

Towards trusted data sharing: guidance and case studies

Data sharing: implications for policy and practice

The following text is taken from *Towards trusted data sharing: guidance and case studies*, published by Royal Academy of Engineering. The full publication can be read online at:

reports.raeng.org.uk/datasharing

Contents

Overview	page 3
Introduction	page 5
Practical challenges	page 7
Policy context	page 10
Implications for policy	page 11
Acknowledgements	page 13
References	page 14

Overview

Creating value from data requires organisations and individuals to have access to data and use it effectively and appropriately. This project sets out to investigate one aspect of this: trusted data sharing. Scaling up data sharing activity, where it meets business and privacy needs, will help to release value but it requires sources of friction to be reduced and it must be done in ways that maintain, and do not erode, trust.

A series of ten case studies illustrates emerging examples of data sharing and some of the key enablers and constraints, including governance, business models, technologies and regulation. A practical checklist for organisations has been created, drawing on lessons from the case studies. Further work is needed to raise awareness of the opportunities, to progress approaches to data sharing and guidance, and to share best practice.

Through this project and other initiatives, the Academy aims to play a role in supporting better use of data in engineering sectors and in the broader economy. Best engineering practice is a vital part of realising the opportunities with its focus on the interface between technical systems, people and organisations. Data must be assembled, structured and managed over its lifecycle so that it meets business or other requirements, for which a robust engineering approach is needed.

Data sharing may be just a small, but potentially significant, part of the requirements and its success relies on the engineering being done well. Recognising the interdisciplinary nature of the challenge, the Academy has drawn on the expertise of a wide range of stakeholders, through the working group, individuals interviewed as part of the case study research and reviewers.

This publication is for organisations that have identified the opportunity to create value through sharing data and wish to collaborate with others to develop solutions. Policymakers in government and other stakeholders who play a role in promoting the many opportunities and developing enablers for data sharing may also draw valuable lessons from this publication.

Data sharing opportunities

The opportunities for organisations to use data to improve products and processes and to innovate are widely recognised, with ensuing benefits for the economy and society. Static, regularly updated or real-time data may be collected, stored and processed by a single organisation, to improve their own business processes or create new products and services.¹

The opportunities to create value increase when data is shared or exchanged across organisational boundaries. For example, one organisation may share weather data with another organisation that needs accurate weather forecasts for its business. Alternatively, an organisation might have valuable data while another has the expertise to create products or services from it, as is the case with companies who use authorised NHS patient data to develop artificial intelligence-based diagnosis tools for use by NHS clinicians. Additional value might be realised if data can be shared securely across sectoral and international boundaries.

The ability to reliably link datasets also creates new value. Organisations may share data that can be linked using a common identifier. For example, child-level data in England is linked across different government departments in order to increase understanding of how decisions in the family court impact on children's educational outcomes.² Value may originate from many organisations pooling similar data, since analysis on the larger, pooled dataset has greater value than that on a single dataset. For example, oil and gas operators share well component failure data to improve decision-making around operations and maintenance.³

In the case of physical infrastructure, data about different assets and their geographical location, along with operational data, may exist as separate datasets. The ability to share and link them would provide the real-time status of an infrastructure system; the power flow in the electricity transmission grid, for example. In cities, passengers are provided with journey options via apps on the basis of linked real-time data from multiple sources.

Enabling trusted data sharing

Data may be commercially sensitive, or it may relate to individuals, with associated privacy requirements as legislated in the General Data Protection Regulation. Concerns about the sensitive nature of data can restrict data sharing. When one organisation's data is accessed and used by others, appropriate frameworks are vital in order to ensure that data sharing meets commercial, regulatory, legal or ethical requirements and promotes trust. Trust might be enabled through the following activities:

- Ensuring that costs and benefits of collecting, storing and using data are fairly distributed.
- Defining and agreeing how data is used, with enforcement mechanisms if an agreement is breached.
- Ensuring people with the necessary skills are managing the data.
- Providing assurances about data quality, provenance and timeliness.
- Using anonymisation or other privacy enhancing techniques to enable access to data while preserving privacy or respecting commercial sensitivity.
- Ensuring that storage and transmission of data is secure.

Ways of facilitating the discovery of and access to data, where it is collected by more than one organisation, are also needed. For example, data platforms may be used to host and manage shared data from multiple organisations, and data may be accessed using application programming interfaces (APIs). Standards help enable interoperability between datasets, but also between data platforms so that it is possible to search for and access data wherever it is hosted.

Ideally, the data format and quality are appropriate to the purpose but in practice, data may need to be cleaned or converted to the right format to make it usable. Approaches to guaranteeing the provenance of data and to data linkage are also required. Good data management is at the heart of successful data sharing.

Data sharing may underpin new business models. New sources of value originate from the novel uses to which shared data is put, or alternatively from the activities that enable data sharing. New roles are emerging such as data enhancer or data broker. Intermediaries with their own business models may catalyse the data sharing opportunity or develop the technologies that enable it.

