4.4 Open Data

Momentum around open data is constrained by the privatisation of public data and increased security concerns. This limits the potential of data to benefit the whole of society.

Context

Open data rests on the principle that a wide range of often publicly funded information should be made freely available for anyone to use at no charge. Its popularity is based on the assumption that, as long as the correct safeguards are in place, it can make governments more transparent, accountable, and efficient, while allowing businesses to use the data to create innovative and helpful products and services.[49]

There are various different types of open data:

- Data made available by governments and other institutions for purposes of transparency;

- Data made available by any organisation to enable innovation, often by private companies to create new paid-for services; open banking with far-reaching legislation such as PSD2 is a good example of this;

- Data intended to empower citizens and other communities to be community aware and self-managed.

A host of international bodies, including the World Bank,[50] OECD,[51] the EU,[52] and numerous UN agencies,[53] all support the Open Data movement. To reflect this, the Open Data Barometer, the Open Data Inventory, and the Global Open Data Index are all seeking to highlight which countries and governments are most open.[54],[55],[56]

Level of Workshop Debate

- High
- Medium
- Low

Opening up vast public digital estates - from maps to chemical compounds – is driving a plethora of innovation – many with positive social and economic effects – think of the likes of CityMapper and OpenStreetMap, which help people plan their routes by integrating data for all urban modes of transport.

It is also contributing to the economy. The European Commission estimates the market value of open data will be around €285bn by 2020. Companies are now joining Governments and public bodies in making data sets available for open use, many as part of 'data for good' initiatives.[57]

However, it's not all plain sailing. In some locations, awareness of the potential of open data remains low, and as was noted in our Ivory Coast workshop, increasing this awareness was seen as *"a pre-requisite to more open sharing."* On the other hand, there are times when open data's potential has been exaggerated, and some assumptions relating to open data are wrong or misleading. For example:

- Making data open doesn't automatically yield benefits;
- Not all information can or should be made accessible;
- Not every stakeholder is able to make use of open data. Although its publication is intended to provide wider access, the reality is that the number of actors that can truly make use of it is small; they require infrastructure, highly technical skills, access to technical assets and capital. Because of this, often these are established institutional and corporate actors, not members of the public;[58]

- Open data does not automatically result in open government.[59] As the Web Foundation observes, "the community continues to struggle to demonstrate the positive impact of open data on good government."[60]

A number of studies suggest that less than a third of the data that is being made available is actually being used.[61] There are many reasons for this, not least a lack of data-handling skills among officials, activists, and journalists. Also, to be truly effective, open data needs to be accessible and of high quality, not just high quantity.[62] However, many data sets that have been published were built for administrative purposes, and are not structured in a form that can be easily sorted, analysed, and matched with other data. As yet, there is no shared definition of what constitutes 'good quality' open data,[63] even though many are hugely optimistic about its potential - McKinsey research suggests that better quality open data could help unlock an annual $3.2tn-$5.4tn in economic value globally.[64]

"As long as there is access to viable data, much can be achieved. It is increasingly recognised as an essential part of transparent and effective government."

Abidjan workshop

What We Heard

In our discussions, there was widespread support for open data. In Europe and North America, open data was highly ranked as a key issue for the future. Elsewhere, across Asia and Africa, it was also embraced. In Abidjan, for example, the view was that *"as long as there is access to viable data, much can be achieved. It is increasingly recognised as an essential part of transparent and effective government."* However, many also agree with a view in Bangkok that *"the public sector does not understand the benefits that can flow from this."*

