

Knowledge & opinion > Reports >

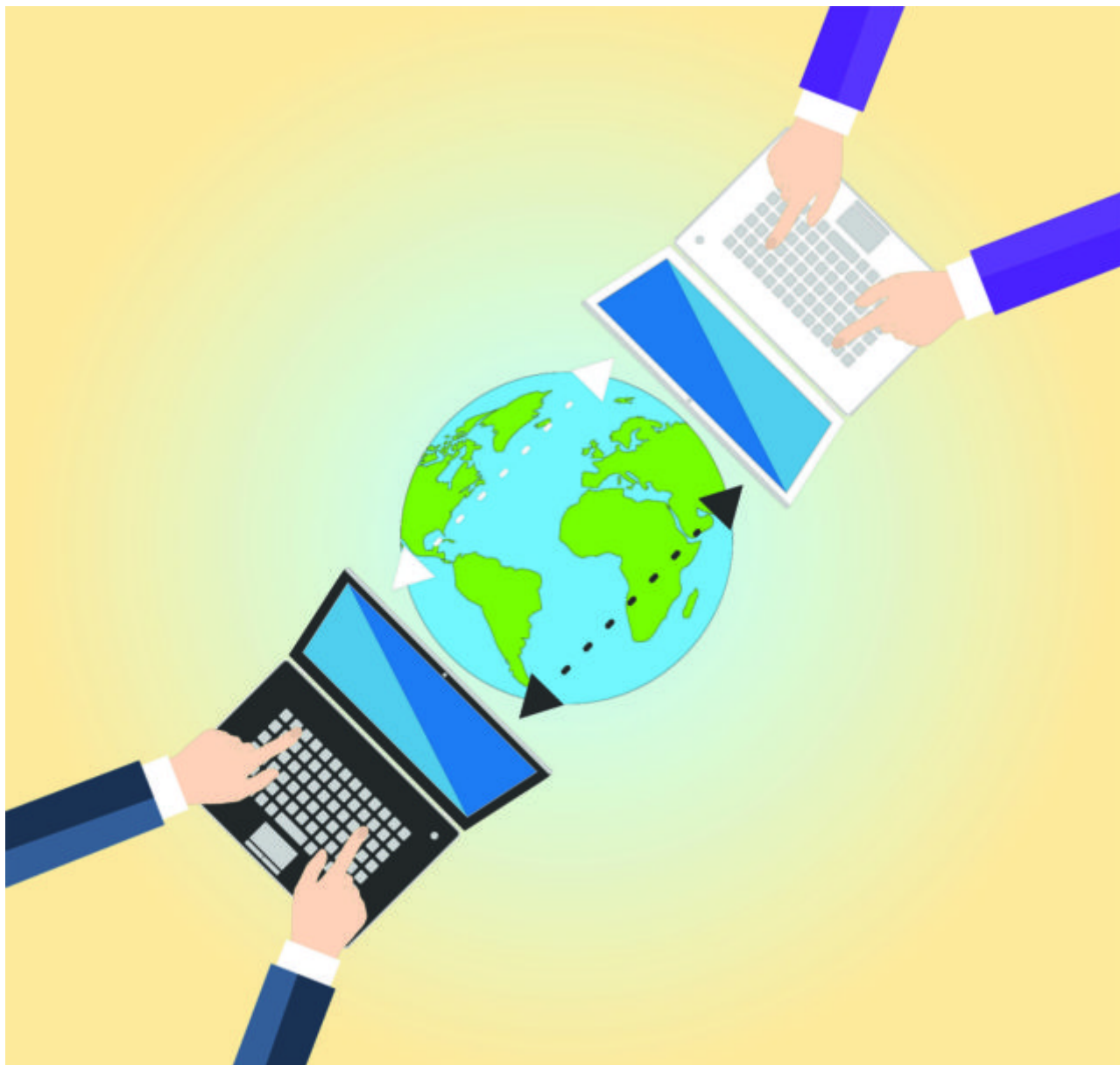


Image: Copyright Adobe Stock

What organisations need in order to share more data: our research

Fri Oct 26, 2018

Share

Data infrastructure **Data publishing and use** **Innovation**

We speak to businesses about their challenges in data sharing and how they can be supported

*As part of an R&D project exploring how to increase access to data while retaining trust, ODI User Researcher **Lisa Perez** spoke to businesses to discover their key challenges in data sharing, and explore how they can be better supported. Here are her findings*

In June and July 2018, the ODI interviewed people who work for organisations which share or publish data, or who have expert knowledge of data sharing practices in specific industries.