Above all, strong oversight ensures the opportunity is realised well and that it meets commercial, regulatory, legal and ethical requirements. Where many organisations come together to share data, ideally all would be represented in the oversight body. The participation of consumer organisations

or patient associations may be appropriate where personal data is used. Oversight mechanisms are sustainable if they can be maintained even when one participating company is purchased by another. Transparency and accountability are vital so that an independent body can scrutinise the data sharing arrangements and the enforcement of any rules.

As data sharing becomes more pervasive and new ways of sharing data emerge, broader questions need to be addressed, such as whether access to data is equitable and promotes or stifles competition, and what the implications are for society and the economy.

Outstanding challenges

This work illustrates the multiple technical, economic and governance issues involved in setting up a data sharing solution. Each of the case studies in this publication have tackled some or all of these issues. They provide insight into how these are being addressed in the real world and possible approaches to finding solutions.

The challenges have not been completely solved, however. For example, there is still uncertainty about defining data rights and ensuring that data acquired by one user is not copied and distributed further. Some business models remain unproven.

The effort required to share data often reflects the imperfect context in which data sharing occurs: inadequate data quality requires data cleaning effort, a lack of standard data formats makes data linkage tricky, and an inadequate legal framework makes the creation of robust data sharing agreements challenging. As these aspects and others are increasingly addressed, the friction involved in data sharing will decrease.

Introduction

This publication illustrates how data sharing can improve access to valuable data, which would otherwise remain locked away in corporate siloes or within sectors.⁴

A data-sharing checklist guides organisations through the main areas for consideration as they work towards a solution. It is aimed at organisations that have identified the opportunity to create value through sharing data and are considering setting up arrangements for data sharing.

Several key practical challenges are discussed, including data management, integration and linkage. The engineering components required to enable data sharing are conceptualised in a diagram of a data-sharing system.

The importance of improving access to data through data sharing is recognised by government in its Digital Strategy and in the Artificial Intelligence (AI) Sector Deal. This report explores the policy context for data sharing and draws out implications for policy.

A series of 10 case studies, across a range of applications and sectors, shows how organisations and people have come together to share data under controlled conditions, and how value is created as a result. The case studies explore challenges around business models, technical design and data curation, and legal and commercial arrangements.

Data sharing

Data sharing has existed in specific sectors and contexts for many decades, although limited by the amount of data available and the number of players involved. New technologies and digitisation have dramatically increased the type and volume of data available. The capability to collect and analyse static and real-time data, along with reduced storage and processing costs, has opened up the potential to create significant new value with and across industries and sectors.

New approaches are emerging that facilitate controlled data sharing. These approaches help to break down data silos and enable access to data and associated data services that add value. They need to address commercial, legal, regulatory, ethical and technological challenges, and require new workforce skills.

Data sharing may occur in many different forms. The main characteristics of data sharing include the following:

- Data sharing generally refers to the exchange of data between two or more organisations, occurring in one or more directions.
- There may or may not be a financial transaction associated with this exchange.
- Reciprocal arrangements sometimes exist, whereby participants share their data and in return receive similar data from other organisations.
- The exchange of data usually comes with conditions about where and how the data is used, for example.
- Raw data, augmented data, derived data or a solution based on data may be shared.
- Data sharing can also refer to the integration of data-based operations between two or more organisations.

Data needs to be managed

A vital element in creating value from data is a robust information management approach, where the business requirements are clearly set out, and the relevant data that meets the requirements is assembled, structured and managed over its lifecycle.

If data sharing is required, the question of how data will be integrated or linked is also key. If the information management is done well, less resource will be needed to share data effectively. These aspects are discussed further in the practical challenges section.

Data sharing is only part of the picture

Data sharing is part of the 'data value chain' (Figure 1), which describes how value increases as data shared and transformed into insights, leading to better decision-making, improved processes, and the development of innovative products and services.

The potential for data sharing is one of many considerations when an organisation investigates whether the relevant data is available to meet business requirements. A project decision tree for data-enabled business change (Figure 2) sets out key points in the decision-making process, and indicates where considerations about data sharing occur in that the process.

