Towards trusted data sharing: guidance and case studies

Case studies

The following case studies are 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

**Case study 1**
**Databox**  
Allowing individuals to control how they share data with other parties  
page 3

**Case study 2**
**CityVerve Manchester**  
A platform of platforms for smart city data sharing  
page 6

**Case study 3**
**One TRANSPORT**  
Federated platforms for sharing suburban and rural data  
page 9

**Case study 4**
**Industrial Data Space**  
Data sharing within a secure and trusted ecosystem  
page 13

**Case study 5**
**DAFNI**  
Data and analytics facility for national infrastructure  
page 17

**Case study 6**
**APROCOONE**  
Integrated product design in the aerospace industry  
page 20

**Case study 7**
**Smart meters**  
Data sharing in the energy industry  
page 23

**Case study 8**
**The Weather Company**  
Data services for sectors where weather has an impact  
page 26

**Case study 9**
**Grampian Data Safe Haven**  
A facility for sharing sensitive health data  
page 29

**Case study 10**
**MK Data Hub**  
A data platform for the city of Milton Keynes  
page 32
Case study 1

Databox

Allowing individuals to control how they share data with other parties

Databox aims to increase consumer trust in the use of personal data by organisations by enabling transparency and control. Individuals generate many different types of personal data from mobile phones, smart meters and media streaming services, for example. In return for letting organisations access this data, individuals are given insights into what they have been doing or receive useful advice. Data may also be processed with other people’s data, with additional benefits to the individual and other parties. Databox enables individuals to limit the ways in which their data is used and understand the implications of any data release.

Summary: the eight dimensions of data sharing

The opportunity:
Databox enables individuals to control access to their personal data by service providers. It allows service providers to pre-process data to ensure shared data is minimised and possibly desensitised.

The data and its use:
Consumer data such as Internet of Things (IoT) data from home services or smart meters can be put to multiple uses.

The business model and value creation:
Consumers will be able to obtain insights from their own data. Commercial organisations will have access to a greater range of data sources. Databox has no fixed plans for commercialisation at the time of this research.

The model for data sharing and the partnership:
Technology facilitates direct interaction between consumers and third-party service providers. It has been developed by a university consortium with industry involvement, and has future commercialisation potential.

People with the right skills and expertise:
Databox has used academic expertise in containerisation and virtual machines, data analytics, human-computer interaction and accountability in the IoT ecosystem, as well as domain expertise from industry partners.

Constraints on how data is shared and used:
GDPR and acceptability from the consumers’ perspective constrain how data is shared and used. Databox enables adherence to Privacy by Design and data minimisation requirements.

The data architectures and technologies:
Databox uses ‘containerisation’ technology to mediate access to data held ‘at the edge’ rather than in the cloud. Its interface helps users understand what an app can analyse based on the data that they are willing to give.

Governance / oversight / enabling trust:
The consumer has control over which data is shared and understands how data will be used. Only the results of the processing are provided to the individual and app developer, after which the data and app are ‘killed’.
Introduction

Databox is a multi-partner research project funded by a £1.2 million Engineering and Physical Sciences Research Council (EPSRC) grant that runs for three years until October 2019.¹ The project came about as a result of a number of converging factors including the emerging data protection regulatory regime, the development of ‘containerisation’, and an understanding of the value of personal data and the need to provide individuals with control of their own data.² The project also built on precedents such as Dropbox, a platform that enables people to share data and gives others access to it. Databox allows individuals to manage, log and audit access to their personal data by other parties, countering the prevailing practices of aggressive data harvesting by certain companies. In return, they receive useful services.

The consortium of universities carrying out the project comprises Imperial College London, the University of Cambridge and the University of Nottingham. These partners bring technical expertise in containerisation and virtual machines, data analytics, human–computer interaction and accountability in the IoT ecosystem. Industry involvement includes the BBC, BT, Internet Society and Microsoft, as well as Telefonica and Open MHealth.

Data is processed locally ‘at the edge’, rather than in the cloud, with resulting computational and social advantages.³ The former arise from reducing the need to transport large volumes of data over communications networks for processing at remote data centres. These datasets may originate from multiple connected devices and other sources. A social advantage is maintaining privacy and safety by reducing the risk of data breaches when data is being distributed via a network.

Both consumers and commercial organisations will benefit. Consumers will be able to obtain insights from their own data, while commercial organisations will have access to a greater range of data sources of appropriate type or granularity, enabling richer and more accurate analytics.

The Databox technology mediates access to the source of data, rather than holding data. Value is extracted from the data by apps that will be developed by third parties. Individuals can specify which bits of data the apps can access. Once the data has been analysed, the results are sent to the individual and the app developer. The data itself is not kept, enabling rich analytics while limiting access to the data.

Databox uses ‘containerisation’ technology provided by Docker.

Other projects with similar aims to Databox include the Hub-of-all-Things and CitizenMe.⁴,⁵

Business model

The project is at an early stage and has no fixed plans for future commercialisation, although there is interest from venture capital firms. Monetisation of personal data is not being explicitly considered as part of the project.

Large-scale analytics, such as providing market researchers with data from multiple people without the need for data brokers, could create value in the future. Utility companies would benefit if, for example, they could access data sources other than smart meter data. As more sources of data become open, value that could be obtained from the box will be increased.

For example, access to financial information would be greatly improved if banks gave their customers the ability to share transaction data with third parties.⁶ An increase in the number of smart devices would also add richness to the analytics that are possible.

Apps would be paid for by the consumer, or by a commercial organisation such as a market research company. A public-sector organisation such as the NHS could create an app to enable data collection for the public good, for example, for health research.

Technical and data curation arrangements

The platform under development is open source, although the apps developed by third parties that process or analyse data will not be. Data may be from local or remote sources, such as online social networks or IoT sensors. The project is developing libraries that provide functionality for other apps to access the data in the form that they need, so that the app itself does not need drivers or other ways of dealing with the data.

‘Data negotiability’ is a central concept: the project team is developing an interface that allows the user to understand what the app can and cannot analyse based on the granularity or type of data that the user is willing to give access to. For example, the sampling rate may influence the ability of the app to infer certain information sufficiently accurately for appropriate decisions to be made. This is an area of research in the field of information theory.

Another important part of the project is how to ensure transparency for the user about who has got access to which data. Capabilities for intensive logging and auditing of access, operations, apps and data sources are being developed. One challenge is how potentially detailed and complex information is communicated to the user. User studies are planned to investigate how best to present such information. The project should enable organisations to be compliant with new data protection laws.
Towards trusted data sharing: guidance and case studies

Case study 1

One area for discussion is whether Databox carries out aggregation and archiving of data itself, or whether it allows an app to carry out these functions. Databox could have an archiving facility for data that is not being stored elsewhere, with a certain level of aggregation capability. However, the intention is that Databox does not become a ‘honey pot’ that attracts potential hackers. Data will be stored on multiple servers.

It is intended that security will be included as a feature in Databox. The project is exploring the concept of Databox as a kind of ‘firewall’, so that it is possible to control the exposure of IoT devices to the internet, and for example ‘whitelist’ destinations for data. Apps are contained and isolated from the network and they write inferences to an export module – another container that is also secure, and the output of the apps can be monitored. The assurance of apps is an intensive process and there are plans to vet every app launched.

Legal and commercial arrangements

It is envisaged that there will be some kind of service level agreement between the consumer and the app developer that underpins what data the app is allowed to access, the data rate, where it will be stored and the service that will be provided.

The research project is not investigating in detail legal issues such as liability, although the legal grounds will need to be addressed before the project becomes operational. Liability will be minimised by ensuring Databox is not created as a ‘honey pot’.

Outcomes and lessons learned

- The value to consumers and organisations will come from the ability to mix data types and sources that cannot currently be brought together. For example, the ability to correlate an individual’s physical activity with their sleep quality will be of value to that individual in enhancing their health and wellbeing.

- A key challenge is ensuring that users understand the value of their data, why privacy is important, what the nature of the data accesses are and the consequences of data accesses. People are not very risk-aware, for example, they may be unaware of, or indifferent to, sharing personal data with companies via connected home devices.

- A further challenge is enabling meaningful engagement by individuals in the management and sharing of their personal data.
CityVerve Manchester is a smart city demonstrator. It uses Internet of Things (IoT) technologies to help Manchester City Council improve the way it designs and delivers services for the people who visit, live or work in Manchester. Projects focus on four themes: culture and public realm; health and social care; energy and the environment; and travel and transport. The ‘platform of platforms’ allows the city and other organisations to access multiple sources of data, with controls placed on how data is accessed and used. It enables the application of data analytics across domains and sectors, and creates the potential for the city and other organisations to develop business models around data and smart services.

Summary: the eight dimensions of data sharing

<table>
<thead>
<tr>
<th><strong>The opportunity:</strong></th>
<th><strong>The data and its use:</strong></th>
<th><strong>The business model and value creation:</strong></th>
<th><strong>The model for data sharing and the partnership:</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Data from multiple sources is brought together to improve city services to benefit the city of Manchester and its inhabitants. A similar solution could potentially be applied to other smart cities.</td>
<td>IoT data plus other sources. It has multiple uses across four themes: health and social care; energy and environment; travel and transport; and culture and public realm.</td>
<td>Primarily, CityVerve is a demonstration of technology; however, the technology provides the tools and methods for future business models around services for the city and citizens. There is potential to monetise data via the platform.</td>
<td>Technology provides a common platform for finding and accessing data by the city authorities and third parties. A public and private sector partnership is led by Manchester City Council with Cisco as the main technology partner.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>People with the right skills and expertise:</strong></th>
<th><strong>Constraints on how data is shared and used:</strong></th>
<th><strong>The data architectures and technologies:</strong></th>
<th><strong>Governance / oversight / enabling trust:</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Technology expertise comes from primary technology partners and SMEs, alongside academic expertise in analytics and domains of application. City partners have domain expertise.</td>
<td>GDPR, regulations on sharing healthcare data and commercial sensitivity are all constraints to be considered.</td>
<td>GDPR, regulations on sharing healthcare data and commercial sensitivity are all constraints to be considered.</td>
<td>GDPR, regulations on sharing healthcare data and commercial sensitivity are all constraints to be considered.</td>
</tr>
</tbody>
</table>
Introduction

CityVerve Manchester is a £10 million project funded by the Department for Digital, Culture, Media and Sport and Innovate UK, which ran between July 2016 and July 2018. The consortium of 21 organisations was led by Manchester City Council. Cisco was the lead industry partner, and worked alongside a number of other technology companies. Manchester Science Partnerships was the business engagement lead, and the University of Manchester was the project evaluator, among other roles.

