Scale'; we have to change and transform our vision, leadership, our data infrastructure and science, our data-enabled innovation, together with our knowledge, culture and practices.

This was not enough. Arup is not an island. Hence we focused 'Data Supernova' on two key topics: data infrastructure for market openness and data enabled sustainable development goals. In this report we focus on how the built world can reduce the cost and friction between our clients and us, partners and the ever-growing market for refined data.

It is a first step, with the ODI towards an open innovation process that can lead to comprehensive changes in our industries.

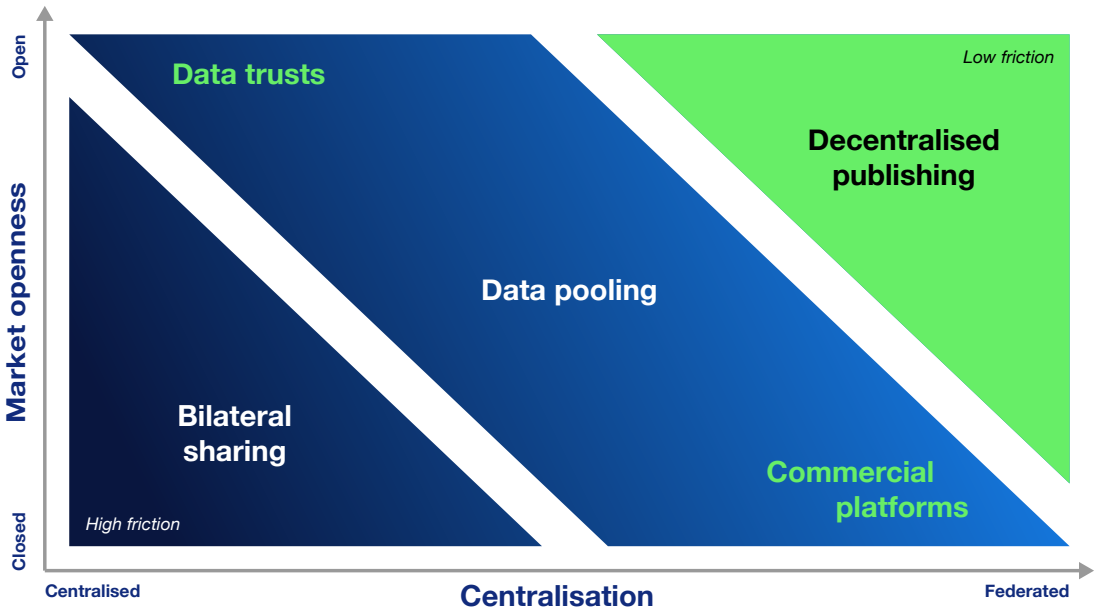
This report provides critical foundational concepts including three approaches for sharing:

- the default model of bilateral data-sharing agreements
- newer approaches for data pooling
- decentralised publishing initiatives

What we envision is that as sectors such as energy and the construction industry move towards market openness, all three of these activities can take place symbiotically and create a data market-place in which:

1. Cost and friction are both reduced, leading to industry wide efficiencies, economic growth and enabling sustainable development goals;
2. Commercial and social platforms for data sharing can prosper; and
3. Data Institutions such as Data trusts or Open Implementation Entities will emerge.

Since this report was published in March 2021, I have developed a visual representation to illustrate 'Data Infrastructure for Market Openness' (see below). I hope readers will find it helpful in further understanding these concepts. It is a model where bilateral sharing, data pooling and decentralised publishing all act in harmony:



It is exciting to see 2021 as the year when these newer approaches begin to deliver value. The decentralised initiative from Icebreaker One is a step change for Open Energy, and there is momentum behind the idea for a data trust for construction or data boards in cities.

I would like to thank our team in Arup and the ODI for another excellent collaboration. The ODI is an organisation we trust as a neutral host for the development of data institutions. They share our interest in designing or catalysing markets to move towards the production of greater social, environmental and economic value from data. As a result, it is an excellent (and agnostic) partner for this project.

We will further explore data infrastructure for market openness and data enabled sustainable development goals at the ODI Summit in November. Please get in touch via data@arup.com or info@theodi.org

Volker Buscher (July 2021)
Arup | Chief Data Officer

An earlier version of this foreword was originally published with the report in March 2021. It was updated in July 2021.

Executive summary

We create value from data when it is used. This means that data needs to be accessible to those who can use it well. It needs to be at the right point on the data spectrum.

Data has the potential to help us tackle a variety of social, environmental and economic problems. Solving the biggest challenges requires sharing data not just within organisations, but also between partners and across sectors. This requires investment in the data infrastructure that can help us to use data well.

Like other industries, the organisations designing, building, operating and maintaining the built environment are looking at how to create value from data. In this paper we explore some of the challenges and opportunities.

There are a range of existing approaches for increasing access to data and open, collaborative models for building and maintaining the necessary data infrastructure. Data-sharing agreements are the default model for sharing data between organisations and for engaging with researchers. They shape how data is used by providing clarity around the rights, responsibilities and permitted forms of reuse. But other models exist.

Pooling data from multiple organisations allows data to be aggregated and analysed in ways that individual organisations cannot achieve by themselves. For example, to support benchmarking and analytics that can increase efficiency or improve safety. With strong governance and investment in centralised infrastructure, organisations can feel comfortable that data is being used within a trusted environment.

But centralised approaches to sharing and governing data are not suitable for all scenarios. Initiatives like open banking, use another approach based around decentralised publishing of data using open standards and a common framework that governs how data is accessed, used and shared. This works well when data is best shared by those who collect and maintain it.

The right approach depends on the purpose of sharing data. We need to be clear on the challenges to be solved before building our shared-data infrastructure.

Smart buildings and digital twins can help us tackle a number of common challenges, including the urgent need to adapt to a changing climate. We highlight how increasing access to data can contribute to tackling those challenges and discuss how building stronger data infrastructure and independent stewardship of data and open standards are important to achieving that goal.