Hurdles and constraints were also recognised. Workshop participants considered that some open data sets are not kept up to date. One Bangkok participant observed that, although there was access to government data, *"it is of poor quality and there is no clarity on how it might be used to drive positive impact."* There are also questions about who should cover the costs of making open data complete, consistent, accurate, and appropriate. San Francisco asked, *"who will pay to clean data?"* And while some see this as a government responsibility, others suggested that those who use it should pay a fee to help cover these costs.[65]
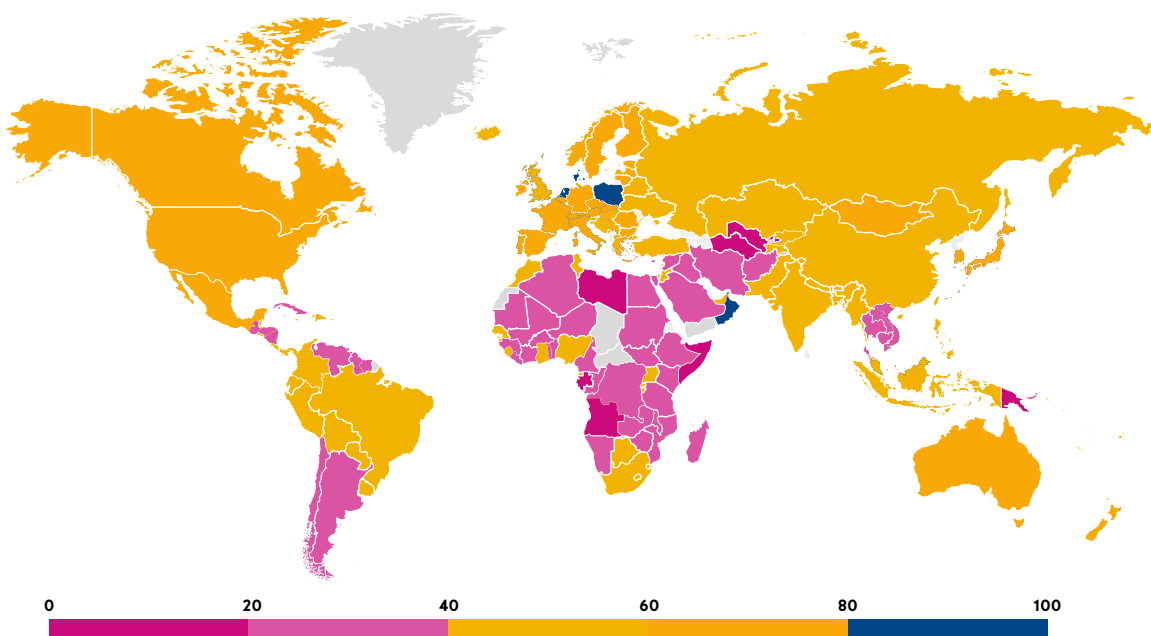
Which Nations are Most Open: The Open Data Investors (2018/9)

A bigger, more heated debate is growing around the 'privatisation' of open data. We heard unequivocal views on how open data is being compromised by aggressive intellectual property stances in some locations

Four key issues that were highlighted during our discussions:

• **Copyright:** As was highlighted in Toronto, some government bodies, including the UK's Ordinance Survey and Canada Post, have spent many years building up expertise and insight, and are exerting copyright over key data sets. As the generation of this data was originally publicly funded, many see that this ring-fencing is against the national interest. Others see it as a legitimate protection of prior investments.

• **Licensing:** As commercially valuable data is aggregated into 'derived data', and new forms of value are being identified, there is a lack of clarity on how (or if) that value should be shared, for example, through licensing new copyright and patents. Mapping apps such as Waze depend on open data, but their business model, which is based on hyper-localised targeted advertising, collects and monetises personal information.[66] In Toronto, it was felt that *"this is a clear conflict between claimed ambition and business model reality."* Another example is private companies repackaging and reselling public railway train timetable data.

• **Privatisation of public information:** New commercial sources of value are being created from public, academic, and government information, and are then being used for private enterprise. In Singapore, discussions cited *"Uber's 'wholesale privatisation' of Carnegie Mellon's autonomous vehicle expertise,"* through the recruitment of many leading academics along with their know-how.[67] Monsanto tried to

patent nature's plants a decade ago, and there have been a host of more recent activities by the likes of Facebook, Microsoft, and Amazon.[68] Tactics include attracting university professors with up to 10 times their academic salaries, extensive computing resources, and the promise of limited bureaucracy.[69] Moving forward, if more public information is made open, there is a concern that private companies will increasingly exploit this opportunity via intellectual property mechanisms.