Data-sharing initiatives already existed in Greater Manchester before the project was initiated, providing existing partnerships and a strong basis upon which to build the project. Two important components are Datowell, an innovative digital technology project that allows doctors and care providers to share patient and client information, and GM-Connect, a data-sharing authority launched in 2016 by the Greater Manchester Combined Authority, that aims to help break down the barriers that prevent public services sharing information.9, 10

CityVerve built on earlier smart city projects by improving the way in which smart applications are linked together, thus helping to break down silos between applications. The project developed a ‘platform of platforms’, a unifying layer that acts as a secure ‘catalogue’ and precludes the need to collect together different datasets from multiple platforms in one repository. Instead, the data is held on a series of ‘federated’ platforms. The ability to use data generated in multiple ways and apply analytics across domains and sectors was a key aspect of the project.

The project investigated a number of use cases in the areas of health and social care, energy and environment, travel and transport, and culture and public realm. The data collected was driven by the use cases, but further value was generated if that data was used by another use case or if other sources of data were used. Data from other smart city initiatives such as Triangulum was also exploited, as were other existing sources.11

For example, researchers involved in use cases on chronic obstructive pulmonary disease aimed to collect data about individuals’ symptoms and experience alongside environmental data such as weather data, with the aim of providing individuals with personalised feedback to help them to manage their condition. Under the energy and environment theme, sensors were retrofitted into buildings and linked to the building management system to analyse energy use and occupancy. Multiple energy-usage data sets across an area were brought together for a study on local area energy management. Cross-domain data analytics was applied by the University of Manchester, drawing on both analytics and domain expertise from Manchester Informatics, the university’s institute for digital research. Individual data hubs were deployed by the various tech companies involved in the project. For example, the BT hub collected transport data. Another organisation, Asset Mapping, carried out a project to identify, monitor and visualise all city assets on a map, using a scalable database.

Other UK smart city projects include MK Smart (case study), Bristol Is Open and Future City Glasgow.12, 13

Business model

The main purpose of CityVerve Manchester was to demonstrate technology that could be transferred to other smart cities. Companies that were involved were interested in creating a demonstrator of reusable technology that they could then find markets for elsewhere. The focus of the project was therefore less about specifically developing business models.

However, the technology provides the tools and methods for new business models to become a reality in the future. It creates the potential for the city or other organisations to generate new business models around data or smart services. For example, new business models and forms of partnership are emerging whereby third parties are offering services to cities, taking on the risk and sharing profits with the city. Companies involved in developing individual data hubs will also be developing their own business models. In addition, the data has value for research organisations.

Technical and data curation arrangements

Cisco’s platform of platforms sits above individual platforms or ‘data hubs’ to provide a single way of accessing and connecting data from different sources. The project addressed the challenge of building a technology that can scale – potentially incorporating an unlimited number of data hubs – and that can be deployed and managed easily and securely.

The platform incorporates ways to enable policies about data access and use to be enforced, for example if the data is commercially sensitive or personal. Individual data controllers specify the rules for data access and use, and the platform will be able to accommodate their various requirements. It will, however, be challenging to enforce policies around linkage and analysis of data. The platform will allow data controllers to restrict access to data if it is not being used in the agreed way.

The project explored capabilities such as data provenance tracking and auditing who has accessed data. A further capability that was explored was local data processing – where data is analysed close to sensors and actuators - allowing local autonomy and precluding the need to transfer
data to and from the cloud. Three aspects of security were addressed: authentication, integrity and confidentiality.

The platform connects data from many different sources, but does not itself copy the data. A central challenge is how to retain control of data if it is copied and used outside the platform, which is common to many other situations beyond CityVerve. For example, it has implications for implementing the ‘right to forget’, one of the principles of the GDPR.

HyperCat has been mandated as the standard that underpins the ‘platform of platforms’ and the catalogue, which contains descriptions of what data exists and where it resides. The standard has been enhanced within the CityVerve project to address additional challenges, such as the introduction of a bi-directional capability that allows data to be written or modified as well as read - therefore enabling control of what is happening ‘in the field’ - and the ability to define real-time data streams of data.

A means of monetising data is being developed, so that if access to data is requested, it is possible to assign a commercial model to the request in an automated way. In future, it is real-time data generated by sensors that will generate most value, rather than static datasets held in databases. An ecosystem could potentially develop that includes data providers and those who sell solutions based on data.

**Legal and commercial arrangements**

Partners signed a collaboration agreement that ensured they were equally connected the project. The agreement addressed intellectual property, liability and governance of the project, data and other assets within the project. A specific part of the agreement was that there was no requirement for the partners to bring the solution to market together. Each partner has a different exploitation plan; private sector partners may be looking for future business; public sector partners are interested in the project’s legacy and citizen benefits. Big corporates are looking for branding and at developing new business models.

A data sharing work package within the project covered technical, ethical and social issues, including privacy. It was led by BT, and involved representatives from several interested organisations.

Sensitive data was subject to bespoke governance arrangements. For example, the chronic obstructive pulmonary disease data from the health study was not shared more widely.

**Outcomes and lessons learned**

- The ability to connect many different sources of data and break silos results in use cases and solutions that are very different from what is achieved in individual silos. The technology has the potential to be applied to any sector.
- The project’s agile approach means that many aspects, such as the legal agreements and technologies, are being developed over the course of the project. SMEs have demonstrated innovation and creativity, and have introduced ideas that have enhanced the project.
- The project has changed the usual relationship between a public authority and private sector supplier from a customer and vendor relationship towards a more collaborative partnership with the participation of SMEs. It requires the sharing of risk and different ways of working. It is vital for ensuring that rapidly-evolving technology can be developed and applied in an optimal way.
Case study 3
OneTRANSPORT
Federated platforms for sharing suburban and rural data

OneTRANSPORT provides an ‘open marketplace’ for sharing data, enabling new data-driven intelligent mobility services to be discovered and created. The project is aimed at suburban and rural local authorities who may not have the resources – whether capital or operational revenue budgets, skills or knowledge - to create and operate their own data platforms, unlike major cities. Potential benefits include more efficient transport network utilisation, improved customer experiences and better infrastructure planning, for example. Cities can also benefit through a better understanding of flows of commuters and people delivering goods into and out of the city. Local authorities pay a subscription to join the marketplace, obtaining access to other local authorities’ data and services in return.

Summary: the eight dimensions of data sharing

The opportunity:
Data sharing via an ‘open marketplace’ by suburban and rural authorities, to enable the development of data-driven intelligent mobility services.

The data and its use:
Dynamic Internet of Things (IoT) data in real-time or regularly updated, plus static data sets. The data can be used by organisations to develop and deliver services.

The business model and value creation:
Local authorities pay to put data on the platform and in turn get dashboarding and visualisation services, data checking and access to others’ data. Third parties can also access the data to develop services.

The model for data sharing and the partnership:
Technology provides a common platform for local authorities and third parties to access data. It is a public and private sector partnership between local authorities, InterDigital (main technology partner) and others.

People with the right skills and expertise:
Interdigital developed the platform and the commercial model, as well as acting as data broker. Arup provided expertise about the transport sector and was involved in project management. Other organisations developed analytics. Local authorities were involved as end users.

Constraints on how data is shared and used:
GDPR and commercial sensitivity are constraints.

The data architectures and technologies:
A federated approach to data platforms, for platforms that share a common standard.

Governance / oversight / enabling trust:
A common data sharing agreement between local authorities and the platform, with 12-month licence agreements created for local authorities. At the end of 12 months, the data is portable and to an international standard.
Introduction

OneTRANSPORT was conceived in 2013 and developed with Innovate UK and private sector funding. It emerged from a recognition that better use of data for transport solutions was required, and for this, technical, commercial and legal frameworks would be needed to enable data exchange across many organisations and individual data platforms. The involvement of the first four local authorities was being funded by Innovate UK, and this part of the project ended in November 2017. Interdigital launched the platform as a commercial data marketplace in February 2018.

A consortium of 11 organisations was involved in the project, including InterDigital, Arup, Imperial College London, a number of SMEs, four local authorities and Highways England. Interdigital was the lead partner and developed the platform and the commercial model, as well as acting as data broker. Arup provided expertise about the transport sector and was involved in project management, developing the business case, evaluating direct and indirect benefits and project dissemination.

Four organisations carried out analytics including Trak, an SME that is transferring expertise in financial sector analytics, previously used to predict share prices but now being used to predict journey times. The Transport Data Initiative, a community of local authorities chaired by Buckinghamshire County Council, provided a forum for sharing lessons, for example about license agreements or data.

Three use cases were developed. The first two use cases enable better management of traffic flows during major events at Silverstone Circuit and Watford Football Club. The third use case gives Oxford City Council the ability to provide information on transport options for travellers coming into Oxford city centre, with the aim of increasing the use of bus services and the park and ride. Travellers are given the necessary information to make better informed choices, and there is also potential to influence modal share by offering incentives to customers to use different modes.