Introduction

Around the world we are facing a range of social, economic and environmental challenges. A safer, sustainable and more resilient future will require us to innovate and adapt the ways in which we engineer and maintain the built environment, including our infrastructure, energy and transport networks.

Some data generated in the built environment sector will be either commercially sensitive or personal. This data should remain private. But there are other datasets that could be made more accessible. Increasing access to data held by both the public and private sector will allow it to be used in ways that increase safety and productivity, drive innovation, tackle climate change and enable sustainable growth. To unlock that value we need to ensure that data is accessed, used and shared in ways that are both trustworthy and sustainable.

In 2019, in collaboration with the Lloyd's Register Foundation, we published 'A manifesto for sharing engineering data'.[1] Our manifesto describes a set of principles to help increase access to data and drive innovation across the engineering and built environment sectors. One year on from the manifesto, it is time to put more of these principles into practice.

This paper explores how the principles relating to the sharing and stewardship of data might be put into practice. We do this by exploring two built environment use cases: smart buildings and digital twins, but there are other examples that could be drawn from across the sector.

To set the scene we explore how other sectors are adapting to use data well and review several approaches for increasing access to data while mitigating potential harmful impacts from its use. These approaches include the creation of data institutions – organisations whose purpose involves stewarding data on behalf of others, often towards public, educational or charitable aims.[2]

1. Open Data Institute (2019), 'A manifesto for sharing engineering data'
2. Open Data Institute (2020), 'Data institutions'

Unlocking the value of data for the built environment

We unlock value from data when it is used. Data needs to be shared and published in ways that build trust and ensure it is used ethically, legally and securely.[3]

3. Royal Academy of Engineering (2018), 'Cyber safety and resilience'
4. Open Data Institute (2019), 'The data access map'
5. Open Data Institute (2020), 'Sharing data to create value in the private sector'

Building a shared vision for using data well requires a common understanding of not just the benefits of sharing data, but also the infrastructure and governance needed to support its use.

A first step is recognising that data exists on a spectrum, from closed, shared to open.

- **Open data** is data that is available for anyone to access, use and share. It is published under an open licence that allows it to be used for any purpose.
- **Shared data** is data that is only available to certain people or groups, such as researchers. Data that is shared will typically be made available for specific purposes that are defined by, for example, a data-sharing agreement. There are many ways in which data can be made accessible, or 'shared', with others.[4]
- **Closed data** is data that is held privately within an organisation, such as employment contracts, policies or sales reports.
- Diagram 1 shows how data from the energy sector would be applied to the Data Spectrum.

By making data as open as possible – while protecting privacy, commercial confidentiality and national security – we can unlock more value from it.

Our research into the benefits of sharing data in the private sector[5] found that sharing data – between suppliers, partners and with customers – creates new opportunities, increases revenue, reduces direct costs and improves efficiency in operations. In addition to direct benefits for individual organisations, data can also be used to solve a range of economic and social challenges, for example helping businesses adapt to a changing climate or responding to the coronavirus (Covid-19) pandemic.

A manifesto for sharing engineering data

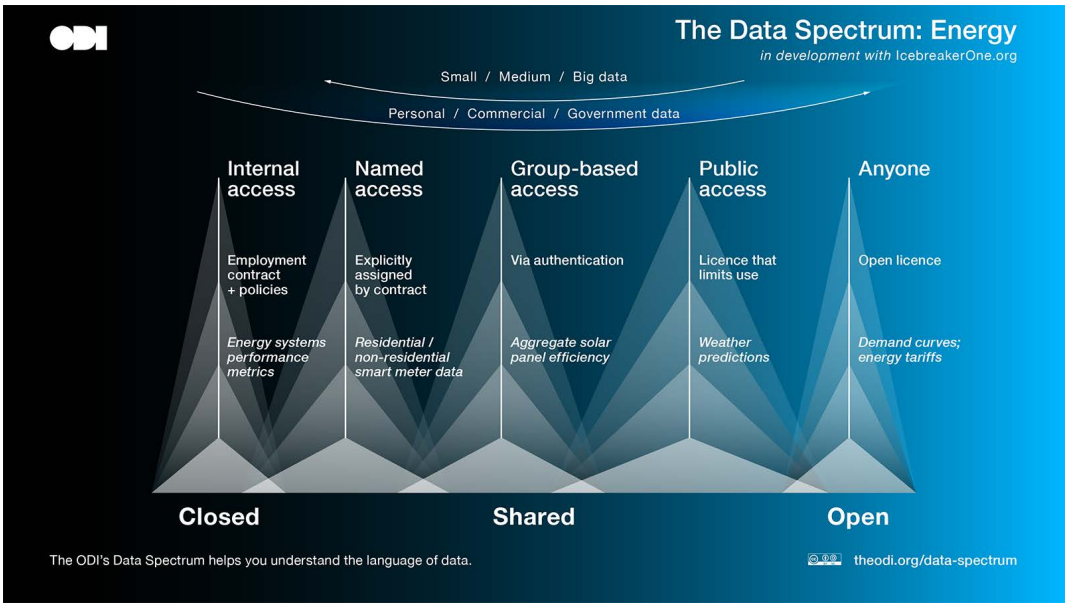
A set of principles for increasing access to data and driving innovation in the engineering and built environment sectors

1. Data is infrastructure.
2. Data must be stewarded.
3. Opening and sharing data unlocks value.
4. Explore new data-sharing models.
5. Use challenges to drive innovation that solves problems.
6. Regulation must adapt to new technologies and uses of data. 7. Building data literacy and skills.
8. Ensure data is used legally and ethically.
9. Share knowledge and insight.

Developed by the ODI and the Lloyd's Register Foundation in partnership with a range of engineering and safety organisations across the United Kingdom (UK), these principles have been endorsed by 18 organisations and programmes.