Figure 1: The data value chain

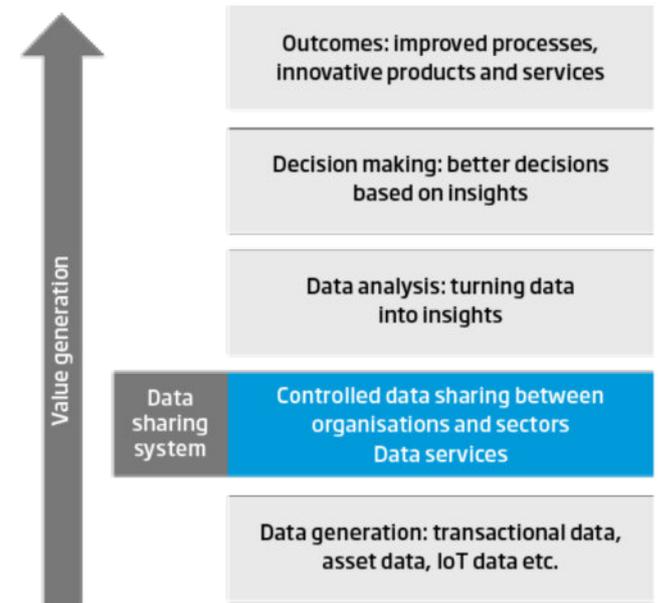
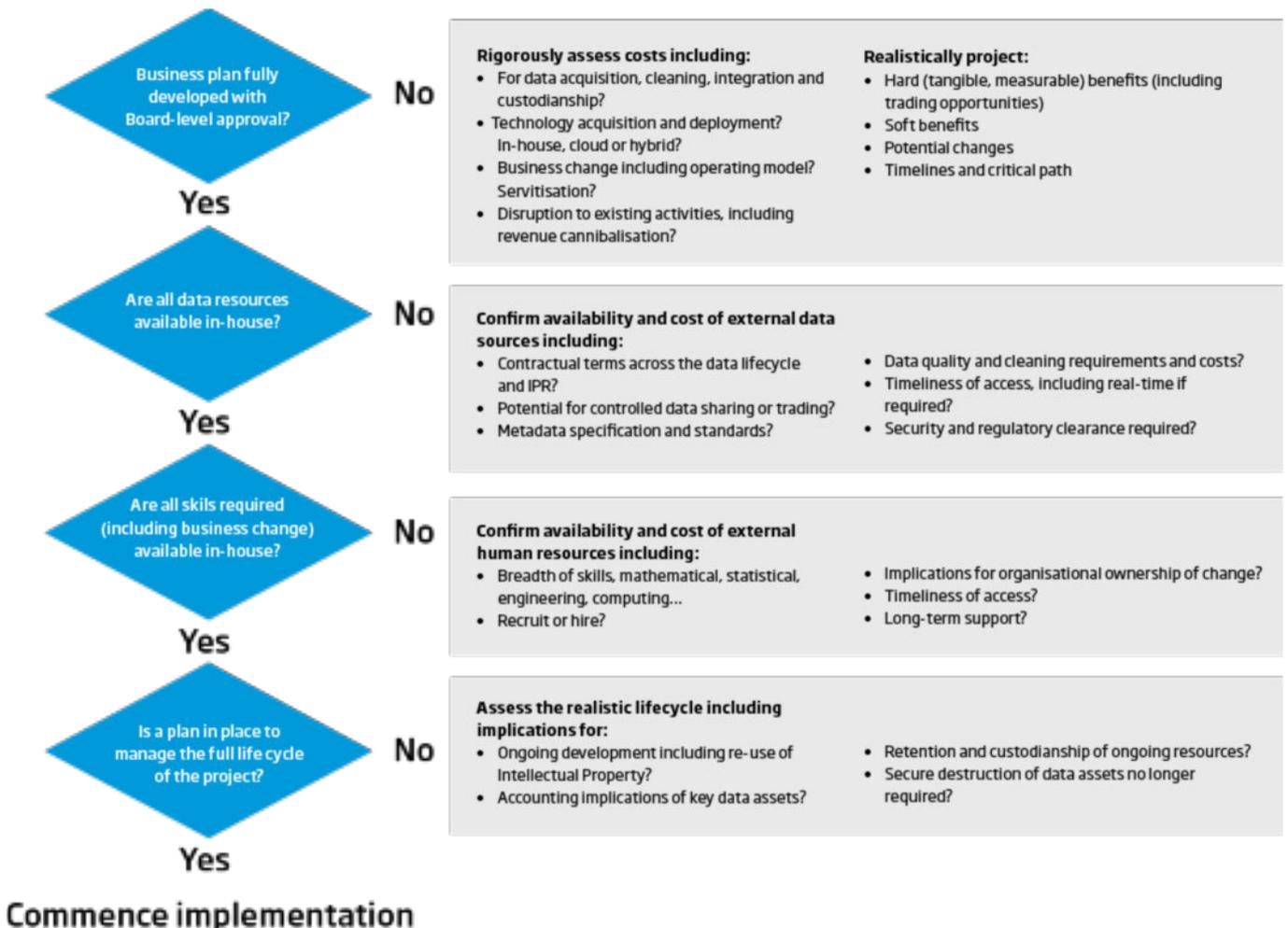


Figure 2: Project decision tree for data-enabled business change



Practical challenges

This section sets out some of the key practical challenges in data sharing. In some cases, these require further development effort to create a working system of data sharing or better data management, rather than more research.

Good data management

Good data management is a critical part of effective data sharing. It must respond to an organisation's business requirements and consider data's full lifecycle. Data sharing may be just a small, but potentially significant, part of the lifecycle management. Considerations such as whether the data is being updated, or how it might be securely destroyed, are relevant here.