The interviews were intended to answer the following questions:

- What challenges do organisations face when sharing data?
- What models are currently in use?
- What new models are being proposed?
- What sectors should we investigate further?

The outcomes of the research will help us to identify sectors and problems which the ODI could work on for maximum impact. This report summarises the key findings from this research stage.

Nine one-hour interviews were conducted over Google hangouts. Interviewees represented the following industry sectors:

- Fintech (1)
- Extractives (2, representing one organisation)
- Transport (2, including one with specific focus on autonomous vehicles)
- Healthcare (2)
- Environment (2, from the Environment Agency)

Participants were recruited primarily through known contacts, direct requests, and social media. The research is exploratory, not exhaustive, and indicates issues and solutions experienced by individuals working in the selected sectors.

The five sectors varied considerably in the degree to which data was shared and in the willingness to share more. For example, the healthcare sector is people-focussed with significant privacy concerns about highly-personal data. Conversely, the transport industry has many demonstrated success stories and a general recognition of the value of opening data.

Despite differences, the sectors shared some similar challenges suggesting the possibility to learn from emerging models and solutions in other sectors. Part I of this report describes cross-sector themes, challenges, and emerging sharing models, while Part II highlights sector-specific issues.

Part I: Overall findings

Challenges

Data discovery

“

There is hardly any structured data available in a, kind of, wider scale or scope

”

Interviewees experienced challenges in finding relevant data, particularly when only limited data is shared or made public and when data is published in unstructured and inconsistent formats in disperse locations. This was a problem in the extractives and healthcare sectors.

Data control and access

“

Who's going to get the value?

”

The need to resolve issues of data control and access among multiple competing stakeholders was seen as a barrier to sharing. In the automotive industry, it's not clear who should control data generated by modern automobiles. Access to the data could be claimed by drivers, passengers, car owners, car manufacturers, software or hardware providers, or the wider transport network. Similarly in the extractives industry, there are differing views on who controls or has access to data about deposits and reserves – the government, the extraction company, the public, or international bodies. The issue becomes further complicated when data includes a personal component. What data should an individual control or own?

Resolving these issues may require an unaccustomed degree of collaboration and sharing among otherwise competing parties and a shift of organisational or industry culture may be required before any technical solution can be effective. Slow progress in meaningful sharing of healthcare data is an example of halting progress. Similarly, the transport sector may need higher degrees of cooperation and data sharing to join separate networks and enable increased transport usage and industry growth.

Managing data as infrastructure will require resolving multiple overlapping data rights each with their own motivations and challenges. Read more in a paper the ODI CEO, [Jeni Tennison recently wrote for the UK Centre for Data Ethics.](#)

Trust

“

There's significant Problems on the PSD2 regulation when it comes to trust, because of the way the payment services regulation and payment service directive was written

”

Trust in data quality is dependent, in part, on the credibility of its source. This is particularly important in extractives where trust is low throughout the sector. Both extractive companies and some governments are assumed to be acting in their own interests rather than the public interest. Companies aren't trusted to released unbiased data and data released by governments who aren't see to be generally credible will not be trusted. However, this varies by sector. In the Autonomous Vehicles sector, data from local authorities may be seen as more trustworthy than data from a private company.

One way Transport for London gained consumer trust for the Oyster card system was by solving customer problems quickly and issuing refunds instantly where needed, thus increasing customer willingness to allow TfL access to individual journey and payment data.

Public trust can be damaged by data breaches, so successful sharing models must include robust data security technology and governance. One interviewee felt that the technical standards for regulated parties in the payment services directive (PSD2) were currently too weak, making a data breach more likely.

Consumers will also lose trust where they feel they have been misled. The NHS experience with care.data showed the consequences of failing to earn the trust of stakeholders by rushing to share data without robust data governance processes in place. The Caldicott enquiry that followed highlighted the importance of rigorous data governance policies in establishing trust. While this example is from the healthcare sector, it could apply to other sectors dealing with highly sensitive data.