Read more about the manifesto and how to endorse it.

Diagram 1 The ODI Data Spectrum: Energy[6]



6. Icebreaker One (2020), 'Data Spectrum for Energy'
7. Open Data Institute (2020), 'Sharing data to create value in the private sector'
8. National Infrastructure Commission (2017), 'Data for the public good'
9. National Infrastructure Commission (2017), 'Data for the public good'

Better use of data from the built environment can help to:[7]

- **increase productivity** in construction and engineering through better collaboration across the supply chain and asset lifecycle
- **enable open innovation**, such as, by including more communities and organisations in the development of services, helping to create new technologies and insights or developing new approaches to designing, building and maintaining aspects of the built environment
- **driving economic growth** by enabling new business models and cost savings across the engineering and construction sector
- **reduce impact on the environment** by reducing waste in construction or engineering processes or through better management of energy and water resources.

Data about the built environment must be at the right point on the Data Spectrum, but it must also be supported by the data infrastructure that enables and governs its use.

The 'Data for the public good' report[8] highlights just that: data is now as much a critical component of national infrastructure as steel, bricks and mortar. Our data infrastructure consists of shared data assets, the standards and technologies that enable its use, the policies and guidance that inform how data is used and managed and the organisations that govern it[9].

Data infrastructure needs to be maintained in the same way that our physical infrastructure also needs to be maintained. To ensure sustainable access to data and to help unlock value over the long term, our regional, national and international data infrastructure will need to be planned, invested in and strengthened over time.

Insights from other sectors

Data and digital transformations are driving changes across all sectors, not just the built environment. There are many examples of how adopting open, collaborative approaches are helping to tackle social, environmental and economic challenges through sharing data.

So how are other sectors working together to understand how to share data and build the necessary data infrastructure and institutions to govern it?

For insights, we briefly look at how both the banking and agriculture sectors have been responding to the opportunities presented by increasing access to data. These examples highlight the importance of finding the right combination of sharing models, governance and collaborative approaches that will help to tackle challenges in the built environment.

Open banking – using data and standards to drive innovation

Following its investigation into competition in UK banking, the Competition and Markets Authority (CMA) mandated the UK's largest banks to launch the Open Banking initiative, whose aim is to 'stimulate innovation across the financial sector, to enable the "unbundling" of complex retail banking products (in particular personal and business current accounts) and to ultimately lead to greater competition overall.'[10]

The Open Banking Working Group was set up in 2015 for two primary purposes. Firstly to explore how data could be used to help people to transact, save, borrow, lend and invest their money, and thus improve their banking experience. And secondly, to begin development of the necessary open standards that would enable access to data.

The CMA determined there would need to be a central implementation[11] entity to develop the standards, build the supporting infrastructure and coordinate and drive implementation.[11] So the Open Banking Implementation Entity (OBIE) was created to oversee the implementation of the regulation, work with the banks to design the technical standard and to support adoption. To be successful, there needed to be an independent, trusted central entity to steward the standards, coordinate across the sector and ultimately, drive change. The creation of the Open Banking Standard guides how banking data should be created, shared and used, facilitating the move to openness, alongside specific challenges to increase innovation.

Spurred on by regulatory changes mentioned above, the banking sector is now beginning to benefit from the adoption of a new standard for open banking. The ability for a broader range of organisations to access data on banking products and customer transaction histories is leading to the creation of more innovative products and services. By granting trusted third parties access to data, customers can more easily switch current accounts, gain access to personal finance dashboards, as well as make more accurate loan and credit-referencing decisions.

10. State of Open Data (2019), 'Data infrastructure'
11. ibid

A recent review[12] of the impact of open banking found that, while the initiative is still early in its development, it has evolved to add value to customers and service providers on broader financial management issues, as well as delivering on its intended purpose: to stimulate innovation. The model initially adopted by the UK is now being replicated worldwide. The approach illustrates the role of regulation in shaping data-access arrangements within a sector, the speed of data innovation once the necessary data infrastructure is available, the ability for innovative models to spread to other countries and the importance of independent stewardship of both data and standards.

Agriculture – using data to enable collaboration

By 2030, the agricultural productivity and incomes of small-scale food producers need to double to achieve food security and promote sustainable agriculture[13]. The rapidly increasing availability of data has huge potential to improve the agriculture sector and address food and nutrition challenges. Supported by government funding, a range of organisations across the agriculture, nutrition and digital sectors are collaborating to increase access to data to address food security and nutrition challenges. Global Open Data for Agriculture and Nutrition (GODAN) Action[14] was a three-year project that began in 2017, to enable data producers, stewards and users to engage effectively with open data and maximise its potential to bring a positive impact in the agriculture and nutrition sectors. The project focused on three key areas of work: standards, impact and capacity building.

Each partner organisation brings unique skills and experience to contribute to the project, which has engaged many different groups to improve understanding and use of open data – training over 1,000 data producers, stewards and users around the world. The project has created tools to facilitate discovery and use of standards for agricultural and geospatial data across diverse communities of practice. It has also made recommendations to improve the discovery, interoperability and use of data across the sector. The GODAN Action project highlights how a range of organisations, working in collaboration, can drive change across an industry and can help to drive change, operating in a similar way to the ODI Peer Networks.[15]

GODAN Action has led to the creation of new data infrastructure, including shared-data assets, standards and new institutions to manage and provide access to that data. One example being the Africa Regional Data Cube, which works to harness 'the latest Earth observation data and satellite technology to help Ghana, Kenya, Sierra Leone, Senegal and Tanzania address various issues relating to agriculture, food security, deforestation, urbanization, water access and more.'[16] Coordinated action between a group of actors has resulted in an approach that is demand led, increases capacity building and is sustainable, which will help to deliver better decision making and progress towards achieving the UN 2030 Sustainable Development Goals.[17] The GODAN Action programme benefited from having a strong emphasis on collaboration to develop standards and create shared-data assets for the sector.