Business model

The platform allows the traveller or transport authority to buy outcomes and services rather than equipment, while the private sector develops the equipment.

Local authorities pay a subscription to put data on the platform, and in return receive dashboarding and visualisation services, a certain level of data checking, and the ability to see other local authorities’ data. The more local authorities that have joined, the lower the cost to an individual local authority and the greater the value as they are able to access a larger pool of data. The platform therefore relies on the development of an ecosystem of local authorities. However, it is a challenge to encourage local authorities to join before there is a critical mass of data, after which the true benefits will emerge. It is only once this critical mass occurs that the private sector analytics providers and app developers can create services that are useful.

The use cases are important in demonstrating the benefits to local authorities. They may not understand the benefits from opening and sharing data, and will also need to justify expenditure on the service. The local authority could be a provider or consumer of data. While some believe that all local government data should be free, if a local authority has invested resources in cleaning or enhancing the data then selling the data may allow them to demonstrate value for money.

The platform developer has a range of challenges: developing the business case, but also developing the platform itself, curating data and providing data services. The platform developer also has the role of a data broker. If a data provider puts a price on data, the broker will pass that through to the person prepared to pay for it, but without the broker paying for the data themselves. It is important that the broker does not interfere with the data so that they are not seen as a gatekeeper to the data. Furthermore, the platform should not provide too many services that might be perceived to be in competition with the private sector who want to access the platform.

The app developer creates services from the data. Their success depends on the number of ‘hits’ to their app, which may in turn depend on the geographical extent of the data. They therefore rely on lots of local authorities joining up and providing their data. Services such as data cleansing or aggregation can be outsourced to SMEs or larger firms. The aggregated data might be placed back on the platform as a value-added data source.

SMEs and others who are creating services need to pay to access data; this limits the number signing up and therefore the administrative burden. The project is considering organisations paying to access data for six months. A different pricing structure for SMEs and larger organisations is needed.

Technical and data curation arrangements

The project attempts to address the challenges of siloed and poor-quality data, creating an open, scalable system that counters the prevalence of proprietary systems in the past.

The platform is using oneM2M, an international standard for machine-to-machine communications that supports a federated approach to data platforms. For example, data from the OneTransport platform could be shared with data on platforms developed by other technology firms if they use the same standard.
Towards trusted data sharing: guidance and case studies

Case study 3

There is potential to bring together data from a diversity of data owners. In the transport sector, there may be hundreds of data owners in England, including local authorities, public transport authorities, logistics companies, airports, universities, private car parks and local authority car parks. In the future, there will also be data from connected vehicles. There is also potential to use data that has not traditionally been used in transport, and to use legacy data, although conversion to the necessary format would require expenditure.

For the first time, oneTRANSPORT has brought together national highway data, local authority traffic data, and supplemented it with temporary Internet of Things traffic sensors. It has demonstrated how that data from legacy traffic sensors and new IoT sensors can be integrated and visualised within the same platform.

The four organisations involved in carrying out analytics trialled different types of analytics on the same data, allowing them to test which is most appropriate for different applications. It showed that with one platform there could be many analytics firms developing proposals for services. Where the analytics work identified missing data, sensors were added to provide that data. The project helped to ascertain what datasets are required to obtain useful information from the data analytics.

Legal and commercial arrangements

The platform reduces the administrative costs associated with sharing data. The local authority has one data sharing agreement with the oneTRANSPORT platform, rather than numerous data agreements with app developers or analytics firms. Conversely, an app developer can access data from four local authorities, say, with one data agreement.

Licence agreements of 12 months are created for local authorities. At the end of 12 months, data is portable and to an international standard. In other cases, agreements may be longer. The intention is to avoid the situation where local authorities lose the rights to their data and have to buy it back.

The combination of a short-term license agreement and open standard gives rise to a risk for Interdigital but benefit for the project: a local authority could take back data and put it with another platform; conversely, these factors provide greater potential to attract others to the platform and build up a critical mass.

Data agreements were developed as part of the project. The local authority specifies the data that they want to be discoverable. There may be certain data that local authorities only want to share between themselves. Alternatively, there may be certain data that they allow everyone to see but at a price. All data is anonymised.

With regard to the data agreements, it is important that the data remains the property of the data owner or originator. However, if somebody uses the data to create different kinds of data and add value, that is then the new provider's data, and they can sell it back or give it away.

Cleansing is adding value, but whether the cleanser then owns the data needs to be clarified. If the cleanser has not been paid by the originator, then that could become their data as they have invested in it. If that organisation has paid for them to clean data and put it back on the platform, then it remains with the data owner.
Outcomes and lessons learned

- The project has shown how 11 partners have successfully worked together without a contract between them, instead using a collaboration agreement for the joint outcomes of the project.
- There is a lot more data than people believe. However, a lot of data is of low value. There is quite an expense involved in maintaining real-time data feeds.
- Local authorities need revenue budgets to keep the data flowing. The main lesson learned is that a sufficient amount of data is needed to carry out the various activities.
Case study 4

Industrial Data Space

Data sharing within a secure and trusted ecosystem

Industrial Data Space (IDS) enables companies to leverage the potential of their data within a secure and trusted ecosystem. It is based on a reference architecture model for data sharing, that supports the secure exchange and simple linking of data in business ecosystems using collaborative governance models. A key aspect is ‘data sovereignty’ – a data owner can share their data within the ecosystem, but can attach usage rights and usage restrictions at the time that the data is shared that can be remotely enforced. The model provides the basis for multiple uses by supply chains within and across sectors, including logistics, transportation and manufacturing. It creates the potential for organisations to develop business models based on data.

Summary: the eight dimensions of data sharing

The opportunity:
Industrial Data Space allows for secure exchange and linking of data by different organisations, while allowing organisations to retain sovereignty of data.

The data and its use:
Multiple sources of data and uses, including uses in logistics, manufacturing and health applications.

The business model and value creation:
It has multiple possible business models.

The model for data sharing and the partnership:
The project is a reference architecture model for controlled and secure data sharing. It is funded by the German Federal Ministry of Research and Education. The Industrial Data Spaces Association, a not-for-profit organisation, provides a forum for ensuring that requirements from various industrial sectors are met.

People with the right skills and expertise:
Twelve Fraunhofer institutes provide expertise in: data management, governance and valuation; discovering and linking data; security; and domain expertise. Industrial partners also provide domain expertise. Software partners provide expertise on how concepts might be implemented in practice.

Constraints on how data is shared and used:
GDPR and commercial sensitivity are constraints.

The data architectures and technologies:
An architecture that provides a blueprint for different implementations of industrial data space, encompassing business, security, data and service and software aspects.

Governance / oversight / enabling trust:
The project has bespoke governance depending on the application.
Introduction

The Industrial Data Space (IDS) project, which commenced in 2014, was prompted by the Germany’s Industrie 4.0 initiative. It builds on an earlier project that identified a key principle of new digital business models: specifically, that they depend on forming ecosystems around a comprehensive end-to-end customer process and require the sharing of data. The project also builds on research on distributed usage control – a fundamental principle of IDS - that has been developed over the last 10 to 15 years.

The German Federal Ministry of Research and Education has funded the project to bring forward the architecture and deploy it as a prototype in use cases. Initial use cases include ‘truck and cargo management in inbound logistics’, ‘development of medical and pharmaceutical products’, ‘collaborative production facility management’ and ‘end-to-end monitoring of goods during transportation’.

The Industrial Data Spaces Association, a not-for-profit organisation, provides a forum for ensuring that requirements from various industrial sectors are met and puts forward use cases. Increasingly, members of the association are software technology companies. It is hoped that they will take up the ideas so that there are multiple providers of IDS software. The association is open to members from abroad.

Twelve Fraunhofer institutes are involved, and the project is coordinated by the Fraunhofer Institute for Software and Systems Engineering in Dortmund. The 12 institutes bring together different fields of expertise; for example, Dortmund has a strong background in data management, governance and valuation. Other institutes have expertise in linked data and ensuring that data can be found, and domain expertise, such as production management, logistics and healthcare. Institutes with expertise in security have built the trusted connector.

A few ‘lighthouse companies’ (members that are well-known and already driving standardisation processes in their domain) are involved in driving forward use cases. The involvement of the software industry is key in reaching critical mass, as they will be implementing the concepts and technologies.

Companies have realised that they need to share more data than in the past, including data that was previously considered too sensitive to share. Companies are also recognising that data has value, so they want to be able to protect it. Data is only exchanged if it is requested from trustworthy certified partners. The data owner determines who is allowed to use the data and in what way. As a result, the partners of one supply chain may have joint access to certain data by mutual consent so that they can innovate, develop new business models, design their own processes more efficiently or initiate additional added value processes elsewhere, either alone or together.15, 16

Business model

One of the motivating factors for companies to join a data ecosystem is the knowledge that new business models will be characterised by interdisciplinary, cross-domain approaches. The project attempts to lay out the ‘ecosystem governance’, that will enable the ecosystem to emerge. Each participant has their own idea of a business model, but they are all relying on each other for the business model to work. The architecture of IDS provides organisations with the freedom to create this system of business models.

The core roles in the ecosystem include data owners, data providers and data consumers. Data owners may be interested in monetising data through IDS. Other trusted organisations may act as intermediaries within the ecosystem, taking on roles such as certification, brokering data, settling financial and data exchange transactions, verifying the identity of organisations wishing to access data, and data service providers. Added value is created by these roles as they promote trust throughout the IDS and provide metadata to the other participants. IT companies play a role here, providing software and services to the other participants of the IDS.

