FUTURE AGENDA
Open Foresight

DELIVERING VALUE THROUGH DATA

Insights from Multiple Expert Discussions Around the World

First published November 2019 by:
Future Agenda Limited
84 Brook Street
London
W1K 5EH
www.futureagenda.org

Contents

1.0 Delivering Value Through Data 4

1.1 Key Findings 6

1.2 Consensus and Disagreement 8

1.3 Our Approach 10

1.4 Our Hosts 13

2.0 Setting the Scene 16

2.1 What is Data? 20

2.2 What Makes Data Uniquely Special? 22

2.3 Debating Data 23

2.4 Getting Value from Data 24

3.0 Overarching Themes 26

3.1 Data about Me 28

3.2 Ownership and Value 31

3.3 Power and Influence 35

3.4 Global vs Regional vs Local 40

3.5 Trust and Trustworthiness 44

3.6 Shared Language 54

4.0 Key Future Shifts 56

4.1 Data and Digital Literacy 58

4.2 Culture, Governance and Privacy 64

4.3 Consent and Control 72

4.4 Open Data 78

4.5 Ownership of Machine Data 86

4.6 Data as an Asset 92

4.7 Data Localisation 100

4.8 Data Sovereignty 106

4.9 Data Quality 114

4.10 A Question of Ethics 122

4.11 The Organisational Response 130

4.12 Accountability and Regulation 140

5.0 Conclusion 150

5.1 Questions 154

Workshop Locations 158

References 160

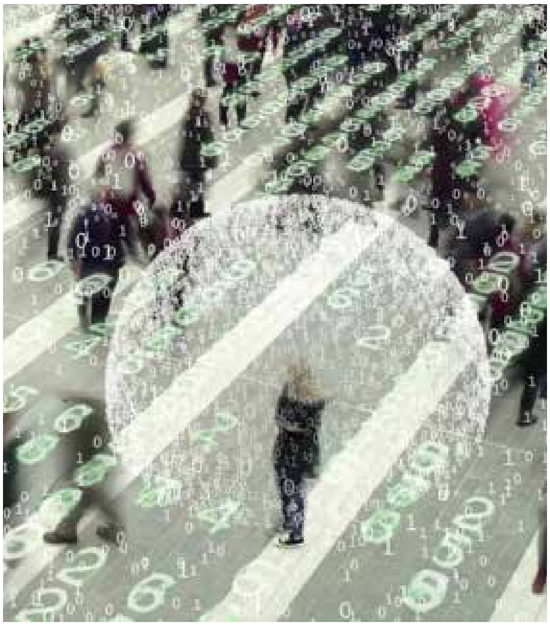
1.0 Delivering Value Through Data

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? 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.

To achieve this, we did three things. First, in each locality we brought together as wide a variety of people of different perspectives and disciplines as possible: Policy makers, corporate professionals, start-ups, NGOs, students, think tanks. Second, we asked participants to identify and prioritise the themes they considered to be most important in terms of opportunity and concern. Third, we asked them to debate in depth the highest priority issues, to identify areas of agreement and disagreement, and map out possible paths forward.

We are not aware of any other exercise of this scale or scope. No other project we know of has carefully and methodically canvassed the views of such a wide range of experts from such a diverse range of backgrounds and geographical locations. The result, we hope, delivers a more comprehensive picture of the sheer variety of issues and views thrown up by a fast-evolving 'data economy' than can be found elsewhere. And, by providing this rich set of perspectives, we aim to help businesses and governments - to develop the policies, strategies, and innovations that realise the full potential of data (personal, social, economic, commercial), while addressing potential harms, both locally and globally.



Our findings provide insights into:

- How priorities differ across geographical locations
- Areas where the actual perception and understanding of key issues differ (for example, over what the risks and opportunities are)
- Areas of broad consensus
- Areas of disagreement as to what issues should be prioritised and how to deal with them

1.1 Key Findings

The broad open-ended nature of this study provided participants with the opportunity to drive their own agenda. In each location around the world, workshop participants could prioritise the issues they felt to be most important. The resulting discussions covered a wide range of subjects from ownership of machine data, through to the potential of Open Data, to and whether or not 'informed consent' is a workable way to ensure fair and, equitable uses of personal data.

Wherever we were in the world, however, six over-arching themes informed virtually every discussion. They were:

1. Data about Me: One of the most valuable but contentious forms of data is data from and about individuals. How to deal with issues related to personal data generated a wide range discussions and dilemmas, including (for example) debates about privacy, the efficacy of mechanisms such as 'informed consent', who should have control over this data, the extent to which individuals are getting a 'fair share' of its benefits, and the degree to which they are able to participate in the new data economy. With the exception for of the need for 'privacy', where there were some strong differences in opinion, there was widespread agreement that these issues are important. Pressure for solutions that 'empower' individuals further is strong, but is there a broadly acceptable solution, and what does it look like?

2. 'Ownership' and Value: In discussion after discussion, we found that 'ownership' and rights to the value extracted from data are inextricably linked in peoples' minds. The assumption is that if we can agree on the 'ownership' of data, then we can sort out who is entitled to what "fair share" of value. But the more these discussions progressed, the less helpful traditional notions of 'ownership' seemed to be. If exclusive 'ownership' is a questionable concept in the context of data, what is the alternative? Many alternative ways of thinking about this (discussed later in this report) were proposed: the debate is only just beginning.

3. Power and Influence: Data is becoming a means of exercising power, as well as a focus for the multiple struggles for power. This power can come in many forms. It could be the power to make decisions that affect peoples' lives by, for example, giving or withholding their access to services. Some organisations' use of data gives them the power to act as 'choice architects', deciding what information is to be presented to people and how. Concentrations of data can create concentrations of economic power, which in turn could affect the distribution of available benefits. Is there are 'right' or 'best' balance or sharing of power between different parties and stakeholders. If so, what does it look like?

4. Global, Regional, and Local: Many workshop participants took it for granted that the reach and influence of global 'Big Tech' firms will simply continue to grow. But there is powerful sentiment, especially in fast-growing regions such as Africa and India, that governments should assert more control over data to protect citizens' rights, develop local economies, and maintain a sense of cultural identity. Some saw this as a necessary reaction to ongoing 'data imperialism', particularly by US- based west coast technology companies. Interestingly, participants in several Western workshops were quick to dismiss these concerns, which suggests mutual misunderstanding between key parties in different regions could intensify.

5. Trust and Trustworthiness: In workshops around the world, there was a widespread sense that very few organisations, if any, can be trusted with data, without any checks and balances. Indeed, apart from some nations where trust in government remains high, there was a common feeling that levels of trust in all established institutions, both government and business, are in decline - just when, arguably, increasing levels of trust are needed. This suggests that there will be increasing pressure for organisations to demonstrate trustworthiness. What do they have to do to achieve this? And policy makers and regulators will come under equally intense pressure to answer the question, 'on what basis can/should organisations be trusted with data?'

6. Shared Understanding: From Abidjan to Bogota, Bengaluru to Stockholm, workshop participants were keenly aware that societies are still struggling to understand what the key issues are, and what to do about them. While 'everyone knows' that data can be extremely valuable, people are much less clear on where this value comes from, or what forms it can or will take. Issues such as data ethics, the potential impact of AI and of machine generated data often seemed dauntingly complicated and difficult to understand. Despite this, there is a huge appetite to find universal approaches in dealing with these complexities. Many agreed that the first step is to establish a common language about data that can help provide clarity about of terms, issues, and implications, in order to point a way forward.