12. Open Banking (2019), 'Open Banking, preparing for liftoff'
13. United Nations (2019), 'Food'
14. GODAN Action, 'About GODAN Action'
15. Open Data Institute (2016), 'Peer networks for open data leaders'
16. Global Partnership for Sustainable Development Data, 'Africa Regional Data Cube'
17. GODAN Action, 'Africa Regional Data Cube'

Three approaches to data sharing and governance

There are numerous approaches to sharing data.[18] Each model has its own strengths and weaknesses.

Choosing the right model depends on a wide range of factors including:

- the length of time over which data will be shared, which might dictate the level and type of technical infrastructure required
- the volume and type of data to be shared, which might require specific governance or technical protection
- the maturity and scale of the data ecosystem.

Our report on sharing engineering data noted 'The engineering sector needs to explore and use the full range of data-access models, creating new institutions where needed, to help build a data ecosystem where important data is accessible and data is used and shared in trustworthy ways.'[19] The same is true for the built environment sector.

In this section we outline some the strengths and weaknesses of three approaches for increasing access to data: data-sharing agreements, decentralised publishing initiatives and data pooling. The last two involve the creation of data institutions – organisations whose purpose involves stewarding data on behalf of others, often towards public, educational or charitable aims.[20] Data institutions come in different shapes and sizes and, like the OBIE and the African Regional Data Cube, they can fulfil different roles in their data ecosystem. But they are becoming an important part of how we move data along the data spectrum while preserving trust.

1. The default model: data-sharing agreements

The typical model for sharing data between two organisations are data-sharing agreements. On the ODI Data Spectrum, data-sharing agreements are defined as 'named access'. These are legal contracts that define a close, direct relationship between organisations. They define how data will be shared, the purposes for which it will be used and the roles and responsibilities of the individual organisations.

Data-sharing agreements are used in a variety of circumstances, such as to share data with suppliers and partners. They are frequently used to enable sharing of data with researchers and startups as part of open-innovation models.

The Data Pitch challenge programme used data-sharing agreements to facilitate sharing of data between a variety of organisations to tackle a variety of challenges[21] such as futureproofing retail supply chains, smart manufacturing and innovative approaches to reduce traffic congestion. The programme successfully demonstrates how simple data-sharing models can quickly unlock more value from data, providing direct benefits to individual organisations.

18. Open Data Institute, 'Mapping the wide world of data sharing'
19. Lloyd's Register Foundation (2019), 'Insight report on sharing engineering data: using data for the public good'
20. Open Data Institute (2020), 'Data institutions'
21. Data Pitch, 'Competition/ Challenges'

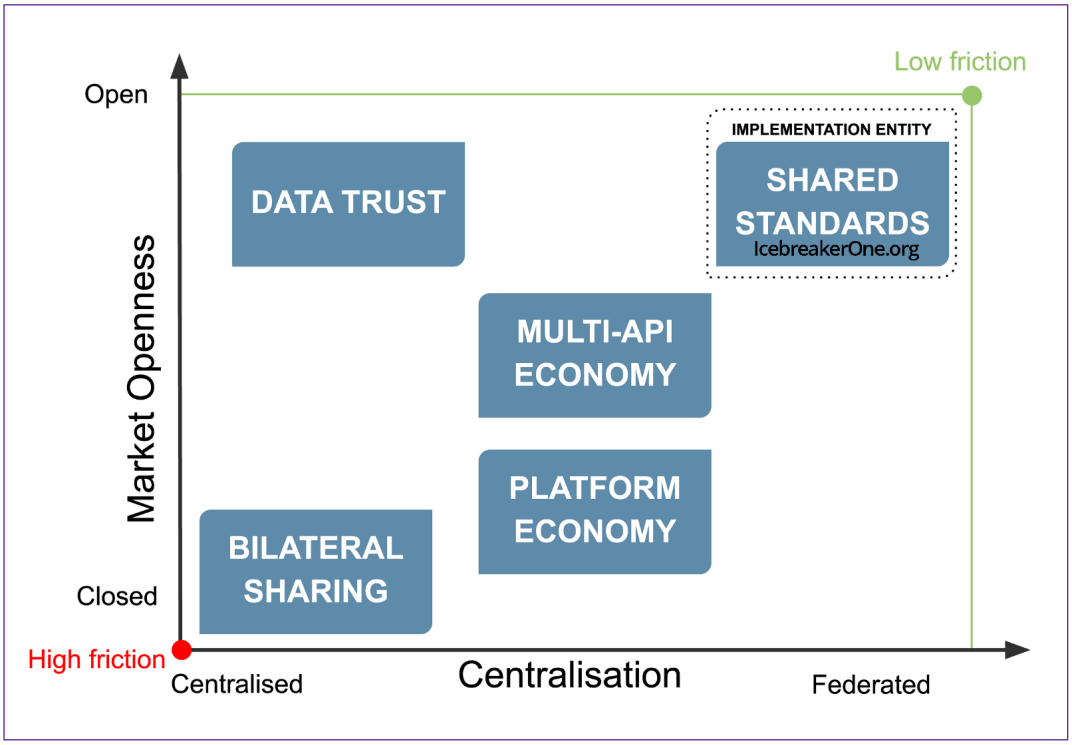


Image credit: CC BY 4.0, by Icebreaker One

Icebreaker One is a non-partisan, global non-profit organisation with an interest in sustainable data sharing and innovation, which aims to help decision makers work towards net zero through the use of data. This diagram shows an exploration of approaches to data sharing, created by Icebreaker One to 'stimulate discussion'.