Technical and data curation arrangements

The reference architecture provides a blueprint for different implementations of the IDS, that can be customised by the participating organisations according to their individual requirements. It consists of the following architectures:

- The business architecture addresses questions regarding the economic value of data, the quality of data, applicable rights and duties, and data management processes.
- The security architecture addresses questions concerning secure execution of application software, secure transfer of data and prevention of data misuse.
- The data and service architecture specifies the functionality of the IDS, especially the functionality of the data services. These may include smart data services such as alerting or monitoring, or basic data services such as data aggregation or mapping.
- The software architecture specifies the software components required for pilot testing of the IDS. Existing technologies are being used as far as possible.

The use of existing domain-specific data standards in IDS is being explored. For example, a use case project in collaboration with SAP and Siemens will show how the
The concept of the administrative shell – developed under Industrie 4.0, and providing the semantic model for the items in factory or warehouse – works alongside the information model developed in IDS. Similarly, in the logistics use cases, heavy use is made of existing standards used for supply chain business processes. So again, the semantic information model is taken from that particular domain.

Usage constraints need to be attached to particular data items to enable data sharing in a ‘sovereign fashion’. A major gap is having a standardised or harmonised understanding of usage rights – the rules for data sharing. In the past, every company has come up with its own data models, and these need to be aligned to develop a common understanding of the rules. IDS is working with the automotive industry to develop this.

The approach to data management should be able to address uncertainties in data quality. There may be a combination of high-quality, structured data and data that is unstructured, volatile and fuzzy. There could potentially be value in bringing both types of data together.

IDS will be one of several channels for exchanging data; other channels might include open data sources that are freely available, or other data markets that make volumes of data available without restrictions, but without any guaranteed quality.

Data analytics is crucial to providing smart services and differentiating providers, so it is not offered by IDS. There is a question about where the boundary lies between the infrastructure and the data services that the infrastructure supports: in future, an increasing number of data services could become part of the infrastructure.

Legal and commercial arrangements

The IDS supports the governance arrangements by providing an infrastructure for data exchange, interoperability, and the use of new business models, while establishing trustworthy relationships between data owners, data providers and data consumers. It acts as a ‘trustee’ for mediation between participants, facilitating the negotiation of agreements and contracts. It offers a decentralised architecture that does not require a central authority.

Participants jointly decide on data governance such as data management processes as well as on applicable rights and duties. The overall architecture envisages services that make sure the data is packed and described according to the information model and usage restrictions attached, and then made available under the restrictions.

It is not the intention to help set legislative frameworks, but the project will follow the legislative process very closely. For example, at the European Commission level there are lots of discussions going on about the rights of data producers.
Outcomes and lessons learned

- IDS potentially ensures that the data owner is realising value from the data that they share with others. It helps to spread the benefits so that, for example, the costs of data creation, capture and management are not incurred by one party while another yields the benefit. The project embarks on the first steps of making a data economy happen, but there are a number of outstanding areas of research including:
  - monitoring data transactions, ensuring they can be billed
  - mechanisms for putting a cost on data in a repeatable or accepted way
  - technical challenges in distributed usage control, in particular enforcing usage rights in a remote environment
    - currently it is not possible to guarantee that data is not used elsewhere
  - accepted information models to describe data goods: semantic information models and domain-specific data models
  - identity management for the IoT ‘things’ that share data
  - management of data flows if multiple platforms are involved
  - the benefits and limitations of different types of platforms along the spectrum from centralised, such as data lakes, to decentralised, such as Hyperledger blockchains.
Case study 5

DAFNI

Data and Analytics Facility for National Infrastructure

DAFNI is a secure facility for assembling, hosting and creating datasets on infrastructure assets and networks, and the human and natural environments in which they are located. It will provide a shared data resource with professionally managed security and access arrangements, to ensure the appropriate level of assurance around commercial confidentiality and the protection of sensitive personal data. It aims to reduce the technical challenges and security risks associated with consuming data used in infrastructure research, simplify the mechanisms for accessing data, and increase the efficiency of data discovery, consumption and utilisation.

Summary: the eight dimensions of data sharing

The opportunity:
Creation of a national infrastructure database and repository of modelling tools to enhance capacity for researchers and practitioners to analyse performance and resilience of infrastructure systems.

The data and its use:
The project will gather infrastructure asset and network data from multiple sources across infrastructure sectors, socio-economic data, geospatial data, consumer behaviour characterisation data to be used in infrastructure modelling, simulation and visualisation.

The business model and value creation:
The project has received capital funding from UK Collaboratorium for Research on Infrastructure and Cities (UKCRIC). Operational funding is expected to come from research projects. UK Research and Innovation operational funding of research infrastructure, government departments and agencies with infrastructure responsibilities and business.

The model for data sharing and the partnership:
The technology provides a common platform for controlled access to data by infrastructure researchers, government and business.

People with the right skills and expertise:
The project is part of the UKCRIC, with academic partnerships led by the University of Oxford and industry involvement. The facility was developed by the Science and Technology Facilities Council.

Constraints on how data is shared and used:
Commercial sensitivity, national security and GDPR are all constraints.

The data architectures and technologies:
DAFNI is a platform that enables the management, validation and quality assurance of data. It provides a multi-model database architecture that allows storage of data types, including geospatial, columnar and network in their respective optimum storage formats.

Governance / oversight / enabling trust:
It is a shared data resource with professionally managed security and access arrangements. These ensure an appropriate level of assurance around commercial confidentiality and the protection of sensitive personal data.
Introduction

DAFNI is a major UK national facility to advance infrastructure system research. It is currently a capital project lasting four years, funded by the Engineering and Physical Sciences Research Council, delivered by the Science and Technology Facilities Council (STFC), and overseen by a governance board of 11 universities, chaired by the University of Oxford. It is both a hardware and software project, providing the platform for a national infrastructure database as well as the capability to carry out infrastructure modelling, simulation and visualisation.

Infrastructure systems modelling is used for a variety of purposes. For example, it may be used to examine the big picture: how infrastructure needs might change in the future, where investment is needed and the benefits. Other types of data such as sensor data might inform real-time modelling that is used to forecast faults in the system, where failures might occur, and to inform maintenance regimes. Modelling can also be used to examine the nature of cascading failures, either because of cyber-attacks or natural hazards, through network infrastructure risk analysis.

DAFNI will not itself develop methods of analysis, but will instead provide the platform to make the analysis more convenient, robust and accessible. The aim is to encourage others, such as government or academia, to supply their data and models to the facility. While utilities companies are unlikely to put their own operational and control models on the platform, it provides a facility for researching and developing operational applications.

The attraction is that the models sit above a properly curated national platform, with high-performance computing and cloud computing resources available. The challenge is to build momentum and achieve positive spillovers that result from data and models residing in a common place.

The software platform is being developed in tandem with a series of pilot projects that are bringing in academic models, building up the capabilities of the database, and demonstrating its use to others. These include projects on optimal maintenance and real-time operation and control. The database will be built up in a pragmatic way focusing first on the datasets required for pilot projects, while also brokering arrangements with data providers.

DAFNI builds on work carried out around NISMOD, a National Infrastructure Systems MODel developed over the last seven years, which made use of the NISMOD-DB++ database developed by Newcastle University. The database contains several types of data including: asset and network data; usage and demand data – including the factors driving demand; building data; geospatial socioeconomic data; and new sources of big data such as consumer behaviour characterisations.

Business model

Once the capital project has been developed, the project will need a sustainable business model to ensure that it has the resources needed for operation, maintenance and upgrade. In future, funding sources might include: continued research council funding and service-level funding: funding from government departments who have migrated models onto DAFNI; and businesses such as utilities companies.

Technical and data curation arrangements

DAFNI provides the processes necessary to ensure that security, provenance and traceability of data are addressed for both researchers and data providers. The database will allow the creation, management and publication of data assets, along with services that support data validation and quality assurance. One challenge is that the data required to understand current infrastructure and how infrastructure needs might change in the future is highly diverse - past work developing the NISMOD-DB++ database has addressed how to deal with this diversity. This is important to ensure that data formats are sufficiently consistent to allow interoperability and usability. The work needed to clean and reformat data to make it usable for modelling is highly labour-intensive.

Asset and network data includes data on power plants, road and rail networks, telecoms, and wastewater treatment facilities. Certain types of data are easier to access than others. For example, there are national databases containing information about power plants, their capacity and technologies, and National Grid makes data about electricity supply and transmission available online. Other types of data are more challenging to access if, for example, they exist in pdf format or they are commercially sensitive, such as telecoms data. Alternatively, different companies within a sector may hold useful information but in varying formats and degrees of accessibility. In the case of the water sector, this is a result of digitisation occurring after privatisation, and the absence of any initiative to develop a national picture of water assets. In some cases, the regulators hold useful information: for example, Ofcom manages to extract a large amount of information. Most geospatial data comes from the Ordnance Survey and contains the location of infrastructure assets, so resource is needed to add attributes about the data's assets.

Census data provides a good resource for obtaining demographic data that informs changes in demand. Geospatial socioeconomic data that form the basis of future scenarios is also useful, including population projections and scenarios around regional economic growth. New sources of big data that allow consumer behaviour characterisation exist, such as those held by the Consumer Data Research...
Towards trusted data sharing: guidance and case studies

Case study 5

Centre, include sales data, market research data and reward scheme data. Data on buildings is also important for determining the future demand for infrastructure.

Not all data will reside on DAFNI; instead, it will provide an index and a means of accessing the data. High volume datasets of secondary significance will reside elsewhere; for example, climatic data will remain on the existing JASMIN database at STFC Harwell. Increasingly, it does not matter where the data is physically stored unless there are strict security requirements. Ideally users should be able to build applications based on data without needing to be aware of its location or format.

Timeliness is a challenge, and there is a spectrum of capability in data suppliers. Some organisations have automatic updates and timed releases. Others have static datasets that were collected at a past point in time, and that are updated in an ad-hoc and undocumented way. In this case, it would be necessary to note that this has occurred.