1.2 Consensus and Disagreement

One key message emerges from debates that were held about these issues: the debate around data globally has multiple potential fault lines, each one of which needs to be addressed if significant progress is to be made. People have different understandings and perceptions as to what is actually happening: they disagree about ‘the facts’. They have different and often (at least apparently) conflicting vested interests. Governments, large global corporations, small local businesses, individuals as citizens and consumers, all want different things. And people are bringing different norms and values to the debate - widely diverging but strongly held beliefs about ‘what is right’ and ‘what is fair’. For example, in Asia in particular, there were multiple conversations around the conflict between eastern and Western philosophies, and how their different approaches will influence, for example, the development of AI and machine learning driven by the data.

This project is all about opening up a better-informed dialogue about the dialogue around the culture and context of data - and about the need for those in power to listen to voices which they may not be hearing. It’s almost impossible to make wise, informed decisions when we don’t fully understand the landscape we are operating within.

Within these broad cross-cutting themes, the issues which inspired the largest number of conversations were:

- The need for greater digital literacy amongst consumers, citizens, employees, policy makers, and regulators;
- The potential to value data as a commercial/economic asset and the implications this may have for how data-based businesses are valued and taxed;
- How to establish a durable regulatory environment that effectively remedies harms and protects users without stifling future innovation;
- The failure of ‘informed consent’ to give individuals control over the use of their personal data.;

Within these discussions, there were areas of global consensus. They included:

- Concern around a lack of transparency about how data is collected, classified, interpreted, and used. This is undermining trust in business and hampering policy makers’ ability to develop robust checks and balances;
- The need for organisations to become more accountable for the ways personal data is used and shared;
- The need to create a common language for data and the issues related to its use;
- The as yet untapped value of sharing data sets that would particularly benefit society. Potential benefits include improved health, transport, and security services. But greater data sharing could also benefit economic growth more broadly.

In contrast, there were also areas of clear disagreement. These include debates around:

- Open data (particularly open public data): Some believe open data offers the best chance of unlocking the potential of big data to solve societal challenges and bring collective benefit. However, others described the exact same efforts as increasing imbalances of power and reward by handing society's most valuable data assets to those most able to exploit them, whether for the public good or not;
- Data 'ownership': While everyone agreed on the need for clearer rules as to who gets what value from data, there was little agreement on what these rules should be, how they should be arrived at, or enforced;
- Personal data privacy and national security: For some, a regulated erosion of privacy is a necessary and reasonable price to pay for heightened national security; for others, even small erosions of privacy set us on the slippery slope towards a kind of society that is to be feared far more than any piecemeal threats to national security;
- Data sovereignty: Many workshop participants, particularly in emerging economies, were strongly in favour of efforts to promote data sovereignty as a necessary measure to address concerns over privacy, consumer rights, domestic law enforcement and cyber security, and national economic growth. But for others, this is considered to be little more than an effort to close digital markets — with the ultimate impact of reducing the efficiency of global services increasing costs for small businesses, and dampening innovation.

1.3 Our Approach

The programme described above followed Future Agenda's Open Foresight model to gather and develop emerging views on a broad theme, to offer new perspectives on the opportunities and challenges facing governments, business, and society. Broadly speaking, this consists of three steps

1) An initial perspective provocation
2) Facilitated workshops
3) Synthesis of emerging views

The initial perspective on the Value of Data was written by Future Agenda's Director of Global Insight, Dr Robin Pharaoh. In it, he points out the scale of change that we are experiencing around the collection of data:

"It is the movement of data collection and analysis, experiment, and discovery from remote and singular processes, to the most intimate and fundamental parts of everyone's personal, social, and economic lives, seemingly without limit and without end, that has driven the idea of data into the heart of contemporary social and political discourse."

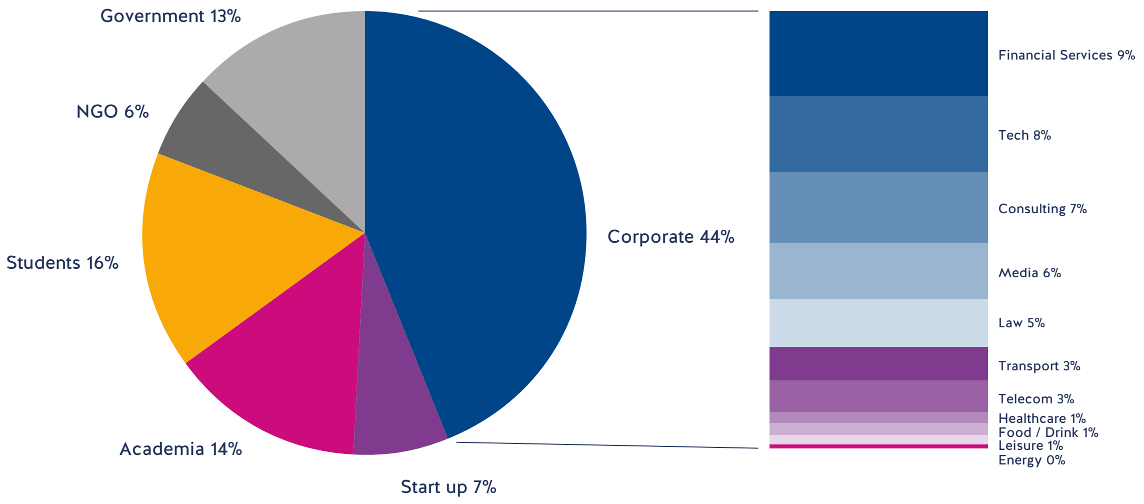
He then argues that "the value of data lies in the uses to which it is put," and goes on to frame questions around what types of data drive the most positive value in what contexts. As the perspective is designed to stimulate debate, he ends with specific questions. These include:

- Who benefits from the value derived from data?
- Who is best placed to use data to drive positive social value?
- What are the trade-offs and downsides of mass data collection, storage, and use, and critically, who is monitoring or accountable for these?

These questions, alongside the insights from the previous programmes, were then used as the point of departure for the subsequent expert workshops which took place across the world.

Workshop format

Workshops were then held in 30 different locations in 24 countries around the world. In all, just over 900 experts from many different industries and sectors took part. Each workshop comprised between 25 – 35 people. In addition to our own network, we worked closely with our sponsors and partnered with think tanks, NGOs, and academics in order to identify participants from as diverse backgrounds and disciplines as possible. As a result, we were able to talk to experts from academia, the technology sector – both start-ups and established players - entrepreneurs, foundations, the voluntary sector, government officials, industry body representatives, corporates, private industry, professional networks, and social media experts.



Future Value of Data - Participants Background

Each event was run under the Chatham House Rule, thus allowing free discourse, so that assumptions could be challenged, new perspectives shared, and insightful and pragmatic views on how change is most likely to occur, actively debated.[1]

All the workshops followed the same process. Starting with insights drawn from the initial perspective and previous discussions, they focussed on identifying the key issues, adding additional views and insights, and highlighting pivotal areas for future innovation and change, globally and locally. The new insights and ideas generated were carried through into follow on sessions to ensure iteration and scrutiny.