Strengths of data-sharing agreements include that they:	Weaknesses of data-sharing agreements include that:
• build direct relationships between partners and suppliers • provide specific, clear controls over use of data, including defining purpose and time limits • are suitable for sharing data with a wide variety of different organisations and researchers.	• they can be difficult to negotiate and agree access • creation of custom legal agreements is costly and time consuming • they are primarily designed for 1:1 sharing of data • they typically draw on bespoke or non-standard models for sharing data, increasing integration costs for data consumers.

22. Open Data Institute (2020), 'Comparing decentralised data publishing initiatives'

23. Icebreaker One, 'Introducing Open Energy'

2. Coordinating access to data via decentralised publishing initiatives

Data consumers often need access to the same data, from multiple sources. Negotiating data-sharing agreements can be time-consuming and costly. Lack of coordination around how data is shared, leads to bespoke technical approaches that further increase costs and add friction.

An alternative approach are 'decentralised publishing initiatives'. These are industry-wide collaborations that involve multiple organisations publishing or sharing data directly from their own infrastructure, using open standards that define how data is being accessed, used and shared within a standard legal framework or via an open licence.

We have documented 14 examples of these types of initiative, which include programmes like Open Banking, OpenActive and OpenContracting.[22] New initiatives like Icebreaker One's Open Energy project[23] are also adopting this model.

This decentralised approach addresses some of the issues with simple data-sharing agreements by defining a standard model for accessing, using and sharing specific types of data across a sector or industry.

The work of developing the open standard, common legal frameworks and shared tools and technology can be undertaken by a group of parties or a single data institution, to coordinate the cross-sector collaboration.

Strengths of decentralised publishing initiatives include that:	Weaknesses of decentralised publishing initiatives include that:
• organisations retain stewardship of data • collaborative approaches build relationships across a sector • the potential for lock-ins is reduced due to the use of open standards and common legal frameworks • they are applicable to a wide variety of use cases across sectors, geographies and data types.	• additional effort and investment is required to drive creation and adoption of open standards • data consumers may still need to find and integrate a large number of data providers • data stewards are required to create their own infrastructure to support sharing and publishing of data.

3. Data pooling

Decentralised publishing of data can help to reshape how data is shared across sectors through the use of open standards, but sometimes it is necessary to aggregate data to unlock its value. Pooling of data from multiple sources can help to create a critical mass of data that can be analysed to create insights that are impossible for individual organisations to do themselves.

Benchmarking – the sharing of data and metrics across an organisation or sector to compare performance – is a common example. By sharing data with a central body it becomes possible to compare data on operations, performance, finances and a range of other indicators, including salaries, diversity and safety practices, allowing companies to benchmark themselves and so make better decisions about how they operate.

These central bodies are data institutions that act as a trusted intermediary, providing the necessary oversight and governance around the shared data, enabling those contributing to the pool of data to trust that it will only be accessed, used and shared in agreed ways.

HiLo, a data institution in the maritime sector, supports sharing of safety and accident data across the sector. By analysing the aggregated data it is able to provide individual insights to the companies contributing data. HiLo has helped to save lives and money: to date it has reduced lifeboat accidents by 72%, engine-room fires by 65% and bunker spills by 25%.[24]

HiLo has helped to save lives and money: to date it has reduced lifeboat accidents by 72%, engine-room fires by 65% and bunker spills by 25%

The UK Geospatial Commission's National Underground Asset Register[25] is another example of data pooling for the benefit of shared infrastructure. The project is creating a single data platform for construction projects, containing information on buried utility assets such as cables, pipes, sewers and ducts for when preparing ground investigations and excavation work.[26]

Strengths of data pooling include that:	Weaknesses of data pooling include that:
• they enable sector-wide alignment around key targets and assessment of whether they are being met • they enable and support analysis and use of data across multiple organisations • contributing organisations can remain involved in ongoing governance and stewardship • they provide a common technical platform for analysis, use and onward sharing of data.	• there are additional costs of setting up and sustaining a central platform and technical infrastructure to aggregate data • risks increase, such as from data breaches, when key datasets are held by a single organisation or infrastructure • benchmarking requires standardised data collection so that true comparisons can be drawn.

24. Open Data Institute (2020), 'Case study: The value of sharing data for benchmarking and insights'
25. GOV.UK (2020), 'National Underground Asset Register project update'
26. ibid

Increasing access to data from smart buildings

Buildings of all types, including offices, shops, schools and entertainment venues are increasingly being fitted with sensors and connectivity that are providing access to new sources of data. This includes data from an increasingly diverse range of sources, not just energy or water usage, but observations from wifi access points, bluetooth beacons, cameras, lighting systems, air quality sensors and door locks.

Who might benefit from increased access to data and how?

Better access to and use of data from smart buildings would have significant benefit for **property managers and building occupants**.

For example, 'data collected by motion and occupancy sensors at a building level [can be used] to regulate air-conditioning and lighting in real time, thereby reducing energy costs and optimizing the internal environment for its intended purpose.'[27] At the White Collar Factory in Shoreditch, London, metering is 'implemented at a very detailed level and is expected to provide valuable information on how the building performs in practice.'[28]

The data can also lead to more efficient building maintenance, as 'the continuous monitoring and predictive capability of Internet of Things (IoT)-enabled buildings can also preempt a repair or maintenance issue by enabling a building manager to take appropriate corrective action before tenants even notice a problem.'[29]

Arup has also been researching the benefit of smart buildings during the coronavirus pandemic, theorising that 'on-site assessments, computational modelling and simulations can quickly identify high-risk areas, such as poor ventilation. This allows operators to recognise the sources of pollution and optimises the building's operation and management.'[30]

Increasing access to data can also support other stakeholders:

- **Tenants** might be offered more control over lighting, heating and ventilation, as well as better services, for example, offering more flexible spaces and easier interactions with spaces such as booking meeting rooms, personal room temperatures and customisable food and beverage options.
- **CEOs** and **COOs** may be able to better understand their building use and connected environmental impacts, allowing them to reduce their carbon footprint and save on capital costs.
- **Researchers** could use data to evaluate and compare building performance. They could also use this to monitor any effect that the proliferation of smart buildings has on meeting energy reduction targets and hold organisations in poor-performing buildings to account.
- With access to standardised data and interfaces, **startups** might offer new tools, products and services utilising the data generated by smart buildings.