Legal and commercial arrangements

There are several ways of commercialising the platform, which includes high-performance computational hardware, as well as database facilities. For example, technically advanced users could pay to use the platform, where they see the benefits of carrying out model development or analysis on the platform.

The platform could sell data products or pay-per-use analyses, requiring another tier of application to be built on top of the underlying data and analytical capability. Alternatively, a user could pay to develop a service on DAFNI and then take it elsewhere. It is intended that the platform will trigger interest from third parties.

Outcomes and lessons learned

- The database will be most successful if it holds a critical mass of data. One of the challenges is to encourage organisations that hold data that is in the public interest or for academic purposes to open it up. Even where national databases are nominally accessible, there may be restrictive licensing and pricing arrangements that preclude access to data.

- There is potential for the database to link to the work on developing digital twin pilot projects. A digital twin is a digital model of national infrastructure that will be able to both monitor infrastructure in real-time and simulate the impacts of possible events. 19

- A further challenge for DAFNI is raising awareness and getting people on board. The approach of carrying out rapid prototyping and agile software development, along with early pilots, is useful so that the benefits can be demonstrated early in the process. It is important to help people understand what the database is – their mental model may be very different from the reality.
Case study 6

APROCONE

Integrated product design in the aerospace industry

APROCONE (Advanced Product Concept Analysis Environment) is a project to develop improved methods of collaborative product design using novel information systems technologies. One part of the project focuses on methods for sharing data to enable improved integrated aircraft design. Through better data sharing, parallel design studies can be carried out by supply chain partners as partners do not need to wait for the latest information to be sent. This has the potential to speed up the design process as well as enabling better product design and cheaper products. Product data is shared via a central platform in a secure way. Companies can control and manage their intellectual property, which helps to promote trust in data sharing.

Summary: the eight dimensions of data sharing

The opportunity:
Sharing data in the aerospace supply chain to create the capability to design aircraft wings in a more innovative and time-efficient way.

The data and its use:
Aircraft and engine design data is used to improve integrated product design.

The business model and value creation:
The project is currently an R&D project with public and private sources of funding.

The model for data sharing and the partnership:
The technology provides a common platform for data sharing between project partners in a supply chain.

People with the right skills and expertise:
The project consortium is led by Airbus and includes Rolls-Royce and other supply chain partners, and academic partners. The partners have expertise in modelling and simulation, integrated product design, and process automation.

Constraints on how data is shared and used:
Constraints include commercial sensitivity and aviation regulations.

The data architectures and technologies:
A platform that allows the secure exchange and sharing of product data.

Governance / oversight / enabling trust:
Commercial frameworks are in development.
**Introduction**

APROCON is a £19.2 million project to develop new methods for aircraft design. It will create capability for UK companies to design aircraft in a more innovative and time-efficient way to meet future market and environmental needs. The development of such a capability puts the UK in a competitive position in future. The project is jointly funded by industry and government, the latter under the Aerospace Technology Institute. The project consortium is led by Airbus and includes Cranfield University, GKN Aerospace, MSC Software, Rolls-Royce, CFMS, and the University of Cambridge.

The project recognises that digital technologies such as the internet and distributed computing can be used in collaborative product development. One focus of the project is methods for sharing data to enabled improved integrated aircraft design. Other examples of collaborative product development include integrated systems and airframe design, for example, or ‘design for maintenance’, through better collaboration between aircraft manufacturers, supply chain and airlines.

The transfer of data has always been at the heart of collaborative design processes. In the distant past, it was generated by physical testing or hand calculations and transferred by letter. With computers adopted in design, they now generate data and transfer it by electronic means. The process has not changed, although the means to do so has. As an example, Airbus and Rolls-Royce will transfer data at certain pre-agreed points during the design process. Airbus might initiate the design and then request information from Rolls-Royce about the requirements for accommodating the engine. Airbus will hand the progressed design back to Rolls-Royce, which will use this information in its engine design studies.

With shared database technology, design and engineering processes can change. The project is investigating how to manage this information exchange from a technical viewpoint, by several shared-use cases that are realistic but not real, given the commercial sensitivity. The technology is already demonstrating transformation changes in these shared-use cases.

**Business model**

It is currently a R&D project with both public and private sources of funding.

**Technical and data curation arrangements**

The project uses SimManager, developed by MSC Software, and ShareAspace, developed by Eurostep. Both support the exchange and sharing of product data in a secure way.

Collaboration partners use their existing analysis tools, making use of a neutral data-sharing platform for which it is easy to add and remove systems and partners. Companies can control and manage their intellectual property, which helps to promote trust in data sharing.

Several areas are being developed as part of the project, including data standards, quality and provenance. The project investigates how best to define and record data provenance, and what data formats are required to create the right level of information that is appropriate to the process.

**Legal and commercial arrangements**

In integrated product design, data handover is typically carried out as a series of discrete transactions within a contract-based process. The different parties will transfer the minimum information required to enable each other to proceed with the design, so as not to give away what is perceived as sensitive commercial information. The different parties do not have detailed insight into other parties’ design studies.

The use of shared databases potentially transforms the commercial framework, since the sharing of information in this way is not compatible with current transaction-based processes. The project is investigating the impact on commercial frameworks required, alongside technical processes. For example, more information might be available to the various parties involved on the shared database, with conditions placed on who has access to the data and how the information is used. Technical and commercial aspects are increasingly coming together, such as the use of data tagging: conditions of use are tagged onto datasets and can be defined down to individual data points rather than at a more global level.

Better data sharing also potentially increases engineering capability further in the supply chain. For example, if a component is manufactured by a company lower down the supply chain, limited data is shared by the OEM, such as geometrical data, loading or stress analysis data. The OEM does not share information about how that data was obtained, or how the component forms part of the aircraft’s design. Potentially, the OEM could give suppliers more freedom to create the needed component, leading to better product design or cheaper products. However, there are commercial barriers to operating in this way.

Regulatory authorities, the Civil Aviation Authority and its international counterparts, tightly control the industry. The impact of current regulation on the new data-sharing environment is a research question that future studies will investigate.
Outcomes and lessons learned

- Engineering as a discipline is transforming, and APROCONE is helping to push this forwards.
- Engineering now depends on access to the information required to develop the product.
- Information management technologies will help to enable the transformation of engineering processes in the future.
Case study 7

Smart meters

Data sharing in the energy industry

Data is at the heart of creating a smarter and more flexible energy system. Data generated by smart meters can be used to inform the planning and operation of the grid, and to empower consumers to reduce and understand energy use. Third parties with permission to access an individual’s data can offer services, such as advice on switching to a different supplier or services that link smart meter data with other types of data. The smart meter communications infrastructure provides the means of accessing smart meter data. All energy suppliers, network operators and other organisations wishing to use this data become a party to the Smart Energy Code, which sets out their rights and obligations, including security and privacy obligations. Consumers can decide how their data is used.

Summary: the eight dimensions of data sharing

The opportunity:
Sharing smart meter and other sources of data with energy suppliers, network operators and third parties in a controlled way.

The data and its use:
Energy consumption data, and in future, data generated by other smart devices will be used to reduce energy use and create a smart grid.

The business model and value creation:
The business case is based on cost-benefit analysis for the various stakeholders, including consumers, energy suppliers and network operators.

The model for data sharing and the partnership:
Technology provides a communications infrastructure and mechanisms for accessing data in a controlled and secure way.

People with the right skills and expertise:
The project is a partnership between the energy industry, government, the energy regulator and the Data Communications Company (DCC). DCC works with service providers and is responsible for the delivery and operation of the smart meter infrastructure.

Constraints on how data is shared and used:
Constraints include GDPR, regulatory duties of energy suppliers and others and acceptability by consumers.

The data architectures and technologies:
Smart devices connected into a communications infrastructure to allow controlled access to data generated by devices.

Governance / oversight / enabling trust:
The Smart Energy Code (SEC) sets out the rights and obligations of those involved at an industry level. Security and privacy obligations exist for DCC users. Companies can access half-hour data only with the permission of the consumer.
**Introduction**

Existing electricity and gas meters in consumers’ homes are being replaced with smart metering equipment as part of a national roll-out. The roll-out is being led by energy suppliers, who are required by their licence to take all reasonable steps to provide smart meters to all of their domestic and small business customers by the end of 2020. Smart metering equipment typically comprises a smart electricity meter, a smart gas meter and a communications hub, alongside an in-home display that allows consumers to see what energy they are using and how much it costs.

As well as providing a link between the smart meter and in-home display, the communications hub provides a link to the telecommunications networks that allow information to be exchanged between smart meters and organisations that use the data. These include competing energy suppliers, energy network operators and other authorised parties, such as third party intermediaries that offer services around energy saving, switching suppliers or load shifting and demand-side response. The design of the communications infrastructure enables access to any piece of data by any party, subject to consumer authorisation, except for sensitive data such as security information. The national infrastructure has been built by Smart DCC Ltd, following a tendering process for the Smart Meter Communication Licence, which is regulated by Ofgem.

Data generated by smart meter readings benefits the various organisations in differing ways. Smart meter readings will automatically be sent to energy suppliers to allow them to generate accurate energy bills, and to inform settlements for the purposes of balancing national energy demand. Network operators use data about the energy system to investigate quality issues in relation to the network and determine outages at particular properties. Operational data helps them to manage networks better. Aggregators sell surplus energy back to the grid, or export energy from microgenerators such as solar panels.

Other parties might provide optimisation services, such as advising a customer on how they can manage demand against a tariff in order to lower costs, or whether to switch suppliers. Other third parties might provide bundled services to customers based on linking energy and other home services data. The data allows usage habits of customers to be determined, where permitted; these may vary according to whether a property is occupied throughout the day, for example.

Smart meters also support two-way communication and provide an interactive capability for energy suppliers, who can automate switching the supply on and off.