Each event also followed the same basic format:

- Group feedback on the insights gained from the initial perspective and from the previous discussions;
- Table debate to agree the relevance of each insight for their market or sector, and prioritise them as areas of high, medium, or low significance, according to their particular perspective;
- Plenary discussion so that all the groups can compare their point of view with others;
- Table discussion to uncover areas or issues which might not yet have been addressed and should be included;
- Participants were invited to vote individually on the insights that they thought would have the most impact in their country or region;
- New groups were consequently formed so that they could explore the most highly rated topics in more detail. This included discussion on the drivers of change, and probably pathways for the future;
- These were then shared and challenged in a plenary session at the close of the workshop.

In this way, experts in each locality were both voting on which issues they see as the most important, and then detailing the future impacts and implications – locally and globally. At the end of each event, a detailed write up was shared with all participants, who were asked to check it for accuracy. If necessary, modifications were then made, and the final write up was shared with all those who had contributed to the other discussions that had taken place around the world.

An interim paper was published in August 2018, which provided an overview of the views from the first 18 workshops.[2] A further 12 workshops were then held between September and December. By the end of the programme, 70 unique insights had been gathered. These were then synthesised into 14 clusters and shared as a presentation in December 2018. Further academic and industry expert feedback and iteration during much of 2019, has added more context and refined the issues which are now integrated into the foresights summarised in this report.

1.4 Our Hosts

The initial sponsor for this programme was the Privacy and Data Policy team at Facebook, who funded just over half the workshops in total and have been involved in the writing of this report. We would also like to thank Unities Ltd. for their support, commentary and expert research during its development.

As with all Future Agenda programmes, we were keen to engage others in the discussions, and therefore we reached out to a wide range of organisations, including other corporates, academics, NGOs, government bodies, and think tanks, to ask for their support. Sometimes this was financial, sometimes it was operational – for example, by providing space for a room for a workshop and helping to identify participants. In total, thirty-four organisations became involved in developing the programme. As a result of their help, we were able to run 30 workshops in 24 countries throughout 2018, and, in so doing, engaged with approximately 900 experts from around the world.

We would like to thank all those who contributed in whatever way for the time and effort they gave us. This report would not have been possible without their generous support.

Dr Tim Jones and Caroline Dewing

November 2019

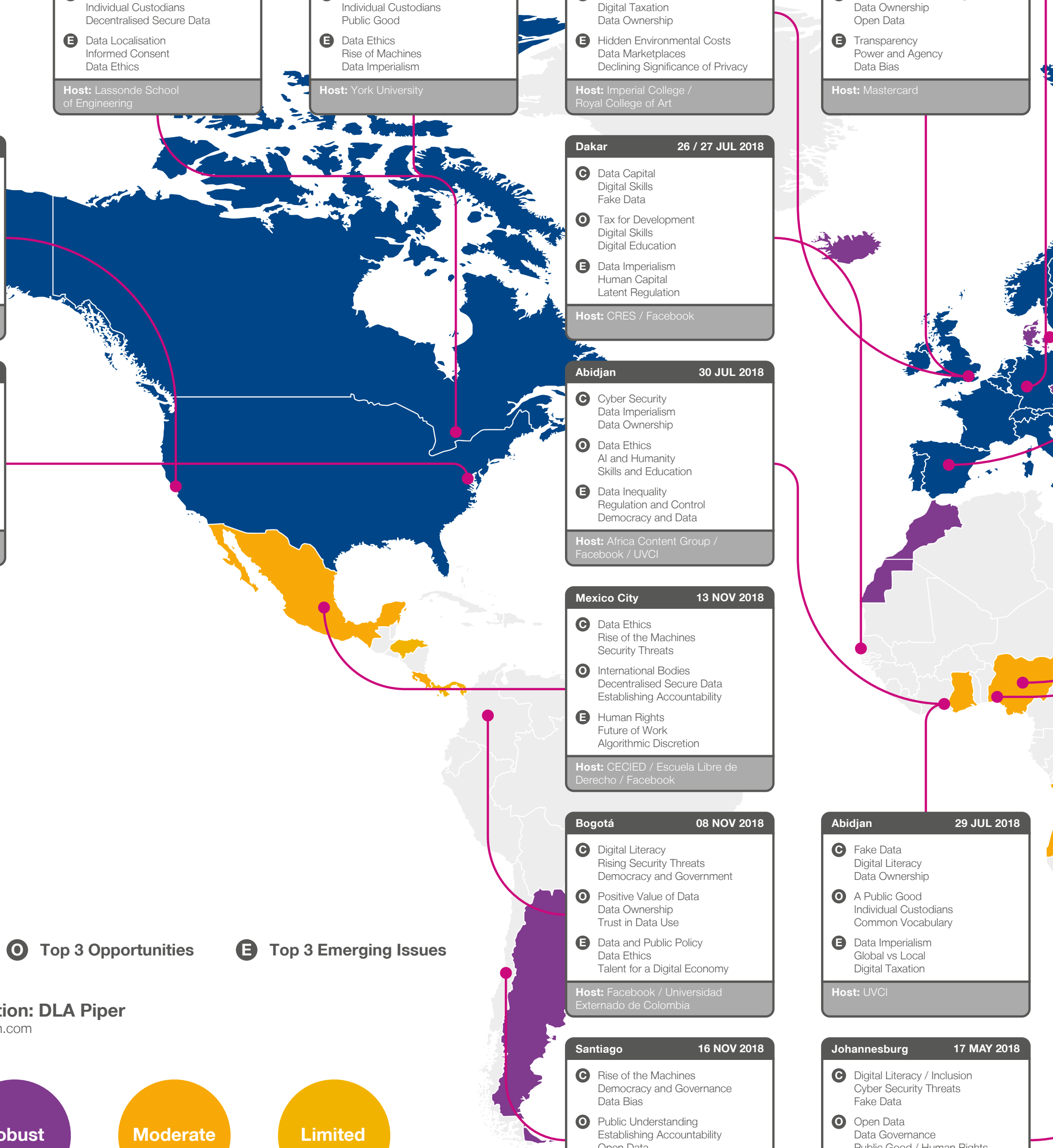
AFRICA CONTENT GROUP
ATENEO DE MANILA UNIVERSITY
BROOKINGS
Cámara de Comercio de Bogotá
CARNEGIE INDIA
CECIED
CIPIT
CRES
Cognizant
DICT
Danmarks Tekniske Universitet DTU
DMCC
Escuela Libre de Derecho Universidad
facebook
ISGPP
IBM
Imperial College London
ispim
KICTAnet
KTH
LKY Lee Kuan Yew School of Public Policy National University of Singapore
mastercard
科学技術・学術政策研究所 National Institute of Science and Technology Policy
Orange Silicon Valley
Pontificia Universidad Católica de Chile
RISTEX 社会技術研究開発センター
Royal College of Art
TACIT Knowledge Alliance
TAL
Universidad Externado de Colombia
香港大學 THE UNIVERSITY OF HONG KONG
UNIVERSITEIT VAN PRETORIA UNIVERSITY OF PRETORIA YUNIBESITHI YA PRETORIA
UVCI
LASSONDE YORK U