27. Deloitte (2016), 'Smart buildings: How IoT technology aims to add value for real estate companies'
28. BRE, 'White Collar Factory: an inspiring BREEAM Outstanding building'
29. Deloitte (2016), 'Smart buildings: How IoT technology aims to add value for real estate companies'
30. Arup (2020), 'Smart buildings: how a virus might lead to healthier buildings'

- **Society** can benefit from buildings contributing less to the climate crisis. Stronger data infrastructure for smart buildings can help to fight the climate emergency and support other sector's initiatives such as the EU's Energy Performance of Buildings Directive.

On a macro level, data can support the built environment sector to ask and answer bigger questions. Are office blocks an efficient use of capital resources for an organisation? Has the coronavirus pandemic changed our reliance on office space? If we have offices, where should they be: city centre, business parks or suburbs? What impact do they have on the environment and on the community around the sites?

Where are the challenges in sharing and using this data?

Currently, in the UK, data from smart buildings is siloed. This begins during design and construction, when digital plans are handed between multiple organisations over the early lifecycle of the building.

These silos continue once the building is occupied and functional – leading to fragmentation across data platforms, services and infrastructure across the sector. As Arup commented: 'Too often buildings' IT packages are more of an after-thought, something to fit in, not an opportunity to be seized.'[31] For the promise of smart buildings to be realised, these vertical silos must be broken down so that data can be accessed and shared more easily.

The increasing amount of data that is used to monitor buildings and workplaces also raises potential ethical and privacy concerns. Data from smart offices could be designed or repurposed as a form of workplace surveillance,[32] leading to harmful impacts for workers. Increased use of data in offices could 'potentially [result] in the disclosure of data that people might not feel comfortable disclosing (such as where they go, what they do, when and with whom they spend time, whether they are healthy and more)'.[33]

Understanding and addressing such ethical and privacy issues within smart buildings are vital. It would be hasty to roll out experimental technologies, such as the IoT and sensors, without understanding the potential negative aspects on people and the environment.

While there are potential benefits from using data to reduce carbon footprints, there is a need for a balanced approach to data collection and use. Not only do the environmental impact of the additional sensors need to be factored in, but also the ongoing financial and environmental costs of storing the data in servers, which worldwide, has a significant impact on the environment.[34]

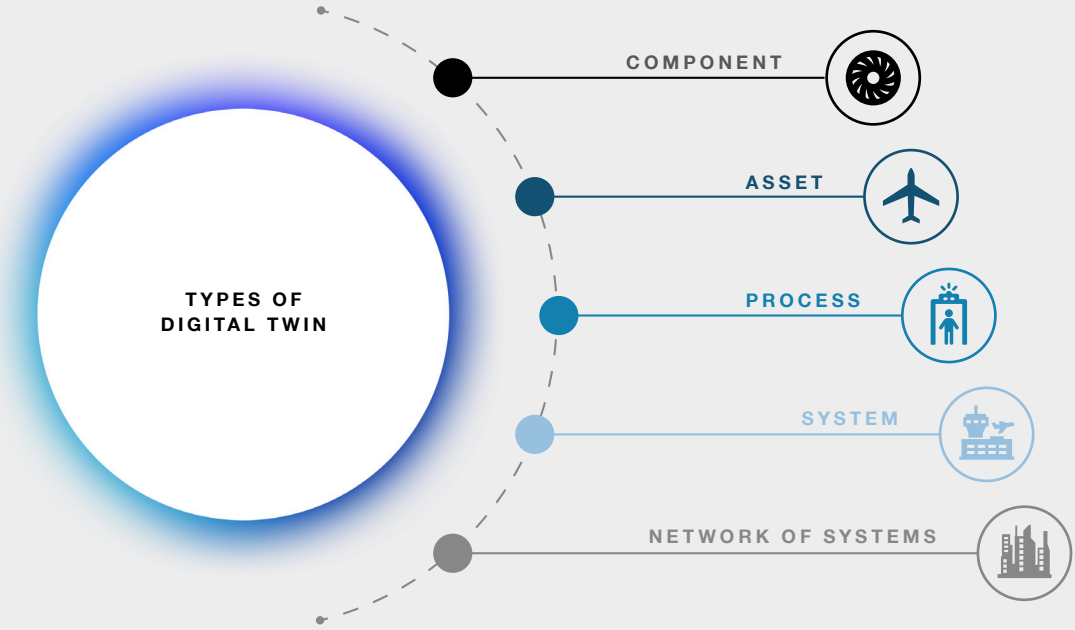
Where could new data-access approaches help?

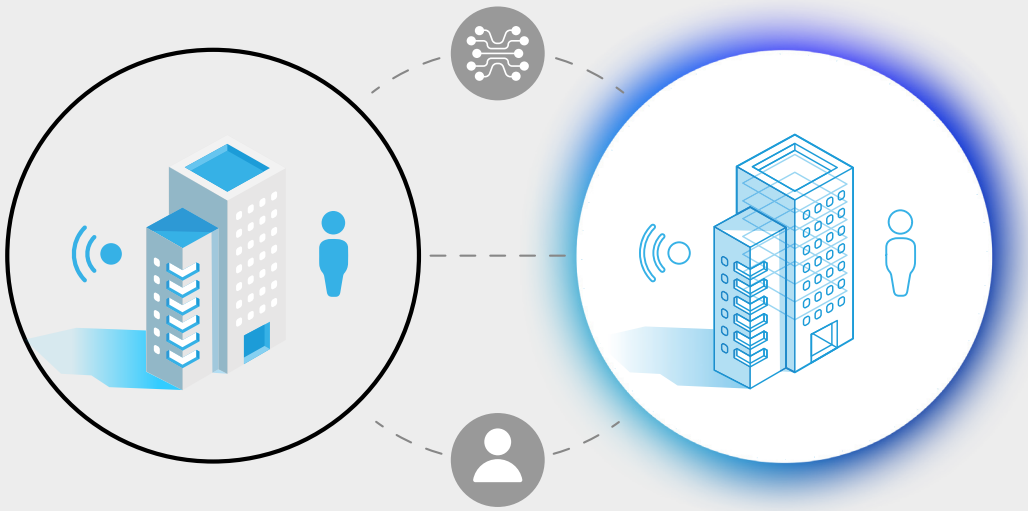
Addressing the above challenges requires the creation of a stronger, more-open data infrastructure for smart buildings. This infrastructure will be based on open standards and clear policies and guidance that will clarify the rights and responsibilities around how data is being collected, used and shared.