The quality of data and metadata is a key consideration, as it in turn affects the quality of the data analytics and the confidence to make robust decisions based on the outputs. In the case of personal data, the General Data Protection Regulation includes requirements around data minimisation, the right to be forgotten and the right to an explanation, which all need to be addressed as part of data management.

If data is shared, it will be vital to ensure that data is being managed in line with data-sharing agreements between parties exchanging and using the data. Commitments to manage and use data appropriately may extend over many years. This is even more important where tensions exist, for example, between the need for an organisation to meet its own aims and the need to maintain privacy or meet other regulatory or legal requirements.

Data integration and data linkage

Ideally, shared data can be considered integrated if it can be treated as if it came from one system. It needs to be accessible and it should be possible to query the data meaningfully in the same way wherever it originated. The use of integration data modelling techniques may be useful here.⁵

In practice, standard data formats may not exist, and therefore approaches that facilitate data sharing and linkage will be required. These can be considered in two stages.

The latter stage requires specialist skills, and indeed data linkage is becoming a skill set in its own right.

Facilitating data sharing and linkage

Stage 1:

Creation and storage of data in a way that makes it easier to share. For example, raw data in spreadsheet form is easier for others to use than data in pdf format. A vital part of this is clear, readable and accessible metadata describing the data.

Stage 2:

Where datasets from different sources are being linked, the development of strategies, methods and tools for combining datasets.

Engineering the enabling functions of data sharing

Internal components need to be engineered to enable data exploration, interoperability, identity management, quality control, and monetisation, for example. These enabling functions must include metadata management, creating data catalogues, managing access to data, and handling contractual obligations to destroy data, for example. Governance arrangements and technologies ideally work within the necessary legal structure, while delivering the intended benefits to participants. The case studies illustrate solutions that involve multiple participants, enabling functions and types of transaction, underpinned by enabling technologies.

Figure 3 (below) attempts to conceptualise the challenge with a diagram of a data-sharing system, illustrating the range of type of participant, enabling functions, transactions (either financial or data) and enabling technologies. Enabling functions are grouped into related activities and colour-coded accordingly. In practice, a data-sharing arrangement will include some combination of these elements, which depends on the application and the nature of the data being shared. Figures 4 and 5 (below) illustrate this for two of the case studies, the Data and Analytics Facility for National Infrastructure (DAFNI) and Smart meters.

Data sharing and digital twin technology

Data sharing is a key element of the plans to create a national digital twin for infrastructure - a federation of digital twins that will enable better decision-making in the delivery, operation, maintenance and use of infrastructure.⁶ A digital twin is a digital representation of entities such as assets,

processes or systems. Data from these entities is fed back into the digital twin, which in turn supports improved decision-making about that entity.

Digital twins may exist for a variety of purposes and operate at a range of scales. The ability to share data in a secure, resilient and interoperable way between many digital twins is a requirement of the national digital twin and requires the development of an information management framework.