Delivering Value Through Data

Insights from Multiple Expert Discussions Around the World

Future Value of Data

Key Insights: 2018



Frankfurt — 08 NOV 2018

C Cyber Security Threats
Data Liability
Data Bias

O Personal Choice
New Business Models
Transparent Algorithms

E Data Ethics
Social Impact
Data Capital

Host: Cognizant

Toronto — 09 NOV 2018

C Rising Security Threats
Digital Literacy
Fake Data

O Open Data
Individual Custodians
Decentralised Secure Data

E Data Localisation
Informed Consent
Data Ethics

Host: Lassonde School of Engineering

Toronto — 12 NOV 2018

C Digital Literacy
Data Bias
Informed Consent

O Open Data
Individual Custodians
Public Good

E Data Ethics
Rise of Machines
Data Imperialism

Host: York University

London — 01 OCT 2018

C Rising Cyber Security Threats
Rise of the Machines
Fake Data

O Open Data
Digital Taxation
Data Ownership

E Hidden Environmental Costs
Data Marketplaces
Declining Significance of Privacy

Host: Imperial College / Royal College of Art

London — 13 SEP 2018

C Data Ethics
Partial Inclusion
Rise of Machines

O Contextual Data Sharing
Data Ownership
Open Data

E Transparency
Power and Agency
Data Bias

Host: Mastercard

San Francisco — 30 OCT 2018

C Digital Literacy
Data Bias
China vs. The US

O Responsible Sharing
Democracy and Government
GDPR Setting Standards

E The Nature of the Firm
Data Ethics
Data Imperialism

Host: Facebook / Orange Silicon Valley

Dakar — 26 / 27 JUL 2018

C Data Capital
Digital Skills
Fake Data

O Tax for Development
Digital Skills
Digital Education

E Data Imperialism
Human Capital
Latent Regulation

Host: CRES / Facebook

Washington DC — 02 NOV 2018

C Digital Literacy
Data Bias
Cyber Security Threats

O Open Data
Data Ethics
Trust in Data Use

E Future of Work
Fake Data
Data Imperialism

Host: Brookings Institution / Facebook

Abidjan — 30 JUL 2018

C Cyber Security
Data Imperialism
Data Ownership

O Data Ethics
AI and Humanity
Skills and Education

E Data Inequality
Regulation and Control
Democracy and Data

Host: Africa Content Group / Facebook / UVCI

Mexico City — 13 NOV 2018

C Data Ethics
Rise of the Machines
Security Threats

O International Bodies
Decentralised Secure Data
Establishing Accountability

E Human Rights
Future of Work
Algorithmic Discretion

Host: CECIED / Escuela Libre de Derecho / Facebook

Bogotá — 08 NOV 2018

C Digital Literacy
Rising Security Threats
Democracy and Government

O Positive Value of Data
Data Ownership
Trust in Data Use

E Data and Public Policy
Data Ethics
Talent for a Digital Economy

Host: Facebook / Universidad Externado de Colombia

Abidjan — 29 JUL 2018

C Fake Data
Digital Literacy
Data Ownership

O A Public Good
Individual Custodians
Common Vocabulary

E Data Imperialism
Global vs Local
Digital Taxation

Host: UVCI

Santiago — 16 NOV 2018

C Rise of the Machines
Democracy and Governance
Data Bias

O Public Understanding
Establishing Accountability
Open Data

E Transparency
Data Ethics
Data Politics

Host: Facebook / Pontificia Universidad Católica de Chile

Johannesburg — 17 MAY 2018

C Digital Literacy / Inclusion
Cyber Security Threats
Fake Data

O Open Data
Data Governance
Public Good / Human Rights

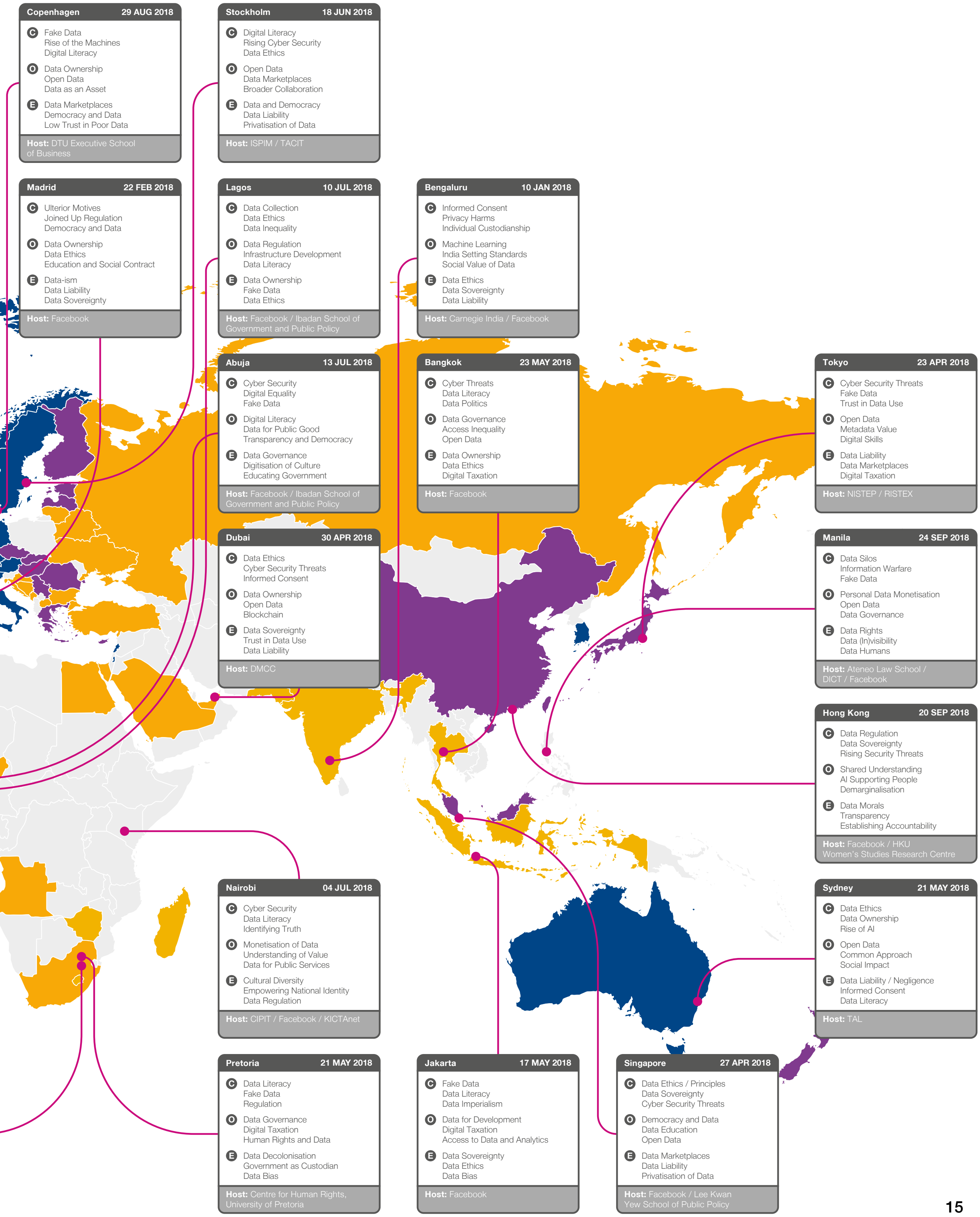
E Data Ethics
Data Sovereignty
Privatisation of Data