Transparency

“

People own the resources, they should know about what their governments do ... so that they can hold both companies and their governments accountable

”

Transparency is often opposed by those who feel they or their organisation will be

seen negatively if data is released. This issue crosses sectors, from government resource departments unused to detailed public scrutiny, to NHS GPs concerned about the use of league tables. Within any environment however, there will be natural allies of data sharing. Revenue authorities may be more interested in opening details of extractive activities than the departments working more closely with extractive companies. When discussions about healthcare data sharing have included both patients and GPs, GPs have been more willing to share data when they directly see patient support.

Business cases

“

We're destroying traditional business models and creating new ones at the same time and that is not straightforward or easy

”

Many organisations lack knowledge of business models that would support greater degrees of data sharing. One interviewee felt that in the rapidly-evolving transport sector there aren't enough successful examples to overcome the reservations of risk-averse managers. In the historically secretive extractive industry, organisations typically only see downsides of sharing data that they have always viewed as commercially sensitive. When extractive organisations do share data, the terms of the sharing agreement are rarely made public so the industry lacks good examples to learn from. Furthermore, examples can take time to develop whereas the downsides of data sharing, e.g. resources required, are usually more immediately apparent.

Personal data

“

If you're doing something on population-level, consent just would not be feasible

”

Many organisations and sectors struggle with how to make effective legal use of personally-identifiable data. The lack of test cases of GDPR has led to organisations being very cautious. People who have historically dealt with confidential data may be reluctant to share that data even when there is a clear legal basis. One interviewee remarked that data protection is often used as an excuse not to share data.

Anonymisation in particular is both a technical and a cultural issue. Data publishers and decision makers may not understand the anonymisation spectrum or feel

sufficiently informed to design or make decisions about data aggregation or anonymisation processes. Interviewees from the Environment and Healthcare sectors raised this issue, but it could affect any sector which deals with personal data.

Possible solutions

Effective data sharing requires the motivation and willingness to do so as well as the technical and management skills to implement it effectively.

Changing market conditions will be one impetus for data sharing. Examples from ATMs and mobile phone networks in the past have shown how organisations shared data in order to make these systems work for customers and grow the market. Legislation and government policy that requires or enables data sharing will also have an effect, though some feel that legislation works too slowly to take advantage of technical innovation.

Successful examples of data sharing can demonstrate benefits to market players and influence sharing throughout the sector. However if details of data sharing agreements are themselves not made public, their impact will be muted. Hence a trend for extractive companies to publish contract terms could be an important step to greater openness in that sector.

Long-term stakeholder engagement to enable culture change was seen by healthcare sector interviewees as essential to increased data sharing in healthcare. Lessons learned in this sector could prove useful to other sectors. Discussions with GPs and patients about data sharing and with clinicians and researchers about data collection, management, and use were suggested. Accountability and data governance needs to be developed together with stakeholders to ensure it works for all.

Data sharing models

Current examples

- The UK Environment Agency now considers data sharing whenever a new dataset is created. They use a checklist to help managers determine which data can be shared with an open license and which needs a more restrictive license.
- Oil companies sometimes share data about their resources in order to attract or reassure investors
- Both transport companies (such as Uber) and extractive companies may share some data publicly or with regulators in hopes of pre-empting increased regulation.
- The Open Banking regulatory framework allows existing financial institutions and new organisations to share financial data for the benefit of consumers
- The NHS does not have centralised data governance, it allows individual trusts to make independent agreements.

New models

New models are being suggested and tried within the transport and autonomous vehicles sector. If market players do not cooperate, a model of competing ecosystems led by a few major players may develop.

The importance of high quality training data for autonomous vehicles shows the potential for public bodies to license CCTV and other relevant data to industry entrants.

Synthetic data sets of edge cases used for training autonomous vehicles are a candidate for sharing amongst developers as it is likely to increase car safety, leading to the market acceptance which will be necessary for industry growth.

Part II: Summary by sector

Healthcare

“

Information governance is 9/10ths cultural

”

Challenges:

- Data quality is not always adequate for research due to varied formats and schemas
- Confidentiality and trust are significant issues and hindered by cases like care.data where communication and acceptance was poor
- There is a lack of consensus on data sharing models – whether consent-based, anonymised, aggregated, or open but managed.

Possible solutions:

- Engagement and discussions with patients, clinicians, researches, and the general public
- Electronic healthcare records designed with clinicians, patients, and researchers in mind.