A new data-access approach could bring benefits, for example to benchmark data for the sector. A new set of initiatives could help to coordinate the creation and adoption of open standards – to help to standardise how data is being accessed, used and shared between different systems, services and buildings throughout a building's lifecycle. These standards should be developed in the open, mirroring the standards development in Open Banking and Open Energy and learning from collaborative initiatives to manage infrastructure such as Project 13. BDNS (Building Device Naming Standard), for the naming of data producing devices such as sensors or lights, is the first of these initiatives towards creating 'Open Property' in a similar way. Once the data is standardised, it can be pooled from various sources, allowing comparison of the performances of buildings across a number of metrics, including occupancy, environmental efficiency, finances and safety. This organisation could perform a similar role that HiLo does in the maritime sector, leading to more efficient, safer and healthier buildings. Whether a data institution is the right fit for this problem will require further analysis, but the cases of Open Banking, HiLo and Project 13 should provide an idea of how it could operate.

A data institution could also act as an independent steward of data to assist in addressing the ethical and privacy concerns around data from smart buildings. Data institutions offer potential to provide strong, participatory governance around data, involving owners, operators, tenants and workers in decision making.

31. Arup, 'Bricks, data and mortar: it's time to build in the intelligence'
32. BBC (2020), 'What are the rules on workplace surveillance?'
33. IEEE (2017), 'Towards Privacy-Aware Smart Buildings: Capturing, Communicating, and Enforcing Privacy Policies and Preferences – IEEE Conference Publication'
34. Computer World (2019), 'What impact are data centres having on climate change?'





PHYSICAL WORLD
A real-world asset and its behaviour.

DIGITAL TWIN
A digital twin is the combination of a computational model and a real-world system.

USER
The user interacting with the digital twin.

INTELLIGENCE
AI enabling the digital twin to perform tasks with minimal or no human oversight, e.g. visual perception, speech recognition, natural language translation, decision making.

DIGITAL THREAD
Refers to information channels connecting the physical and the digital asset.

Image credit: © Arup

Sharing data to enable the development of digital twins

Pairing physical roads, bridges and buildings with a 'digital twin' – 'a realistic digital representation of assets, processes or systems in the built or natural environment' – could be the key to more efficient decision making, design and maintenance and increased safety. It is important to emphasise that 'digital twins are a methodology, not a technology. It is an approach and way of working, enabled by connecting the physical and digital worlds'.[35]

The Centre for Digital Built Britain (CDBB) and the Digital Framework Task Group developed the Gemini Principles,[36] to articulate a set of proposed principles to guide the creation of stronger data infrastructure for the built environment. The principles are organised under three overarching themes: purpose, trust and function and are intended to create alignment in the industry, stress the benefits of sharing data for the public good and provide a context within which key questions can be identified and addressed.

Increased data sharing in the built environment sector has the potential to positively impact multiple parts of society. For the economy, there are potential improvements to national infrastructure productivity, security and measurements of outcomes, alongside increases in collaboration, innovation and safety. Greater data sharing in itself could release an additional £7bn per year of benefits across the UK infrastructure sectors.[37]

Who might benefit from increased access to data and how?

According to the CDBB, the National Digital Twin programme has the potential to benefit society in a far greater way than any other investment in the built environment sector,[38] including everything from better health and wellness in office environments to improved air quality in our dense urban environments.[39] Digital twins will improve transparency, create better outcomes for the customer (often the public) and deliver higher satisfaction to the user.

- **Local and national governments** are better able to understand how different events will impact their infrastructure. A digital twins approach will create more useful insights, enabling better decision making. For example, the 'Breathing City' project is improving the wellbeing and safety of urban populations and provides valuable insights to educate society on the impact of pollution through the collection of environmental data to inform a digital twin approach.[40] Pollution, traffic and footfall data are built upon geospatial data to create a digital model of Leeds.

35. Open Data Institute (2019), 'Towards a web of Digital Twins'
36. ibid
37. Deloitte (2017), 'New Technologies Case Study: Data Sharing in Infrastructure'
38. Centre for Digital Built Britain, 'National Digital Twin Programme'
39. Arup (2019), 'Digital twin: toward a meaningful framework'
40. Open Data Institute (2020), 'Case study: Creating a digital version of a city'

41. Water Industry Journal (2019), Managing incidents with the 'digital twin'
42. BBC (2017), 'Autos – The jet engines with 'digital twins''
43. Open Data Institute (2020), 'Sharing data to create value in the private sector'
44. Open Data Institute (2020), 'Designing trustworthy data institutions'

- A collaborative project between Newcastle University and Northumbrian Water Group is an example of how **infrastructure owners** can better respond to incidents in the water network, such as burst pipes or heavy rainfall thanks to a digital twins approach. The 'Twincident'[41] model uses geospatial data, such as land-cover maps and light detection and ranging (LIDAR) maps to create a model of Newcastle and surrounding areas. This is integrated with water company infrastructure data and complemented with dynamic weather data and environmental data such as traffic and air quality data to build a full picture of the city.
- There are considerable benefits to **businesses**, through the creation of a new market, improved efficiency to the entire value chain, reduced uncertainty and better risk management. For example, **engineers** at General Electric & Rolls Royce use sensor data in their engines, to measure everything from the temperature of the exhausts, to the speed of the turbines to improve performance.[42] Similar methods are used by NASA when building their space shuttles and in Formula One cars.
- In a broader sense, a National Digital Twin will create less disruption and waste, more reuse and better efficiency, all of which create **environmental** benefits through a more circular economy in the built environment.