Figure 3: Data-sharing system

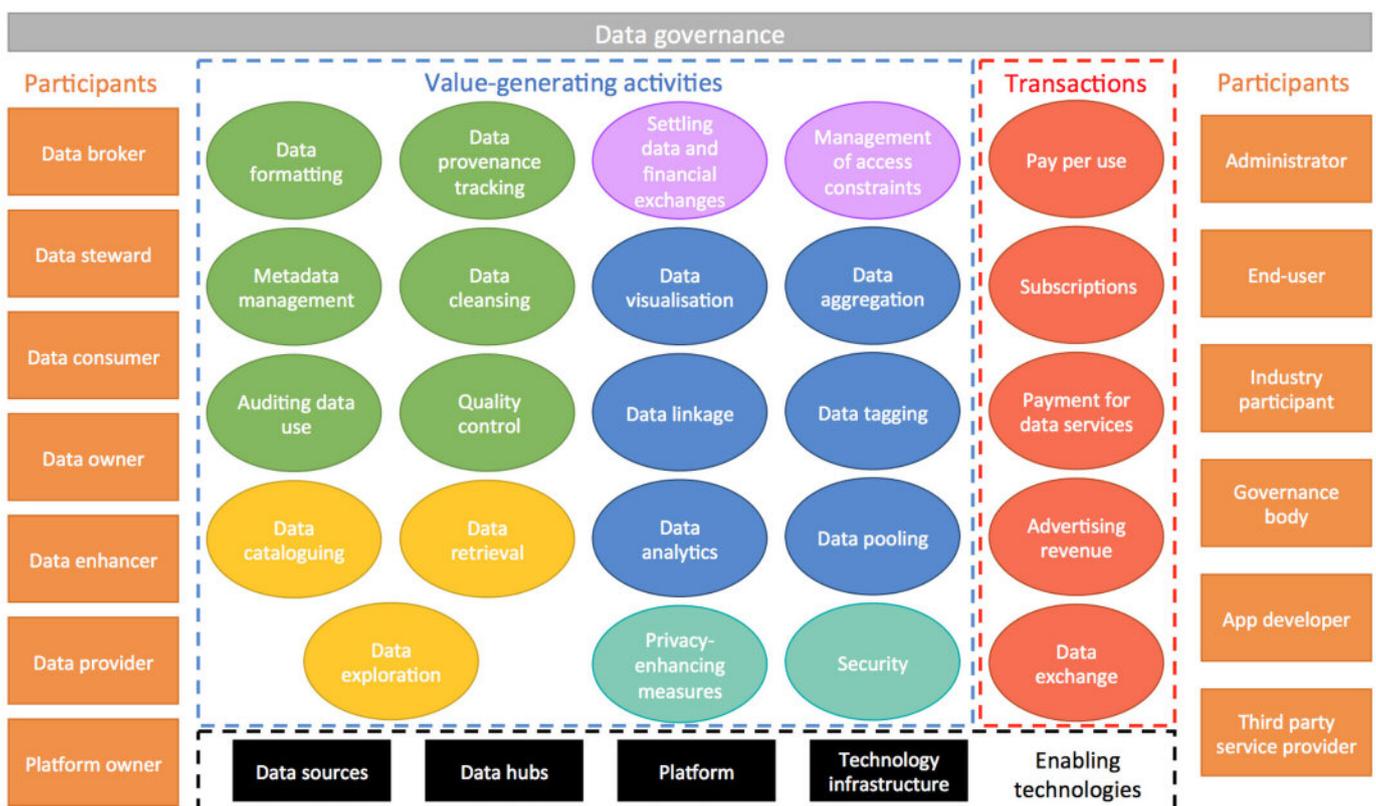


Figure 4: DAFNI

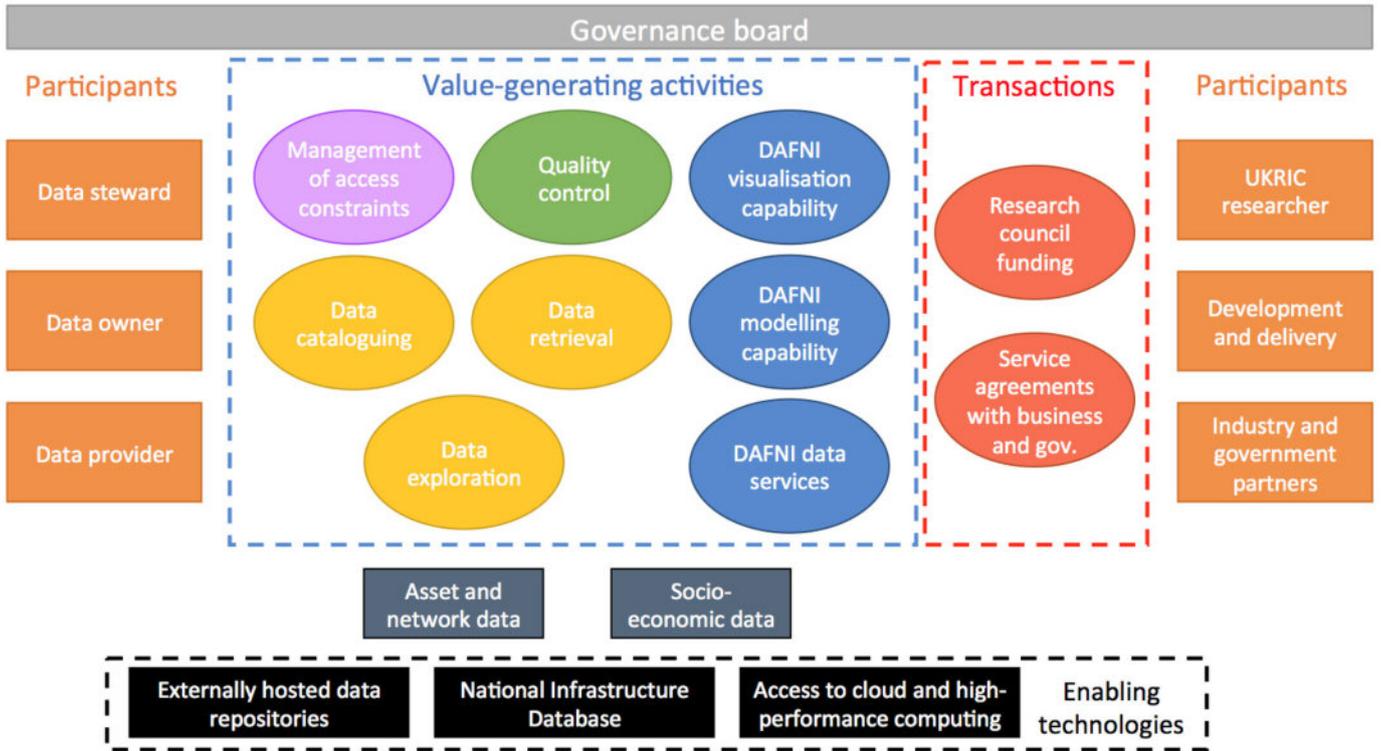
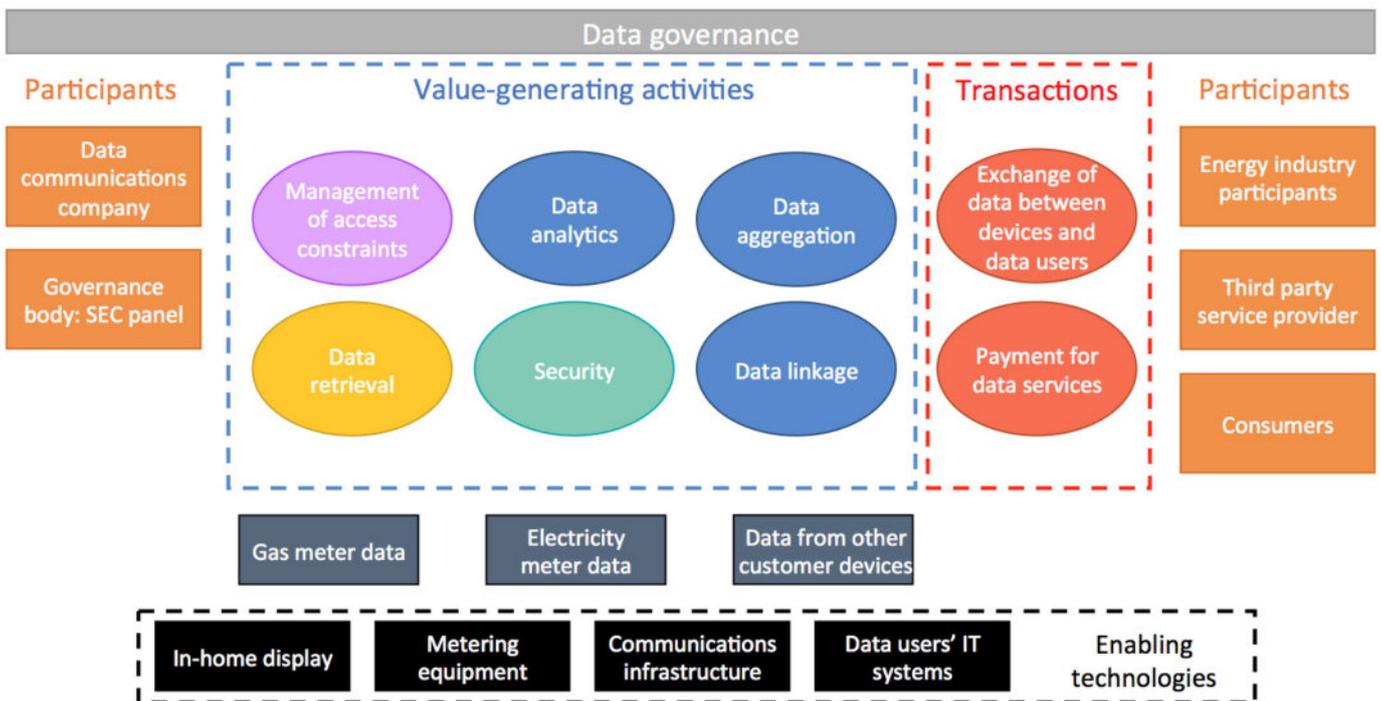


Figure 5: Smart meters



Policy context

Data sharing is widely recognised by government and other stakeholders as a key enabler for unlocking the value of data:

- The government's Digital Strategy emphasised the importance of supporting the data economy through 'unlocking the power of data'.⁷ It recommended the creation of a strong data infrastructure - assets, technology, processes and organisations that allow data to be opened up and shared - and recognised the potential for businesses to make better use of the data that they hold.
- The World Economic Forum Global Risks report identified the importance of having a predictable legal framework in order to facilitate the exchange of data between countries and stakeholders, to realise the full economic potential of digitisation.⁸
- More recently, the independent review of artificial intelligence (AI) highlighted the need to increase the ease of access to data in a wider range of sectors, to continue developing and applying AI.⁹
- The AI sector deal acknowledged the need for the right data infrastructure by tackling the practical and cultural barriers to sharing both publicly and privately held data, and to explore data-sharing frameworks that might protect sensitive data, ease access to data, and ensure accountability.¹⁰
- HM Treasury recognised the need for data to flow freely between economic actors to realise its true potential value and identified 'enabling safe, legal data-sharing' as a key challenge.¹¹
- The free flow of data across international boundaries is also a key component in creating a thriving data-driven economy, as has been recognised by the European Commission and by business.^{12,13}

Implications for policy

Barriers to data sharing - and to the emergence of 'data ecosystems' - are widely recognised. They include:

- the difficulty of creating clear contractual agreements regarding ownership and use of the data and the allocation of value
- technical issues such as ensuring adequate data quality, and enabling data integration or linkage. This are discussed further in the section on practical challenges
- legal and regulatory issues including around what data can be shared¹⁴
- rights management, which is of particular importance, especially where the data propagates and attribution is necessary once datasets have been combined and sent on.