Successes:

The Moorfields agreement with Deepmind publicly shows the benefits of sharing anonymised medical data

Impact:

- Resource efficiency
- Treatment effectiveness

- Development and testing of new treatments.

UK Environment Agency (EA)

“

Share what you can to the extent you can

”

Challenges:

- Third-party licenses (e.g. Met Office, OS) limit data that can be made open
- Few dedicated internal resources to help managers make decisions about data publishing
- Reluctance to find solutions for sharing data with personally identifiable content.

Possible solutions:

- Published examples of people and organisations using EA data help to embed and strengthen the data sharing culture and motivate staff to open more data
- Clarity on how to share data containing personally sensitive information, and GDPR test cases will enable more data to be confidently and safely shared.

Successes:

- The EA's open data risk management checklist helps managers decide if they can release data under an Open Government License or if a more restrictive license is necessary
- The Shoothill Gauge Map is an example of open data leading to the development of good quality third-party products
- Data sharing is now considered whenever a new dataset is created.

Impact:

- Increased trust in Environment Agency activities due to transparency
- Increased public engagement and support
- Reduction in resources needed to process Freedom of Information requests as data is published proactively.

Fintech

“

I'm concerned you'll find out that we've created a really nice fine-grained flexible consent model which empowers the individuals and the individuals just won't care

”

Challenges:

- End-user identity management is currently cumbersome. Users reduce cognitive load by not engaging with privacy statements and terms, and then lose trust when something goes wrong
- PSD2 doesn't currently allow regulated organisations to display their consumer brands, resulting in consumer confusion and lack of trust.

Possible solutions:

- A centralised identity system like those in place in Belgium and Scandinavia (BankID), may enable more informed customer behaviour as the authentication process becomes less of a chore
- PSD2 API extensions for finer control or what data is shared will enable users to allow access to more services without significantly increasing their exposure to data breaches
- A regulator may be needed to manage the friction in handovers between financial institutions and other regulated service providers.

Successes:

Detailed regulations and a clear liability model are already in place.

Impact:

Greater competition and innovation leading to better products and services for consumers

Extractives

“

Lots of decision makers simply don't trust information in public domain because, the industries themselves retain this culture of secrecy overall

”

Challenges:

- The industry has a long history of secrecy both within companies and their government counterparts
- Some governments lack the skills and resources to manage and release the data that they have access to
- Extractive companies are assumed to only release data in their own self-interest, and data that is released is rarely granular. Larger companies typically only need to release a small fraction of information they have in order to

comply with regulation

- A global portfolio approach to risk means that windfalls in one location subsidise failures elsewhere in the world, ensuring that the interests of extractive companies and the localities they operate in are rarely aligned.

Towards a solution:

- Regulation and enforcement
- Help and training for government resource regulators
- More awareness and use of what is already in the public domain
- Support of the trend toward open contract terms
- Initiatives to improve the findability of that is already in the public domain.

Successes:

There is some progress on open contracts

Impact:

- Open data can help governments better manage their assets and address climate change
- Citizens can fairly evaluate government management of natural resources

Transport & Autonomous Vehicles (AV)

“

More data, more choice and a more effective alternative to just driving yourself everywhere

”

Challenges:

- Data sharing is seen as a commercial and legal risk
- Lack of established business models. A decision to release data may feel like a “leap of faith”
- Transport users outside London are believed to be resistant to sharing payment and location data with transport companies.

Possible solutions:

- Clear business cases and examples, particularly in the private sector
- Clear benefits to transport users of sharing payment, journey, and location data
- Policy leadership
- Sharing of AV AI edge case data to improve safety while minimising commercial risk
- Services that turn CCTV data into data which is more abstract and less personally identifiable while still being useful for AI system training
- A data commons stewarded by a non-profit organisation to ensure easy

- A data commons stewarded by a non-profit organisation to ensure easy access for all.

Successes:

- Cashpoint and mobile phone networks demonstrate how data sharing can enable market growth
- Effective sharing within the London transport ecosystem
- Uber releasing data for planning purposes.

Impact of increased data sharing:

- Joined up public transport leads to fewer car journeys and the resulting environmental and health benefits
- Safe autonomous vehicles
- More effective transit options for rural areas whether through more efficient bus services or shared and/or autonomous vehicles.

Get involved

If you'd like to find out more about our research in this area, or would like to work with us, please contact research@theodi.org

Related

Recent related activities

BLOG, NEWS 11/10/2022