Where are the challenges in sharing and using this data?

To build an effective digital twin approach, a range of data is required and that data must be shared openly across multiple sectors in a machine-readable format. Agreeing on standards, a common language for different stakeholders across different domains, takes time and coordination and will impact the ability to connect twins, which may result in issues for owners of single assets, such as vendor lock-in and lack of evolution.

Retrofitting ageing complex infrastructure with the technology necessary for a digital twins approach may be more difficult than creating new physical, digital-ready assets from scratch, one example of which is wifi black spots in old buildings. Physical infrastructure changes at a much slower pace than the worlds of software and data. Integrating the two in an ecosystem of digital twins may be very hard, slow and costly as a result.

Buy-in to the digital twins approach will be important. Some organisations may implement overly complex, unnecessary systems and lose faith in digital twins or a critical mass could choose to stick with current technologies rather than invest in a digital twins approach. Asset owners need to find the right level of complexity to aim for when developing twins; not all twins need to involve real-time data at large volumes and complex predictive models.

We can also look at digital twins not in isolation, but as data and models that can be shared, connected and integrated, creating a whole greater than the sum of its parts, which has its own set of challenges. Data sharing between organisations is often done on an ad hoc basis, through data requests, hack events and even email. This level of data sharing is not enough to organically support a set of connected twins.

ODI research[43] has highlighted a number of barriers to increasing access to data in the private sector, which touch on a range of cultural, commercial, legal and technical barriers. There are high barriers to entry in choosing a digital twins approach, due to the scale of digital twins projects. Data infrastructure is created by many partners who need to collaborate on sharing data, requiring a level of trust. ODI research[44] found that trust is a crucial lever for increasing data sharing. To share data openly, organisations will need to trust the institutions they are sharing data with and data users need to know whether to use data from it.

Where could new data-access approaches help?

For many of the challenges detailed in this paper, in particular in regard to connecting digital twins, a data institution may be able to help by stewarding accessible, readable, trusted data from across different sectors. There has already been movement towards these goals; the CDBB are working on standards and an information-management network to support access to data for the National Digital Twin programme.[45]

One role that a data institution could take would be in maintaining a set of shared data assets that act as reference data libraries containing sector-by-sector reference data. These data libraries would be stewarded on a sector-by-sector basis, as each sector will have a bespoke set of reference data. The UK is a sector-siloed nation and data sharing will be, at least initially, more coherent within sectors than across them. There are examples of these organisations that already exist and are actively stewarding data for different sectors, such as Xoserve,[46] the Central Data Service Provider for Great Britain's gas market and Smart DCC, which has 'built and maintained the secure national infrastructure that underpins the roll out of smart meters across Great Britain'.[47]

Crucially, these libraries need to be aligned around common standards to ensure they are interoperable and reusable across a larger connected ecosystem of digital twins across sectors. Sharing data across sectors is a prerequisite for the success of digital twins. An organisation in each sector, charged with managing the repository of shared data and associated metadata, would provide a simple way of sharing and communicating between sectors. Some organisations are already moving forward in this space, the Data and Analytics Facility for National Infrastructure (DAFNI)[48] is building a platform that will support the development of essential UK infrastructure by providing a central point for data, modelling and visualisation with data from across sectors.

45. CDBB (2020), 'The pathway towards an Information Management Framework – A 'Commons' for Digital Built Britain'
46. Xoserve, 'Xoserve: Central Data Service Provider for Britain's gas market'
47. Smart DCC, 'About DCC'
48. DAFNI, 'Data & Analytics Facility for National Infrastructure'

Final thoughts

Data needs to be at the right point on the data spectrum. Data has the potential to help us tackle a variety of social, environmental and economic problems. Solving the biggest challenges requires sharing data not just within organisations, but between partners and across sectors. This requires investment in the data infrastructure that can help us to use data well.

It is not enough to focus on the purely technical aspects of data infrastructure, we must also develop the policies and guidance that govern how data is being accessed, used and shared. This will ensure equitable access to data and minimise harms from its use.

In this paper we have outlined three approaches for sharing data that are already being successfully applied in other sectors. We have also briefly explored how these approaches might apply to support the operations of smart buildings and delivering on the promise of digital twins. Choosing the right approaches for sharing data and deciding on what data infrastructure needs to be built depends on the problems and challenges to be solved.

Sometimes, what is needed is a cross-sector initiative to develop open standards and support their adoption and use. In other cases, to pool data from multiple organisations or to manage shared data assets, what might be needed are new data institutions that can provide independent stewardship around that shared data.

More work is needed to understand the right mix of approaches for smart buildings, digital twins and other areas of the built environment. What is common is the importance of openness and collaboration.

Our manifesto for sharing engineering data provided a framework for thinking about how to drive open, data-enabled innovation in the built environment sector. But the next steps require more practical work. There is a need for pilots, prototypes and open innovation.

Through our data institutions programme, we look forward to working with our members, our partners like Arup and the Lloyd's Register Foundation to explore how best to increase access to data in the built environment.

If you would like to endorse the manifesto, share your insights from applying these principles or explore how to scope or pilot data institutions then get in touch at info@theodi.org.

49. Open Data Institute (2020), 'Data institutions'

Get in touch
theodi.org | @ODIHQ