Host: Facebook / IBM Research

C **Top 3 Challenges** O **Top 3 Opportunities** E **Top 3 Emerging Issues**

Level of Privacy Regulation: DLA Piper
https://www.dlapiperdataprotection.com

● Heavy ● Robust ● Moderate ● Limited



Copenhagen 29 AUG 2018

C Fake Data
Rise of the Machines
Digital Literacy

O Data Ownership
Open Data
Data as an Asset

E Data Marketplaces
Democracy and Data
Low Trust in Poor Data

Host: DTU Executive School of Business

Stockholm 18 JUN 2018

C Digital Literacy
Rising Cyber Security
Data Ethics

O Open Data
Data Marketplaces
Broader Collaboration

E Data and Democracy
Data Liability
Privatisation of Data

Host: ISPIM / TACIT

Madrid 22 FEB 2018

C Ulterior Motives
Joined Up Regulation
Democracy and Data

O Data Ownership
Data Ethics
Education and Social Contract

E Data-ism
Data Liability
Data Sovereignty

Host: Facebook

Lagos 10 JUL 2018

C Data Collection
Data Ethics
Data Inequality

O Data Regulation
Infrastructure Development
Data Literacy

E Data Ownership
Fake Data
Data Ethics

Host: Facebook / Ibadan School of Government and Public Policy

Bengaluru 10 JAN 2018

C Informed Consent
Privacy Harms
Individual Custodianship

O Machine Learning
India Setting Standards
Social Value of Data

E Data Ethics
Data Sovereignty
Data Liability

Host: Carnegie India / Facebook

Abuja 13 JUL 2018

C Cyber Security
Digital Equality
Fake Data

O Digital Literacy
Data for Public Good
Transparency and Democracy

E Data Governance
Digitisation of Culture
Educating Government

Host: Facebook / Ibadan School of Government and Public Policy

Bangkok 23 MAY 2018

C Cyber Threats
Data Literacy
Data Politics

O Data Governance
Access Inequality
Open Data

E Data Ownership
Data Ethics
Digital Taxation

Host: Facebook

Tokyo 23 APR 2018

C Cyber Security Threats
Fake Data
Trust in Data Use

O Open Data
Metadata Value
Digital Skills

E Data Liability
Data Marketplaces
Digital Taxation

Host: NISTEP / RISTEX

Dubai 30 APR 2018

C Data Ethics
Cyber Security Threats
Informed Consent

O Data Ownership
Open Data
Blockchain

E Data Sovereignty
Trust in Data Use
Data Liability

Host: DMCC

Manila 24 SEP 2018

C Data Silos
Information Warfare
Fake Data

O Personal Data Monetisation
Open Data
Data Governance

E Data Rights
Data (In)visibility
Data Humans

Host: Ateneo Law School / DICT / Facebook

Hong Kong 20 SEP 2018

C Data Regulation
Data Sovereignty
Rising Security Threats

O Shared Understanding
AI Supporting People
Demarginalisation

E Data Morals
Transparency
Establishing Accountability

Host: Facebook / HKU Women's Studies Research Centre

Nairobi 04 JUL 2018

C Cyber Security
Data Literacy
Identifying Truth

O Monetisation of Data
Understanding of Value
Data for Public Services

E Cultural Diversity
Empowering National Identity
Data Regulation

Host: CIPIT / Facebook / KICTAnet

Sydney 21 MAY 2018

C Data Ethics
Data Ownership
Rise of AI

O Open Data
Common Approach
Social Impact

E Data Liability / Negligence
Informed Consent
Data Literacy

Host: TAL

Pretoria 21 MAY 2018

C Data Literacy
Fake Data
Regulation

O Data Governance
Digital Taxation
Human Rights and Data

E Data Decolonisation
Government as Custodian
Data Bias

Host: Centre for Human Rights, University of Pretoria

Jakarta 17 MAY 2018

C Fake Data
Data Literacy
Data Imperialism

O Data for Development
Digital Taxation
Access to Data and Analytics

E Data Sovereignty
Data Ethics
Data Bias

Host: Facebook

Singapore 27 APR 2018

C Data Ethics / Principles
Data Sovereignty
Cyber Security Threats

O Democracy and Data
Data Education
Open Data

E Data Marketplaces
Data Liability
Privatisation of Data

Host: Facebook / Lee Kwan Yew School of Public Policy

2.0 Setting the Scene

It has never been easier to gather and store information. Data is now a key raw material of business, government, and society. For governments, business, and even private citizens, data is cheap, widely available, and relatively easy to access, and its use influences almost all aspects of how our society works. Organisations and governments use it to conduct their operations - whether that is the delivery of services such as financial advice or healthcare, and the sharing of information in the media, or to make decisions such as what products should be made available to who and so on. People use data across the board: to access and use services, stay in touch with friends or family, administer things, make decisions, and to undertake multiple tasks such as take exercise or even (increasingly) to find romance.

Beyond this, data is opening up new frontiers in science and the humanities, from extending our knowledge of how the universe is built, to creating new understanding around climate change, to discovering the impact of a specific teacher on a specific pupil's performance. All this suggests that data has potentially enormous value.

What is surprising, however, is just how little consensus there is around exactly what this value is, or where this value comes from. Many fundamental questions remain unanswered and are subjects of continued debate and controversy.

Some of these questions include:

- Why exactly is data valuable? What are the benefits we derive from it?
- Is this value of data mainly financial/monetary, or does it derive from direct utility - the things we can do with it? If so, what sorts of things? Does the value derive from the ability to gain new insights, or perhaps to streamline or automate processes?
- What are the different ways in which this value is generated?
- Are the benefits being shared fairly? If not, how can a fairer share of benefits be achieved?
- What are the main barriers and obstacles to realising the full potential of data, and what is the best way to address them?
- Are there also harms as well as benefits? If so, what can we do to alleviate them?

Without a collective understanding of the distinctive characteristics of data and the multiple different ways it can be put to use, opportunities to extract value may be missed.

To this, we need to add the complexities of different cultures, different types of technology, and hugely differing stages of technology development and adoption around the world. This changes peoples' and societies' experiences of using data, their perceptions, and their priorities. The same uneven development also makes it unlikely that the benefits of data will be uniformly shared. And without a collective understanding of the distinctive characteristics of data and the multiple different ways it can be put to use, opportunities to extract value may be missed. Different innovations, perspectives, priorities, and initiatives from multiple sectors and across geographies, means that we are likely to see a myriad of future pathways. However, if we explore the different approaches to valuing data, and try to understand the rationale behind them, we may be better prepared for change.

This report highlights those issues and questions which were most debated during our research, and provides an overview of the different points of view around the world. We do not intend to cover every aspect of the Value of Data, or every argument or counter argument made in relation to the points we raise. However, we hope that we have managed to help clarify the issues that were important to those we engaged with.