The ability for energy suppliers to access data at any time, rather than at certain times when manual readings are done, represents a large shift. Energy suppliers can use data from smart meters to record electricity use every half hour, which provides a more accurate picture of consumption and demand. This enables them to match their energy purchases more accurately and to bill customers based on the cost of generating and supplying the electricity they use. Currently, energy suppliers cannot buy energy in half-hour quantities, but in future it should be possible.

In future, in addition to smart meters, electric vehicles or other types of smart device such as smart appliances could be connected to the smart meter infrastructure, broadening the types of data generated and the uses to which data transmitted via the infrastructure can applied.

Future expansion of the infrastructure would enable others, such as water companies, to improve their metering. Smart metering equipment could potentially be used to collect property information, such as temperature or humidity measurements, to spot where there are health risks to vulnerable people.

**Business model**

The government’s impact assessments have identified the key monetised and non-monetised benefits for the main stakeholders. Consumers benefit from energy savings through enhanced energy efficiency behaviour. In future, consumers may be able to measure the electricity generated by microgeneration. Energy suppliers will make a range of operational cost savings by removing the need for site visits to take meter readings and improving business efficiency. Future business benefits include time-of-use tariffs. Network operators will be able to improve electricity outage management and resolve any network failures more efficiently once a critical mass of smart meters has been rolled out.

A major cost is the metering equipment including its installation, operation and maintenance, which will be borne by energy suppliers. Other upfront costs include supporting IT systems and the DCC’s operations. Industry participants will also need to upgrade their systems to integrate into the smart meter network. The government estimates that the expected net benefit across both domestic and non-domestic sectors will be £6.2 billion, although there are some uncertainties around the scale of benefits.

**Technical and data curation arrangements**

The main role of the DCC is to provide the infrastructure that enables communications between smart devices and DCC users. The DCC user sends a service request, and receives back a service response or alert. The service request might update the tariff held on a gas or electricity meter, display...
messages to a customer or read measurements of energy use at a specified frequency.

Data is passed from the device to the user in the form of an encrypted packet, to which only authorised users have access. The data is not stored, analysed or accessed by the DCC. There is a strict security framework since the infrastructure is part of critical national infrastructure.

Smart meters are to some extent limited in the data that they can send; for example, they are unable to stream high-frequency real-time data. They are more suited to sending smaller quantities of data.

DCC users need to create their own systems to deal with the storage and analysis of data. The advent of half-hourly measurements potentially greatly increases the volumes of data available. Energy suppliers have always had to deal with large volumes of data; for example, EDF Energy have been receiving half-hourly data for 100,000 customers who require above 100kWH. In the future, 27 million EDF customers will potentially be providing half-hourly data from smart meters. The necessary systems to support billing and tariffs need to be in place to support this volume of data. In addition, smart meters generate data through alerts that might identify tampering or outages. Analytics are applied to filter out insignificant alerts. Other safety related alerts, such as a gas leakage spotted through a gas meter, require quick identification and a rapid response.

Legal and commercial arrangements

All energy suppliers, network operators and other organisations wishing to access smart meter data become a party to the SEC, which sets out the rights and obligations of those involved at an industry level. The SEC is overseen by the a panel whose membership represents the interests of large and smaller suppliers, electricity and gas network organisations, other SEC parties and organisations representing citizens. The SEC makes sure the infrastructure is working well and ensures that any changes do not cause disruption. It also oversees security.

Certain organisations within the energy industry, such as domestic energy suppliers and network operators, are obliged through their licence to become DCC users. Others can apply to become DCC users if they would like to access smart meter data. For an energy supplier such as EDF Energy, data is used to fulfil regulatory and legislative duties such as balancing the energy system, generating bills or forecasting energy demand. EDF Energy has internal systems in place that help it to fulfil these duties around data, and monitor how it is used to ensure that the appropriate procedures are being followed.

For other uses such as producing a tariff plan or linking up with another energy service, the energy supplier must declare the purpose of the data.

Security and privacy obligations exist for DCC users, which require organisations to undertake security and privacy assessments that demonstrate compliance with obligations. Privacy assessments will be all the more vital with the arrival of GDPR; for example, there will be conditions on how long users can hold the data for.

Smart meters can store the amount of energy consumers have used in each 30-minute period, but by default, energy suppliers are only able to access daily meter readings unless the consumer has given permission for them to access to more detailed and frequent data. Consumers can decide whether energy suppliers share details about energy consumption with other organisations or use meter readings for sales and marketing purposes.
Case study 8
The Weather Company

Providing data services for sectors where weather has an impact

The Weather Company (TWC), an IBM business, is a data-centric company delivering insights to support decisions. It services many different sectors where weather has an impact, including transport, oil and gas exploration, energy and utilities, automotive, finance, agriculture and retail. For certain sectors, data services are critical to their profitability. Other customers are developing revenue streams out of data. TWC has a variety of businesses for which data monetisation is fundamental. These include both business-to-business (B2B) and business-to-consumer (B2C) companies. Various types of weather forecast are generated and supplied to many different users and weather data can be blended with other types of data to create insights for customers. TWC also provides bespoke services to help customers generate insight from their own data.

Summary: the eight dimensions of data sharing

The opportunity:
Using weather data combined with other sources of data to support decisions made by organisations. Diverse opportunities exist in many different sectors and applications.

The data and its use:
TWC collects weather data plus application-specific data, which has multiple uses.

The business model and value creation:
The company has various business models, including a B2C business where data is supplied free and revenue is from advertising, a B2B business where weather data and related insight are delivered to organisations and a bespoke business.

The model for data sharing and the partnership:
TWC has a common platform for all weather data and other types of data.

People with the right skills and expertise:
Data scientists, business development managers and lawyers need to work together closely. Typically, TWC provides platform and analytics expertise and meteorologists, while customers provide knowledge of their business.

Constraints on how data is shared and used:
Constraints include GDPR and commercial sensitivity.

The data architectures and technologies:
A shared ‘data lake’ with a catalogue for data product and analytics development with a common governance model.

Governance / oversight / enabling trust:
The company has a mature and flexible governance capability to deal with different business experiments and developing terms and conditions. Governance underpins all activities and is beneficial to the business, for example processes for checking quality and provenance.
Introduction

TWC has several different businesses. One well-established business uses a global data platform to deliver weather information to many smart devices through various free apps - these include Weather, Weather Underground and Storm Radar, as well as systems built into cars. The platform absorbs very large amounts of data from personal weather stations, satellites and national weather infrastructures. It cleanses and blends this data through a combination of weather prediction algorithms to produce the different types of weather forecasts consumed by different sectors. The free mobile apps such as Weather, Storm Radar and Weather Underground are a B2C business that is based around advertising. Growth is achieved by bringing more people to the apps and understanding context about their usage such as location and time of day, since both increase the advertising revenue.

In addition, there is a B2B business delivering weather data and related insight to organisations. In this business, TWC does not just sell raw weather data. It sells the output of blending additional data and pushing it through analytic algorithms to directly support its customers’ business decisions. This business spans many industries, including transport, oil and gas exploration, energy and utilities, automotive, finance, farming and retail. For example, in aviation, airlines use TWC services for planning flights - weather data allows airlines to better estimate how much fuel to load.

In agriculture, farmers can better understand when crops are stressed. In energy and utilities, weather data is used in outage prediction and recovery operations, in order to reduce the cost of storms and equipment failure.

Finally there is the more bespoke business where a single company asks for help generating insights on their own data. TWC provides a managed cloud service to host the development and operation of its data-driven services, combining the company’s data with IBM internal data, open data and possibly third-party data. There are different models of revenue generation in this business. It can be a one-time contract to help a company understand the value of its data and where to create value from it; TWC can augment data that a business partner continuously shares; or TWC can provide a data service to a business based on a blend of the company’s data, IBM’s data and IBM’s analytics algorithms.

Both the B2B and bespoke business do their development work in a shared data lake. Data services are delivered on the same platform as the B2C business. However, they still provide isolation for private client data, even in these shared environments.

Business model

In the B2C and B2B businesses, data monetisation is a fundamental part of the business, so there are dedicated managers developing data-oriented products. The bespoke business is more opportunistic and ideas come from many sources.

TWC works with both customers and business partners, and there is a mixture of data and value sharing between them. There are a variety of different business models, which depend on the customer, covering advertising, benchmarking, value exchanges and sale of insights. For sectors such as agriculture and aviation, the data services are critical for them to be profitable. Some customers are beginning to create additional revenue streams out of their data and thus a greater number of value-based pricing of the data services could emerge in the future.

For example, vehicle manufacturers are considering how they can monetise the data they gather from intelligent and autonomous cars to build additional revenue streams.

Data products go through a similar process to software products – there is an idea that is tested, development work done and piloted. Feedback from pilots leads to further development or in some cases the project may be stopped. The product is rolled out as a small pilot, after which the consumer base is grown. Data scientists, business development managers and lawyers need to work together closely. Typically, TWC provides the platform and analytics expertise and meteorologists, while customers provide knowledge of their business.

Technical and data curation arrangements

The platform is proprietary and has been built using open source software augmented with IBM governance products and internal components. It can accommodate real-time data and continuous feeds, such as radar data, which are necessary for weather prediction. Data about the habits of consumers comes through mobile devices and is used to support the advertising business. For most B2B business, data about weather, location and human data is blended to deliver specific value to each industry. Raw records, organised datasets or other forms of information are all used. All new data is catalogued, and GDPR is driving the cataloguing of historic data.

Data sets are curated in as automated a way as is possible. When a data scientist identifies a new source of data, a ‘pipeline process’ follows whereby data is catalogued, its provenance is captured and quality checks are run on the data. First, data enters a ‘quarantine zone’ where it is manually checked and then verified against Terms and
conditions (T&Cs). If it passes these checks, the data scientist and others at TWC can access it if they have been granted permission. When new data is derived, or external data is to be used in a product offering, there is a more complex approval process involving lawyers, business development and data scientists, particularly before commercialisation.