This is not a new concern. It was raised as far back as fifteen years ago, when information published from the publicly funded Human Genome Project was "privatised" by companies like Incyte Genomics, that by 2005, had patented 2,000 human genes.[70] Several believe that, in a world where online authorship is increasingly multi-layered and collaborative, and where patents are protecting digital business models as much as technology, the original intent of intellectual property regulation is not working. Open data sets, they argue, should not be patentable, nor should they be subject to other forms of intellectual property, such as copyright.

"We want the bowl of candy out in the open, but we don't want people to steal from it."

Copenhagen workshop

- **Privatisation of government bodies:** Lastly, there is also evidence that some governments are "handing over" public assets, including associated intellectual property and public data, that should remain open to private firms. The potential privatisation of government bodies, such as the Land Registry in the UK and air traffic control in the US, are two current test cases.[71] Commentators believe that there may be many more in the pipeline globally, especially in the fields of environmental and resource information.

There are, however, legal questions about how to share anonymised data from governments and companies in a safe, ethical way, against a backdrop of public mistrust. Some felt that open data advocates might have been too naive in their activities - the scandal around Cambridge Analytica made this clear. As a workshop in Denmark commented,[72] *"we want the bowl of candy out in the open, but we don't want people to steal from it."* It has certainly been a learning process. Data trusts, separate legal entities designed to help organisations extract value from anonymised data, are one way of limiting the risks and allaying concerns about how sensitive data is held by third parties. They also allow individuals to become trustees, and so have a say in how their anonymised data is used.

Further issues were identified around the sometimes-fuzzy borderline between open data and personal data. In particular, the use of open data can make it more likely that identifiable characteristics may appear. Researchers from Belgium's Université catholique de Louvain (UCLouvain) and Imperial College London have built a model to estimate how easy it would be to de-anonymise any arbitrary data set.[73] A data set with 15 demographic attributes, for instance, "would render 99.98% of people in Massachusetts unique." This was discussed in Toronto, where there was concern that the use of government-held, aggregated data around health and social services could, for example, be used alongside data gathered while individuals move through the transport systems and within urban spaces, to re-identify individuals, and that the resulting insights could be used without the explicit consent of the those involved.[74] In order to minimise risk, appropriate levels of access and control need to be established. It should be possible to provide access to relatively basic data, such as high-resolution population data to humanitarian organisations in a conflict zone, for example, but not to the conflicting parties, such as the government forces and "rebel" forces who may use it to cause further harm. The question here is who or which organisation is best equipped to decide who gets access to what.

"More robust regulation is needed, including the ability to drive aggregation and anonymisation. If this is not possible, then the use of this information may only be reserved for academics who adhere to higher standards for data use than many in industry."

Copenhagen workshop

Implications for Data Value

Looking to the future, it seems there will be growing demands for greater clarity about exactly what data should be opened up, for what uses, and by who. Different types of information may require different types of use. Many in our workshops agreed that the purposes for which data is used, and the method of storage, should be open to scrutiny by cyber security experts. Regular transparency reports on who has access to such information would also go some way to reducing the risks.

"Who will pay to clean the data?"

San Francisco workshop

In Copenhagen, it was suggested that we need to define what we mean by the open use of commercial, sensitive, and non-sensitive data:

- For **commercial** data, where private companies and public bodies are both contributing information, a common ambition can encourage the opening up of data. *"The sharing of clinical trial data, to improve the benefits from drug development, is a good example of this."*

- Additional rules may be needed for **sensitive and personal** data, where privacy and security are paramount. *"More robust regulation is needed, including the ability to drive aggregation and anonymisation. If this is not possible, then the use of this information may only be reserved for academics who adhere to higher standards for data use than many in industry."*