2.1 What is Data?

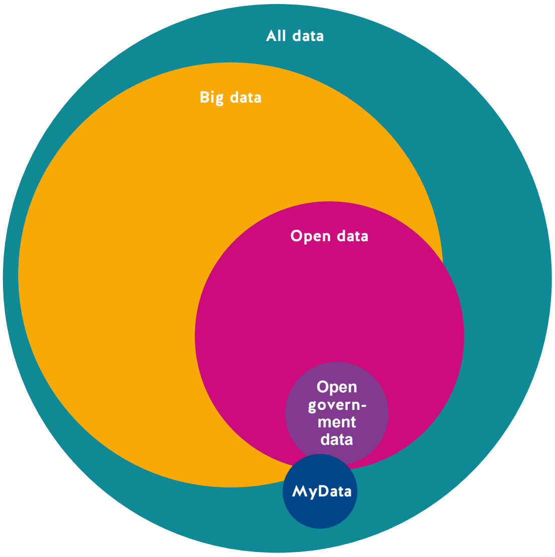
Before addressing these questions, we need to clarify terms. Some, if not all, hotly debated issues relating to data, stem from the fact that people are talking at crossed purposes. They haven't clarified what data actually is and they end up using the same word to describe different things. Many discussions about data may presume a shared understanding between stakeholders that may not actually exist. For the purposes of this report, the following definitions may help.

All data begins with a **data point.** This is a discrete unit of information, such as the temperature (Celsius or Fahrenheit) at a particular location at a particular time. A **data set** is a collection of data points. For example, it could be the different temperatures in a particular location at different times, or in different locations at the same time. Combining different data sets, (temperatures in the same place but at different times; or in different places at the same time) helps us get a richer understanding - such as how temperatures vary between night and day and across seasons in one location, or the differences between this location and other locations. When combined, they could (for example) become vital elements of the study of climate change.

Meta data provides information about other data. Keeping with our example about temperature, this could be data about the device used to measure the temperature, or who did the measuring. Other examples of metadata include the time of receipt of an email or phone call, or the location where a picture was taken, but not the email, the phone call, or the picture itself.

According to the definition made by European data protection rules (the General Data Protection Regulation, or GDPR), **personal data** is any information relating to an identified or identifiable natural person (a **data subject**). This is a living individual who can be identified, directly or indirectly, by reference to an identifier, such as a name, an identification number, location data, an online identifier, or to one or more factors specific to their physical, physiological, genetic, mental, economic, cultural, or social identity. **Non-personal data** is any data that cannot be connected to an individual, such as a company email address, info@, or indeed a company registration number. It can also include **pseudonymous** or **anonymous data** that is personal data that has been, ideally but not definitely, irreversibly de-identified.

Structured data is made up of clearly defined data which is formatted so that it fits into a formal database and is easily searchable; while **unstructured data** is not organised in a pre-defined manner and is usually not as easily searchable; it could include formats like audio, video, and social media postings.



Relationship of Different Data Types

Big data usually includes data sets with sizes beyond the ability of commonly used software tools to capture, curate, manage, and process data within a tolerable elapsed time. It includes structured and unstructured data. It is normally defined by its volume, its variety, its velocity (the speed at which it is generated), and its veracity (where it originates from and how complete it is). While there is much debate about big data in day-to-day life, the vast majority of data-driven activities and processes are driven by '**small data**': discrete bundles of specific data points that are uniquely relevant to the task in hand. These currently make up most of the activities and transactions which use data.

Data comes from a vast range of sources. For example, in our daily lives we create a massive amount of digital information about ourselves. This ranges from data that is captured via technological means, barcodes, online systems, credit cards, and so on; it is 'volunteered' by individuals, sometimes formally through form filling or informally on social media; it is also generated as a by-product of operations such as banking, measuring pollution, or using security cameras. Some data is **personally identifiable**, such as bank details, while other data is **non-identifiable**, for instance, statistical data about pollution or traffic flow. Proxy data is data used to study a situation, phenomenon, or condition for which no direct information is available – for example, scientists use the measurement of tree rings as **proxy data** to estimate climate change variances.

Coincident with the revolution in the collection of data, we are experiencing dramatic changes in the technologies of **data capture, storage, analysis,** and **transmission**. Together, these technological advances are turning our society and economy from a data desert to data ocean. This is not only about the quantity of data available, but also around the ability to leverage it across organisations and industries.

'Data products' are created by the aggregation of data to establish a new, higher level data point that can be used for a particular purpose - for example, a profile, identity, or credit reference. Good examples of this are Google search and Amazon product recommendations, both of which improve as more users engage.

But 'data products' are just one of the many and varied uses of data. Data is essential for all record keeping and administration, and for organising and coordinating activities. For example, individuals use data when they say, 'I'll meet you at [this place] and [this time]', while retailers and producers use data to organise and manage their supply chains. Many industries' core operations are data driven. Without data, banks cannot operate, because they cannot track who is moving money, how much money, or where. Data is also used to measure and monitor our environment, be that temperature recordings, health records, or economic statistics. Interrogating data helps reveal patterns, trends, and variances that previously weren't visible. This in turn helps us develop knowledge and understanding - and the resulting insights help us make better, more informed decisions.

2.2 What Makes Data Uniquely Special?

All of these varieties and differences are important when discussing data and how to unleash its value. But there is one special thing about data that unites them all. As a ‘resource’, data is different to traditional physical resources. Like knowledge and ideas, when data is ‘used’, it doesn’t get ‘*used up*’. This means the same piece of data can be used for multiple different purposes by multiple different parties. And far from being a depleting resource, it is an accumulating one.

The implications and ramifications of these unique characteristics of data are vast, as this report will show. They up-end our notions we often see as ‘fundamental’, such as that of ‘ownership’. They force us to challenge many of the assumptions that lie at the heart of economic analysis. They transform both the possibilities data creates and the dilemmas it generates. They change relationships between stakeholders, be they individuals, communities, networks, organisations, government, or wider society.

2.3 Debating Data

Given the multiple different types, forms, and varieties of data, and the equally multiple range of different uses, when debating issues relating to data, it is almost impossible to avoid the classic fable of the blind men and the elephant, each one touching a different part of the beast, drawing entirely different conclusions about its overall nature. To help them describe what they are experiencing, they naturally turn to analogies. These can be useful, as they capture certain aspects of the way that data behaves in relation to certain aspects of the economy and society, at certain times. But they can equally mislead, causing us to associate data with the wrong things, leading to wrong, even dangerous, conclusions. Here are some of the analogies that were discussed in our workshops.

Data as 'the new oil': Well, data is mined and refined, like oil. Vast hordes of it can make its owners (or 'controllers') very wealthy and powerful, like oil. We might even go to war over it, like oil. But there are also many ways in which data is not like oil. Unlike oil, data is not a finite, exhaustible resource. We have just seen that, unlike oil, when data is used, it doesn't get used up. Indeed, in many cases, data is replicable or reproducible, with the very process of using data creating *new* data - for example, the meta data about what data has been used, and for what purposes. Also, unlike oil, the material costs of extraction, collection, and movement of data are not high, and are falling rapidly. And data ownership is not easily defined, unlike oil.

These differences are important, since they point to a set of end-points for the data economy that are completely different to those of an oil economy, and so demand a different set of societal responses.

Data as currency: Data can certainly serve as a medium for exchange, as it does when a consumer, for example, shares their personal data in exchange for so-called 'free' services. It can also be used as a store of value, even in quite a literal (albeit unstable) sense when it comes to crypto-currencies. So yes, data can be used like a currency in some circumstances.

But describing data as currency really doesn't tell us much. It just tells us that data has exchangeable value in certain contexts. In that sense, many things operate like currency. The economic value of data might have risen in recent times, and more people might be aware of that value, but the same might also be said of quinoa. Describing data as currency can simply edit out many of its most important features.