Where data is sourced from private weather stations, much of the data needs to be validated for quality by cross-correlating it with other sources – for example, the presence of trees or buildings may affect measurements. Checks are also carried out on data from customers to ensure that it is free of personal identifiable information.

Standards are used for commonly shared data, such as location. However, much commercial data is proprietary and is unique to the system that created it.

The data catalogue and associated processes are essential to the business model. The data platform has been engineered to support large volumes of data, be resilient and able to scale up when big weather events lead to high demand on weather services. The shared data lake for data product and analytics development provides a common platform for governance.

Security is ensured by having dedicated experts carrying out monitoring to detect data breaches and penetration testing. In addition, products cannot go into production unless they have the right security controls in place.

Legal and commercial arrangements

T&Cs on data sharing is a critical part of contract negotiation, for acquiring data and insights as well as selling it. The governance capability at TWC is mature and flexible to cope with different business experiments.

A two-way agreement is required, with obligations and permissions on both sides. The agreement covers the data lifecycle and quality. The quality of data must be appropriate for its use, and the agreement addresses the data that it is blended with and its timeliness. Insufficient data quality may cause a new venture to fail. Data checking is automated wherever beneficial. Sometimes the contract is an agreement to exchange data between parties and no money changes hands. Contracts have Service Level Agreements (both internal and external) and the company operates legally within those parameters.

Data governance underpins all activities and is beneficial to the business. Individuals work more effectively if they operate within the governance processes because they are designed to improve efficiency. Strong controls exist on new data acquisition, particularly if it forms part of a contract with a customer. More lightweight curation is necessary for data science experiments. However, as these move closer to commercialisation, checks and processes become more stringent. The data lifecycle is managed through the catalogue. GDPR will require significant amounts of new infrastructure.

Outcomes and lessons learned

• Often companies do not know the value of their data. Business leaders and decision-makers need to have a mature understanding of the value of data they hold; for example, there have been situations where a sales team is focused on selling hardware and physical services and gives away the data for free, even though the data (or any insight that can be derived from it) is the most valuable asset. It is important to treat data as both an asset and a liability, and an ethical approach and respecting T&Cs are fundamental aspects - even open data has T&Cs. It is challenging to establish ownership rights on derivative work and to understand the chain of custody. There needs to be lateral flow of data and the associated coordination, rather than a siloed and command and control organisational structure. For example, one part of business may need to incur the cost of collecting data to benefit another part of business; even inside a business T&Cs may be required.

• The value of the data depends upon its use and not the cost of collecting or processing it. A business makes a choice about whether to sell raw data, augmented data, derived data or a solution; the higher up in the pyramid, the higher the value, effort and risk. Data has a lifetime that can be short - this is particularly true in the advertising business, for example and this has to be factored into the pricing. It is challenging to ascertain what the value of data should be in advance, but a piloting stage can be used to validate an approach. TWC helps customers to explore the monetisation of their data.

• The barriers to participation are unique to each industry. For example, healthcare has its own unique regulations. Different industries have different views on the value of data and the conditions under which data should be collected and used. Several elements of the platform have helped to enhance trust, such as standardised T&Cs and quality measures. It is important to get a people-focused governance programme in place as soon as possible.

• “Always respect data – like the ocean, it can bite you … as well as serve you.”
Case study 9

Grampian Data Safe Haven (DaSH)

A facility for sharing sensitive health data

The Grampian DaSH allows the secure processing and linking of health data for the Scottish population when it is not practicable to obtain consent from individual patients. A virtual network allows researchers to remotely access but not download the data, which is held on different servers. Strict controls are placed on where data is stored, who can access it, the type of analysis applied to the data and the results that are extracted. The Chi number, a unique identifier that is present on certain datasets in Scotland, enables linkages between different datasets. Users of the data are primarily academic researchers, although industrial partners are increasingly interested in accessing the data.

Summary: the eight dimensions of data sharing

The opportunity:
Secure processing and linking of unconsented healthcare data for research purposes.

The data and its use:
Healthcare data, social data and other types of sensitive data will be used in mainly healthcare research, plus some non-healthcare applications.

The business model and value creation:
The project was initially funded by the University of Aberdeen and NHS Grampian CSO investment. It is not for profit. There is a standard access charge and hosting fee, as well as fees based on staff time to support a particular project.

The model for data sharing and the partnership:
DaSH technology provides a means for remote access to datasets in a highly controlled and secure way. Partners of the project include NHS Grampian, the University of Aberdeen, academic institutions and industry.

People with the right skills and expertise:
The project has a clinical lead, DaSH technical lead, research coordinators, quality assurance specialist, analysts and programmers.

Constraints on how data is shared and used:
Constraints include regulations on healthcare data, GDPR and ethical approvals.

The data architectures and technologies:
A virtual network that allows researchers remote access to data, with heavy controls on access and exporting data and results.

Governance / oversight / enabling trust:
Oversight is provided by the DaSH steering committee with a lay member, the Caldicott guardian and the privacy advisory committee. The usual ethical approvals are required by researchers.
Introduction

DaSH is a joint facility between NHS Grampian and the University of Aberdeen, in response to national guidance on improving the safe handling of linked data sets for research. In 2011, as part of the Scottish Health Informatics Programme, a blueprint was established for enabling researchers to access unconsented healthcare data. One safe haven was set up for each of the four major health research boards across Scotland to create a federated network that underpins a system of trust for sharing data. There are five safe havens across Scotland – one national safe haven and four regional ones.

There had always been an interaction between NHS and university researchers, so creating the safe haven environment was a natural development. It provides access to important healthcare datasets, such as the Scottish mortality records and hospital admissions data. The Chi number is a unique identifier that exists to ensure that patients can be correctly identified and is allocated on first registration with the system. It enables linkages between different datasets, since the numbers in different datasets can be matched. In addition, the University of Aberdeen and other universities have large cohort datasets. For example, the Aberdeen Children of the 1950s cohort study provides data about school tests scores and social background, as well as current data about the participants of the original studies. It is possible to carry out studies on the factors in early life that affect health later in life.

DaSH is a virtual network that allows researchers to access the data held on remote servers. Once logged in, researchers do not have a connection to the internet and cannot download or print the data or results of any data analyses. They have access to pre-agreed datasets and statistical packages that they use to run the analysis, including any bespoke code that researchers have written themselves that can be run within a statistical package. The data itself is pseudo-anonymised. Once the researcher is ready to publish their findings, they request the data that they want to extract from the safe haven. Patient-level data cannot leave the safe haven; instead, aggregated results are allowed with a minimum of 10 records. If researchers are interested in smaller groups, they need permission from the custodian of the dataset.

To date, the majority of researchers accessing the safe haven have been academic. However, a number of current bids involve industrial partners such as artificial intelligence companies that would like access to healthcare data to train algorithms that are used in artificial intelligence solutions for hospitals. In the future, there is potential to link to databases containing genetic information. Linkages have already been made to the University of Edinburgh’s Generation Scotland database that contains consented genetic data from human biological samples. DaSH is investigating how to best enable access to very large datasets to train algorithms. The model is also of value in non-healthcare applications. For example, DaSH has been used by the DVLA with MOT data, where there was concern that the location of MOT centres could potentially cause an individual’s address to be identified. DaSH is working in partnership with a similar facility in India with mutually trusted levels of security and governance, so that cohort data can be shared between the two facilities.

Business model

DaSH was mainly funded by the University of Aberdeen and NHS Grampian, with some funding from the Chief Scientist Office in addition. It is a cost recovery service, so costs are included in research grant applications wherever possible. There is a standard access charge, and a hosting fee is offered. Researchers are charged a fee that depends on the time spent by DaSH staff in guiding researchers, preparing data, carrying out linkages and the size of the datasets.

The facility helps to attract research funding to the university. If commercial companies are involved, the facility is required to follow the relevant university guidelines.

Legal and commercial arrangements

Terminal services are used to gain access into remote desktops. The networks of the various organisations involved are separated to ensure the data is secure. There are strict controls on where and how data is stored, and who can access it. The storage of the data is split between the NHS network and the University of Aberdeen network, with the sensitive identifiable information stored on the NHS network, and ‘payload data’ stored on university servers. The researcher accesses the university server. Data is pseudo-anonymised to make it possible to relink it. However, for each project, data is pseudo-anonymised in a different way, which means that an individual patient’s data cannot be linked between multiple different projects.

Only trained ‘approved’ analysts can access identifiable information. For highly sensitive data, one analyst sees only the identifiers and another sees only clinical and demographic data. In addition, linked datasets are stored on separate servers and access to linked data is restricted. No patient level data leaves the DaSH. Each project has its own Data Linkage Plan and a Data Management Plan.

While the facility is primarily for accessing unconsented linked data, consented or anonymised data can be accessed in a similar way. The facility has also been used in a situation where a data custodian has a large, anonymised dataset but prefers to retain some level of control and not give it directly to the researcher.
Towards trusted data sharing: guidance and case studies

Case study 9

There are some data inconsistencies, such as wrong CHI numbers or incorrect names, which makes fully accurate linking of data challenging. There are ways of checking whether the linkage is correct. Some clinical checking of the raw data is carried out, the extent of which depends on the dataset. Clinical staff will also check the analyses to ensure that the results are as expected.

**Technical and data curation arrangements**

The staff of DaSH act as facilitators, guiding researchers through the process of identifying the appropriate data to answer their research questions, putting them in touch with the custodians of the datasets they will use and advising on how to obtain the relevant approvals.

DaSH has a Caldicott guardian, an individual that provides oversight of the arrangements for the use and sharing of clinical information, and a privacy advisory committee. Individual researchers need to obtain ethics approvals for their particular projects, but any review will be proportionate and take into account the governance arrangements for DaSH.