- And for the majority of **non-sensitive** and public data sets, improving accessibility and increasing public awareness and data literacy will be essential.
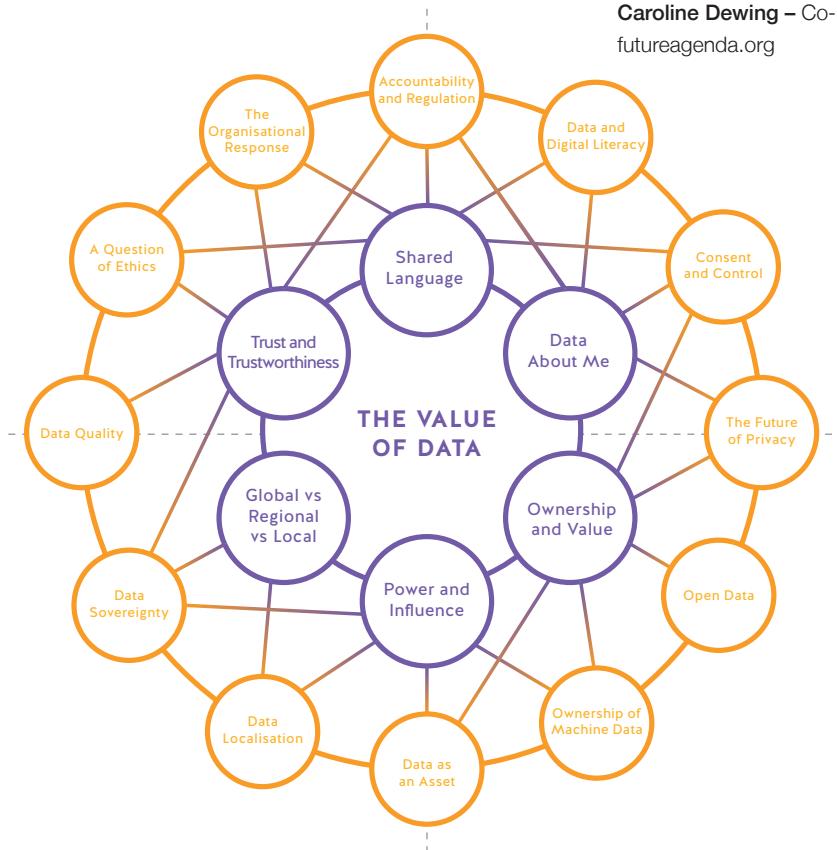
Context

This is one of 18 key insights to emerge from a major global open foresight project exploring the future value of data.

Throughout 2018, Future Agenda canvassed the views of a wide range of 900 experts with different backgrounds and perspectives from around the world, to provide their insights on the future value of data. Supported by Facebook and many other organisations, we held 30 workshops across 24 countries in Africa, Asia, the Americas, and Europe. In them, we reviewed the data landscape across the globe, as it is now, and how experts think it will evolve over the next five to ten years.

The aim of the project was to gain a better understanding of how perspectives and priorities differ across the world, and to use the diverse voices and viewpoints to help governments, organisations, and individuals to better understand what they need to do to realise data's full potential.

From the multiple discussions 6 over-arching themes were identified alongside 12 additional, related future shifts as summarised in the diagram below.

About Future Agenda

Future Agenda is an open source think tank and advisory firm. It runs a global open foresight programme, helping organisations to identify emerging opportunities, and make more informed decisions. Future Agenda also supports leading organisations, large and small, on strategy, growth and innovation.

Founded in 2010, Future Agenda has pioneered an open foresight approach bringing together senior leaders across business, academia, NFP and government to challenge assumptions about the next ten years, build an informed view and establish robust growth strategies focused on major emerging opportunities. We connect the informed and influential to help drive lasting impact.

For more information please see:
www.futureagenda.org

For more details of this project contact:
Dr Tim Jones – Programme Director, tim.jones@futureagenda.org
Caroline Dewing – Co-Founder, caroline.dewing@ futureagenda.org

Details of each of these, a full report and additional supporting information can all be found on the dedicated mini-site: www.deliveringvaluethroughdata.org