Data as 'a periodic table': One suggestion, made in Singapore, that subsequently gained widespread support, was that data should be considered like the periodic table: *"Data is similar to the elements on the periodic table. They can act independently, but interact with each other to create new combinations."* Moreover, from a value perspective, different organisations have varied ideas of the value of the individual elements, depending on where they are looking from and what their field of activity is. But when a number of elements are combined into a compound, then it can be more useful with more value to more people, organisations, and society. As a metaphor, this seemed to work for many.

> "Data is similar to the elements on the periodic table. They can act independently, but interact with each other to create new combinations."
>
> Singapore workshop

2.4 Getting Value From Data

Generically speaking, participants talked about different types of value that data is used to generate:

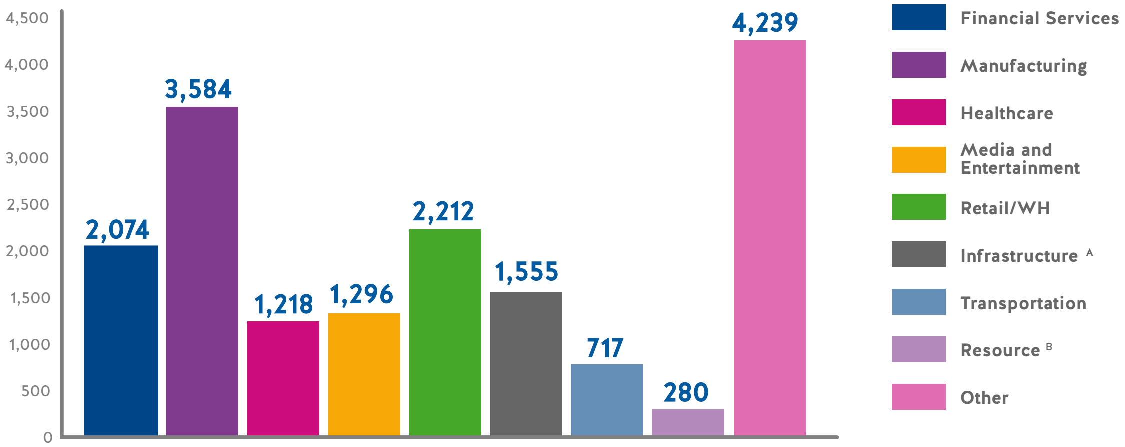
- use value: all the different uses to which data is put, as described above (administration, organisation and coordinating, analysis, decision-making)
- exchange value: the different ways that people can make money directly from data by selling, renting, trading, or other ways of charging for access to data

Stated boldly like this, the concept of value may seem simple. But, of course, it generates much complexity. In societal terms, much debate about 'value' is a proxy of what people think is 'good' versus what they think is 'bad': subjective judgement is never far away from debates about value. Other perspectives include those of different stakeholders, such as are we talking about value to a customer or to a firm? More sophisticated analyses highlight the difference between *potential* value and *realised* value. According to these arguments, value resides not in a product or service per se, but rather in a human being's experience. For example, a company's offering, whether it takes the form of a product, a service, or some mixture of the two, can be described as unrealised value (i.e. a store of potential value). This value is only realised when a customer uses it, and this is invariably an act of co-creation in a particular context.

In terms of practical application, value can be both positive and negative. Sometimes, data can provide unequivocally positive value, such as when large data sets are used to build smart energy or water grids, to improve travel safety, or to search for new cures for diseases. On other occasions, data may be used to generate unequivocally negative value, such as identity theft, cyber-attack, data blackmail, or the proliferation of false information ("fake news").

Other uses seem to allow for mixtures of both positive and negative value at the same time - or positive vs negative depending on how you look at it. Connecting people on a massive scale, for example, can enhance human relationships, allow ideas to flourish, and give voice to those who may not otherwise have one. But it can also enable bullying, criminality, or terrorism, give strength and credibility to bad ideas or ideologies, and encourage mob rule. Data-harvesting for surveillance purposes can help the development of new kinds of consumer products, be used for the purposes of delivering targeted advertising, or for feeding sophisticated algorithms that underlie the efficient delivery of services (policing, insurance, access to government services, etc.). But they can also be seen as incursions of, and threats to, privacy and civil liberties.

Much debate about 'value' is a proxy of what people think is 'good' versus what they think is 'bad'.



Source: IDC's Data Age 2025 study, sponsored by Seagate

[A] Infrastructure includes Utilities, Telecommunications

[B] Resource includes Oil and Gas (Mining), Transportation of oil & gas through pipelines or shipping, Resource industries, Petroleum and coal manufacturing/refining

Where the Data Lies: Global Enterprise Data by Industry (2018)

As will be discussed later, some suggest that a multi-capital view of data value should be used in line with what is proposed and adopted for the Integrated Reporting of an organisation's activities.[3] Others are proposing methods of valuing data against UN Sustainable Development Goals.[4] The IMF has recently hosted conferences exploring how to measure the value of an organisation's data. And in a world of 'free' and 'open' data, it is also looking at the implications of capturing digital impact within national accounts and GDP.

We will discuss these rich and varied views and nuances in this report, but the core of it remains relatively simple: the value of data (both positive and negative) lies in how it is used and/or how it is exchanged, and the effect that this can have on the economy, society, and individuals.

3.0 Overarching Themes

Throughout the programme, six major cross-cutting themes emerged. They were discussed in different ways and there is a degree of overlap between them, but no matter what particular data-related issue was discussed, one or all of them is likely to be a key feature of the debate.

These themes are:

1. The issues relating to the collection and use of personal data – **data about me;**

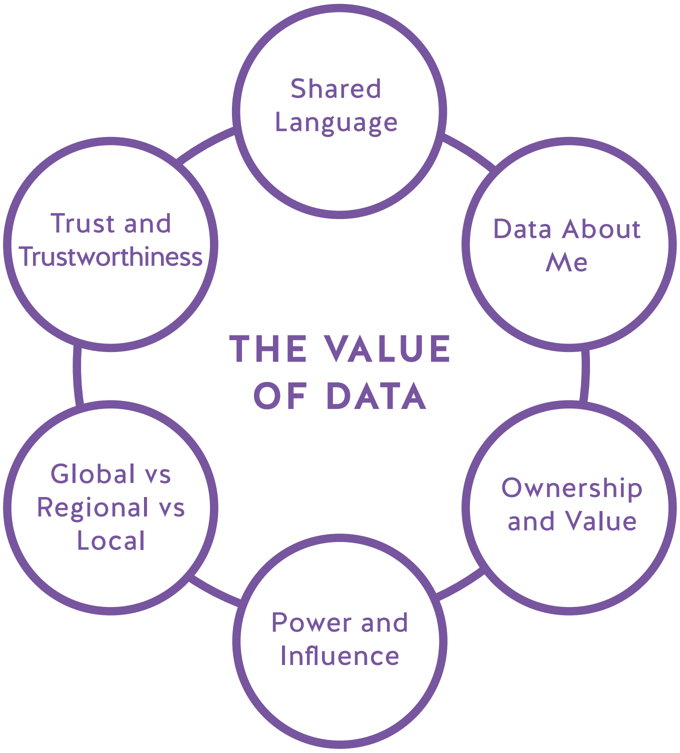
2. Topics linking to **ownership and value;**

3. Issues concerning the exercise of **power and influence;**

4. Matters relating to the level at which we are operating, such as **global versus regional versus local;**

5. Differing perspectives on **trust and trustworthiness;** and

6. The need for a **shared language** that avoids misunderstanding and confusion, and helps to clarify and advance the debate.