Approvals may also be needed from the data custodian, academic institution and NHS R&D. In addition, DaSH has a steering committee that includes a lay member. If there are major changes to how DaSH is used, such as the involvement of industrial partners, the committee is consulted.

Researchers have to fulfil certain requirements in order to be able to access data: they have to be an approved researcher with the appropriate research experience and hold a valid Information Governance Training Certificate, which gives access for a limited length of time. Industrial organisations would need a researcher on their project team. DaSH staff monitor who has accessed the facility and keep audit logs.

**Outcomes and lessons learned**

- The close relationship between DaSH and the health board, partly a function of the small size of each, has contributed to the success of the facility. It was initially challenging to bring on board researchers to the new system of governance, when they had got used to a less rigorous system. However, the facility has enabled many students and other researchers access to data, and has allowed them geographical flexibility; for example, if a researcher relocates to a different country as part of their research programme, they are still able to access data.

- An expansion of the service is the hosting of a safe haven within a safe haven: it has been possible to bring a physical machine that allows researchers access to Administrative Data Research Network data, which they would otherwise have had to travel to Edinburgh to access.

- With hindsight, it may have been more useful to have a uniform approach to setting up the five safe havens; instead they have all been set up in slightly different ways. It may have also been preferable to aim for ISO27001 information security management accreditation from the start.
Case study 10

**MK Data Hub**

A data platform for the city of Milton Keynes

The MK Data Hub is the technical, data infrastructure of the MK:Smart project. This project recognises that smart city infrastructure creates benefits such as reduced congestion and resource use, and helps the city to be competitive. The data hub supports the collection, integration and use of large amounts of data from many sources relevant to city systems. It is a data platform for the city, enabling data from a diverse range of data providers to be accessed, and for developers to create new applications. The project has created a ‘data marketplace’ where open data can coexist with closed data, which may be private or commercially sensitive. Data providers remain in control of the redistribution policies (the data licences) and the access permissions they want to apply.

### Summary: the eight dimensions of data sharing

<table>
<thead>
<tr>
<th>The opportunity:</th>
<th>The data and its use:</th>
<th>The business model and value creation:</th>
<th>The model for data sharing and the partnership:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enabling access to data from diverse sources to create new city services and businesses.</td>
<td>IoT data plus other sources have multiple uses across a number of themes including energy, water and transport.</td>
<td>MK Data Hub is an R&amp;D project based on a public private partnership, with potential for third parties to create business models from the data. Next steps include scaling up the benefits and expanding use cases.</td>
<td>The technology facilitates access to data by partners and third parties. Partners include Milton Keynes Council, the Open University, BT and Catapults.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>People with the right skills and expertise:</th>
<th>Constraints on how data is shared and used:</th>
<th>The data architectures and technologies:</th>
<th>Governance / oversight / enabling trust:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Academic expertise in knowledge technologies sits alongside technology and domain expertise. Domain expertise is provided by the local authority and utilities partners.</td>
<td>Constraints include GDPR and commercial sensitivity.</td>
<td>Integrated databases that allow the searching of and controlled access to data.</td>
<td>MK Smart has a core steering group.</td>
</tr>
</tbody>
</table>
Introduction

Milton Keynes has always been a ‘city of the future’ – an exemplar of master planning and innovation. With the MK:Smart project, it has taken this concept to a different level in terms of scale and ambition, building on the existing innovation cluster of universities and companies that are located in the city. The city has recognised the strategic advantage of using leading-edge technology to improve quality of life and support economic growth. The intention is that the smart city will itself generate jobs, but also attract the people who could build on the smart city infrastructure, creating startups and further jobs in a virtuous loop.

The original project, which ended in June 2017, was led by the Open University and involved a consortium of 13 organisations, which included BT as the primary infrastructure partner, three universities, utilities companies and SMEs.24 A number of associate partners, including Samsung and Huawei, participated in the project, bringing in additional private funding to supplement the core funding provided by the Higher Education Funding Council for England (HEFCE). The project has created a business ecosystem comprising 150 organisations with different levels of involvement. The local community were involved in the project through an organisation called Community Action: MK, which facilitated the generation of ideas and feedback from the community – 13 of the ideas were financed. A number of outreach activities involved local school children and a massive open online course was developed that attracted 40,000 students worldwide.

Additional funding and the continued interest of industrial partners has allowed the work to continue beyond the end date.

BT has worked on developing an ecommerce model that allows data to be monetised. The Open University has worked on the means of controlling access to data. After the project ends, the existing data hub will be kept as an R&D facility managed by the Open University, which can be accessed by Open University researchers and others interested in experimenting with the technology. BT will develop these technologies separately in commercial settings.

MK Smart also includes three sector workstreams – energy, water and transport. Milton Keynes was built on a grid with the assumption that there would be mass car ownership. Mobility has been a challenge since not everyone has access to a car. As the city continues to grow, congestion is becoming more of an issue and traditional measures such as road widening and traffic lights need to be supplemented by smart measures to improve transport. City growth is also putting a strain on water resources, and smart technology could be used to moderate demand for water. The energy workstream is compatible with Milton Keynes aspirations to be a ‘green’ city, building on initiatives around the smart grid and infrastructure for electric vehicles.

Technical and data curation arrangements

The project team has worked with the council to make its data available through the data hub. There is a huge variety of data, both static databases and live streams, including: local and national open data; data streams from key infrastructure networks such as energy, transport and water; other relevant sensor networks such as weather and pollution data; and data crowdsourced from social media and mobile applications.

The data hub is a data repository, but with a key difference: a large effort has gone into making it easier for application developers to access the data. It is automatically integrated so that it is possible to search all databases once, rather than each of the databases individually. The data catalogue supports the discovery of the data, as well as the description and traceability of the relevant metadata, such as the sources and rights that apply to the data.

Not all of the data is open. For example, companies may be willing to put energy consumption data on the data hub but want to control access to the data. A mechanism has been created to precisely manage provisions and permissions. For example, it is possible specify who can have access to data and the conditions of use. A particular challenge is dealing with the relationships between the licences of multiple sets of data that are being used together. The Open University has explored the use of ‘computational licensing’ which provides a mechanism for formally characterising the licensing condition so that when two or more datasets are combined, the system can check the conditions are compatible and suggest a use.

Mechanisms for monetising data are been developed as part of the project, which will potentially bring benefits to the ‘owners’ of data. The intention is that pricing structures can be varied according to the type of user and quality of data; for example, the conditions may specify that research organisations can access data for free for research purposes, but there is a cost for commercial organisations. Data providers would also be able to charge based on the data accuracy, timeliness and granularity, for example. The hub would be able to use accountability mechanisms that allow contractual terms to be imposed, and data users to be compensated if the contractual terms are broken. Security measures are being developed to prevent hacking and spoofing of data.25

All data sources have been accepted, resulting in highly varied degrees of data quality. A minimal amount of data checking and cleaning has been carried out, and some standardisation of formats, to ensure data can be accessed through an application programming interface (API). This process also generates metadata about the provenance and licencing conditions. It is recognised that data quality must be adequate...
for obtaining meaningful analyses. The Open University has done some research on automatically analysing data quality by looking at different datasets at the same time to detect discrepancies. The issue of data quality is an outstanding challenge that requires further research effort.

**Legal and commercial arrangements**

The project is HEFCE funded. As it is not possible to run commercial services on the JANET network, work has been carried out on an R&D basis for the last 3.5 years. If an SME wants to work with the project, they can be supported in developing ideas and solutions but if they want to launch commercially they need to find a solution other than the data hub.

Once the hub is running as a R&D facility, it will not have the same level of service level agreement as such a facility would have in commercial setting.

**Outcomes and lessons learned**

- While MK:Smart has focused on prototyping solutions, the next phase is about scaling up. There is still great potential to develop intelligent mobility solutions from real-time data. It will be important for the project to identify new business models, as well as developing the technical aspects of the project.

- The first large-scale commercial development emerging from the project has been developed by Vivacity, a spin-off from MK:Smart, that has attracted funding from Innovate UK and private investors. It is a smart parking scheme in Milton Keynes that makes use of intelligent camera technology and allows drivers to book a parking space via an app.

- The ecommerce element of the smart city is technologically possible and has been implemented and tested. However, the concept of a data marketplace still needs to be fully proven. It is likely that this will be more successful in a particular context, such as a smart city data hub or industrial data hub.
Tow ards trusted data sharing: guidance and case studies

Case studies

References


2. Containerisation is a method for deploying and running distributed applications that uses multiple isolated systems rather than launching an entire virtual machine for each application. A virtual machine (VM) is an operating system (OS) or application environment that is installed on software, which imitates dedicated hardware.


5. CitizenMe www.citizenme.com/


7. A ‘whitelist’ could, for example, be a list of permitted IP addresses.

8. Ingenia Article - Issue 75, June 2018, Should we trust connected devices? Paul Taylor FREng


10. GM-Connect: Greater Manchester’s Strategic Data Sharing Initiative

11. Triangulum is one of currently nine European Smart Cities and Communities Lighthouse Projects involving Manchester, Einhoven and Stavanger. In Manchester, the focus is on renovating historical buildings and providing an autonomous energy grid that supplies heat and electricity for the student quarter in the city. http://www.triangulum-project.eu/

12. Bristol is Open www.bristolisopen.com/

13. Future City Glasgow futurecity.glasgow.gov.uk/

14. Smart Service Welt and Acatech, March 2016. Smart Service Welt - recommendations for the Strategic Initiative Web-based Services for Businesses


18. DAFNI: Bridging the gap between academic research, data analysis and strategic infrastructure planning, www.dafni.ac.uk


20. Eurostep provides innovative software and solutions for secure PLM collaboration with ShareAspace. www.eurostep.com


24. www.mksmart.org/data/


26. vivacitylabs.com/