The case studies illustrate that some very important work has already been done within specific projects on vital areas such as reference data architectures, mechanisms to enforce constraints around how data is shared and used, and methods for preserving personal privacy. Other challenges include enabling viable business models and ensuring interoperability of data and data platforms. However, not every challenge has been fully solved.

Trusted data sharing

Mutual trust among both big and small players is important, and therefore mechanisms that enhance trust among all transacting parties are vital. These mechanisms should enable transparency and control.¹⁵ They go beyond establishing intellectual property rights and securing data-sharing agreements and include: technologies and architectures that facilitate security or privacy; robust processes for checking quality; mechanisms for ensuring the various parties are compliant with terms and conditions; and ensuring various parties having the skills and resources to deliver their parts of the agreement. A user-centric approach,

with easy-to-understand tools and policies, also helps to facilitate participation and enable trust. Trust mechanisms need to be put in place as quickly as possible.

The development of robust and trustworthy frameworks for sharing data would make these practices more acceptable to companies and other types of organisations, and the public - a key consideration if personal data is being shared.

The concept of a 'data trust' has emerged as a possible mechanism for enabling trusted data sharing, although its precise definition and extent of its applicability is a matter for debate at the time of publication. Data trusts could encompass some combination of legal terms, governance arrangements and technologies used to access data, or it could purely be a legal structure that sets out the relationships between the different parties.¹⁶ Data trust pilots will help to refine the definition and test the practical implications of the concept.

Central hubs for data sharing

The case studies demonstrate that the use of a central hub can facilitate trusted data sharing. The central hub needs its own business model and governance framework to be sustainable and to underpin trust. It may help to catalyse the formation of an ecosystem of data owners, providers and consumers, alongside third parties who may take on myriad roles such as data brokering, data cleansing, data aggregation, certification and analytics.

The development of an ecosystem in which a range of players can participate in the governance is preferable to one that is governed by a single big player.

Data sharing agreements

The development of standardised terms and conditions will minimise the resources needed to develop data-sharing agreements between parties, as long as they reflect a standardised or harmonised understanding of data usage rights and responsibilities. Such a harmonised understanding

is vital, so that 'two or more parties in any sector can partner in data sharing agreements, shape the agreements according to their needs and enable multiple organisations to work together to solve a common problem'.¹⁷

Data agreements must address a broad set of issues including: data quality, timeliness and lifecycle; compliance with the governance rules; and enforcement of usage rights.¹⁸ Defining the requirements for data quality, and ensuring these requirements are delivered, remains a central challenge.

Data ecosystems

Data ecosystems where data is pooled between organisations are found to have a number of features:¹⁹

Features of data ecosystems

- (a) clearly defined boundaries that enable the identification of a legitimate user;
- (b) rules regarding how data resources can be used offline or outside the transaction;
- (c) opportunities for contributors to participate in the development or improvement of the platform;
- (d) effective monitoring by a group of core users or a third party accountable to the core users; and
- (e) rules that define how the resources are to be used and the penalties for misuse.

In addition, an independent oversight body would ensure that individual organisations pooling their data have a measure of confidence in the running of the overall system. These attributes are likely to extend beyond data pooling models to other models of data sharing, such as those illustrated by the case studies.

Government's role

Government can play several roles in helping facilitate the data-sharing ecosystem. It already plays a role in setting standards and making government data accessible and usable. Government can also encourage private sector organisations to collect and release data, leading by example and sharing best practice. It can improve access to the skills that are required for identifying opportunities for data sharing, and for developing and implementing data sharing models.

This requires business skills, and the soft skills required to lead, be part of multidisciplinary teams and work in partnership. It also requires technical skills for developing appropriate architectures and applications of technologies, and for data engineering and linkage. It can create the parameters as a regulator, and fund platforms and infrastructure for data exchange, as well as pilots for new models of data sharing.²⁰

Acknowledgements

The project was overseen by a working group:

Professor Jim Norton FREng, Chair of the working group

Professor Mandy Chessell CBE FREng, IBM

Professor Mark Girolami, Alan Turing Institute

Professor Aija Leiponen, Cornell University

Alastair McMahon, O2 Data and Insights

Hetan Shah, Royal Statistical Society

Dr Mike Short CBE FREng, Department for International Trade

Dr Liane Smith FREng, Larkton Ltd

Dr Jeni Tennison OBE, Open Data Institute

Professor Jeremy Watson CBE FREng, University College London

Steve Yianni FREng, TWI and AIRTO

with support from the following staff:

Dr Philippa Westbury, Royal Academy of Engineering

Dr Alexandra Smyth, Royal Academy of Engineering

Dr Alan Walker, Royal Academy of Engineering

The Academy would like to thank the following individuals who were interviewed as part of the case study research:

Dr Hamed Haddadi, Imperial College London (Databox)

Professor Chris Taylor OBE FREng, Carmel Dickinson and Dr John Rigby, University of Manchester (CityVerve Manchester)

Nick Chrissos, Cisco (CityVerve Manchester)

Tim Gammons, Arup (oneTRANSPORT)

Professor Dr Boris Otto, Fraunhofer Institute for Software and Systems Engineering (Industrial Data Space)

Professor Jim Hall FREng, University of Oxford (Data and Analytics Facility for National Infrastructure)

Martin Aston, Airbus (Advanced Product Concept Analysis Environment)

Ashley Pocock, EDF (Smart meters)

Dan Wolfson, IBM (The Weather Company)

Katie Wilde, University of Aberdeen (Grampian Data Safe Haven)

Professor Enrico Motta, Open University (MK:Smart)

The Academy would like to thank the following reviewers:

Ruth Boumphrey, Lloyd's Register Foundation

Alastair Cook, Critical Insight Security Ltd

Professor Jon Crowcroft FREng FRS, University of Cambridge and Alan Turing Institute

Professor David Hand FBA, Imperial College London

Professor John Loughhead CB OBE FREng, Department for Business, Energy and Industrial Strategy

Paul Martynenko, British Computer Society

We would also like to thank:

Dr Pantelis Koutroumpis and Dr Llewellyn Thomas, Imperial College London and Gordon Blunt, Gordon Blunt Analytics for their contributions to the project

References

1. Royal Academy of Engineering and IET (2015), *Connecting data: driving productivity and innovation*, <https://www.raeng.org.uk/publications/reports/connecting-data-driving-productivity>
2. Office for Statistics Regulation (2018), *Joining up data for better statistics*, <https://www.statisticsauthority.gov.uk/publication/joining-up-data/>
3. Royal Academy of Engineering and IET, *Connecting Data - case studies: where are the success stories?* Case study: Energy - optimising performance in the energy industry, iQRA: Wood Group Intetech <https://www.theiet.org/sectors/information-communications/topics/connected-data/articles/connecting-data.cfm?>
4. This was a key finding in the joint RAEng/IET report, *Connecting data: driving productivity and innovation* (2015)
5. Government Office for Science and Council for Science and Technology (2018), *Computational modelling: technological futures*, Box: challenges and opportunities for modelling the built environment, p.83 and p.84 <https://www.gov.uk/government/publications/computational-modelling-blackett-review-and>
West, M. (2011), *Developing high quality data models*, Elsevier.
6. Blog: Mark Enzer, Chair of the Digital Framework Task Group, July 2018, www.cdbb.cam.ac.uk/Blog/2018JulyBlogEnzer
7. Department for Culture, Media and Sport and the Rt Hon Karen Bradley (March 2017), *UK Digital Strategy*, www.gov.uk/government/publications/uk-digital-strategy
8. World Economic Forum Global Risks Report 2016, 11th Edition, www3.weforum.org/docs/GRR/WEF_GRR16.pdf
9. Professor Dame Wendy Hall DBE FREng FRS and Jerome Pesenti (October 2017), *Growing the artificial intelligence industry in the UK*, www.gov.uk/government/publications/growing-the-artificial-intelligence-industry-in-the-uk
10. Department for Business, Energy and Industrial Strategy and Department for Digital, Culture, Media and Sport (April 2018), *AI Sector Deal*, www.gov.uk/government/publications/artificial-intelligence-sector-deal/ai-sector-deal
11. HM Treasury (August 2018), *The economic value of data*, www.gov.uk/government/publications/the-economic-value-of-data-discussion-paper
12. EU Digital Single Market, *Free flow of non-personal data*, <https://ec.europa.eu/digital-single-market/en/free-flow-non-personal-data>
13. GSMA (September 2018), *Cross-border data flows: realising benefits and removing barriers*.
14. Thomas, L.D.W and Leiponen, A (2016), *Big data commercialisation*, IEEE Engineering Management Review, 44(2), December 2016.
15. Thomas, L.D.W and Leiponen, A (2016), *Big data commercialisation*, IEEE Engineering Management Review, 44(2), December 2016.
16. ODI (July 2018), *What is a data trust?* <https://theodi.org/article/what-is-a-data-trust>
17. HM Government (2018), *AI Sector Deal*.
18. See, for example, ICO, Data sharing code of practice [currently being updated following implementation of the 2018 Data Protection Act] https://ico.org.uk/media/for-organisations/documents/1068/data_sharing_code_of_practice.pdf
19. Thomas, L.D.W and Leiponen, A (2016), *Big data commercialisation*, IEEE Engineering Management Review, 44(2), December 2016.
20. Thomas, L.D.W and Leiponen, A (2016), *Big data commercialisation*, IEEE Engineering Management Review, 44(2), December 2016.