3.1 Data About Me

Rising concerns about personal data collection and use cover many issues. Pressure for solutions that inform and 'empower' individuals is growing.

In our workshops, much of the debate about data focused on personal data. This is not surprising. By definition, personal data relates most closely and directly to individuals' lives in many ways. Data about an individual may reveal intimate details about their lives. It could be - and is - used to bring them many benefits in terms of innovative, personalised services. But it could also render them vulnerable, especially if it gets into the wrong hands (for example via identity theft), or used 'against' rather than 'for' them (discriminating against individuals or groups of people based on what data reveals about them).

Personal data is also where debates about power and fairness is most acute. Huge amounts of money are being made by some profit-seeking companies via their collection and monetisation of the data of billions of individuals. Many individuals feel powerless in the face of these corporations and their intense concentrations of data power.

Such issues exercised the minds of many workshop participants, who wanted to analyse exactly what is going on in relation to the collection and use of personal data - and to find positive ways forward. It wasn't easy - partly because issues relating to personal data can be far more complex than they appear at first sight - starting with definitions.

Many people, when they talk about personal data, refer to very obvious bits of data such as name, address, contact details, payment card details, medical data, or personal purchase history. But the European General Data Protection Regulations (GDPR) go much further, defining personal data as: "any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier, or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural, or social identity of that natural person."

By potentially including data points such as cookies ('online identifiers') and location data, this European definition of personal data casts the net much wider than many anticipate. As we will see in our discussion of the 'Internet of Things' and 'machine to machine' data, if it generates data that relates to an identifiable individual (for example, their usage of a device) in some jurisdictions, it will be seen as personal data. The border lines between 'personal' and 'non-personal' data are therefore not as clear as they may seem, especially when issues such as anonymisation and pseudonymisation are added to the mix.

This is important when we come to discuss the potential value of personal data. While much data 'about me' may include data that could be personally identifiable, there is also much data about people and their behaviours which is statistical in nature (i.e. not identifiable), but which is the source of important insights and of great potential in helping improve peoples' lives.

Many complex issues are therefore raised by how personal data is currently being collected and used. These include whether individuals know about or understand what data is being collected and what it is being used for, whether they would be comfortable about this collection and use if they did know, whether such collection and use of data infringes individuals rights to 'privacy', and whether they are receiving a fair share of the financial and other benefits that their data helps generate.

Multiple solutions are being proposed. These include:

- Ensuring greater transparency
- Questions about 'who we trust'
- User education
- Calls for regulation to empower individuals in their dealings with organisations
- Calls for regulation to restrict organisations' ability to collect or use data or exercise 'data power'
- Proposals to redistribute power and control by, for example, providing individuals with personal data stores which enable them to collect and control their own data independently of the organisations they deal with

> "There is a need to find a balance between protection of personally sensitive data, and the value of sharing."
>
> Bangalore workshop

It's not surprising, then, that many workshops focused their attention on issues relating to personal data. We will return to them in detail in specific chapters, but these quotes provide a flavour.

What We Heard

There was broad agreement that issues around the control of personal data are increasingly part of the public debate. In Dakar, it was observed, *"whatever happens, people still need to be at the centre of the system, not the machines. This will be difficult, because artificial intelligence is becoming more and more dominant."*

As understanding grows, many in our workshops felt that we are witnessing a swing away from corporate power, back to the individual. In Singapore, there was recognition that there is a conflict between what consumers understand to be ownership, and what companies understand to be access, but that *"people are taking data back – there may be a shift in power to control by the individual."* This sentiment was supported in Johannesburg, but with the proviso *"… it will depend on where ownership comes to rest."* In Tokyo, the view was that *"data will increasingly be owned by individuals and not by the government or corporates."* On the other hand, some felt that the whole issue is a bit of a red herring. In a student workshop in Pretoria, they proposed that *"no one should own data."*

Discussions around personal data highlighted a number of cultural differences. For example, in Europe, where privacy is held in high esteem, the view from London was that *"privacy is real – individually and nationally. We need a lack of compromise on this."* However, in Tokyo, the view was that *"most people don't really care about privacy – despite what the experts think."*

When it comes to the consideration of the value of data, the view in Bangalore was that *"there will be growing awareness of the value of personal data, and this will empower individuals…. But the appetite for monetisation will lead to more collaboration. There is a need to find a balance between protection of personally sensitive data, and the value of sharing."* In Copenhagen, they felt *"we have a willingness to sell data too cheap – it is a trade-off."*

3.2 Ownership and Value

Many link ownership with the right to extract value from data. But traditional notions of ownership don't apply, so new models are sought and tested.

In the discussions, there was a strong desire for clear rules and frameworks to establish who is the rightful owner of what data; the common assumption being that once ownership becomes clear, so do the related rights, benefits, responsibilities, and so on.

In some cases, 'ownership' of data is obvious: for example, data generated by an organisation in its internal processes is 'owned' by that organisation. However, generally speaking, data doesn't 'work' in the same way as traditional tangible forms of private property. Very often it is co-created by two or more parties via transactions, interactions, and communications, thereby creating two or more potential 'owners'. Because data can be used without being 'used up', the same data can potentially be re-used by many parties for many different purposes. Data can also be replicated many times over for close to zero cost, which makes it economically limiting, or simply very difficult to enforce traditional proprietary restrictions on the uses of data.

"Data is not created by an individual, it's a joint effort; but it's not realistic to think that ownership is the proper debate to be having. There are multiple owners of data: think of bank transactions...Ownership is an inaccurate term; it's too loose to frame the question."

Bangalore workshop

Rights and Responsibilities

These complexities are driving the search for alternative ways of framing the debate by, for example, focusing on questions of rights of access and use, and on custodianship rather than 'ownership' per se. The workshops identified and distinguished the role of multiple actors in the supply of data: originators, custodians, processors, and users. A great deal of the discussion focused on defining the rights, responsibilities, obligations, and opportunities for each of these roles. The issues and dilemmas are particularly acute when discussing personal data, where, aside from complexities arising from data co-creation, issues of human rights often overlap and/or clash with narrow, legal notions of private property. This debate is also becoming increasingly important with the Internet of Things, where multiple parties, such as device manufacturers, device users, and devices themselves, all play a part in generating data.

Distributing Value

Many of the liveliest debates in several workshops concerned the distribution of value among these actors. Separating the 'ownership' and use of data by other parties was a recurring theme.

As a result, the emerging concept of data custodians was discussed at some length. It was suggested that 'data custodians' could have twin roles for which they would be rewarded: keeping data stores and sources secure (similar to a safe deposit box in a bank vault); and access and pricing control (similar to a literary agent). Some argued that the originator and custodian should essentially be the same actor, where all the data is both controlled and owned by the originator; others felt that the role is better suited to that of an intermediary or independent platform.

Managing Value

Although data manager business models are still emerging, the idea that some of us will gradually be willing to pay for our personal data to be looked after, shared against agreed preferences, and where appropriate, monetised, was often discussed. Whether there is a standard approach or whether there are different platforms with varied models for different sectors, cultures, and types of data, are as yet open questions. Many believed that if our personal data is worth something, then we should be able to see this, benefit from this, control it more effectively - and so also choose who else can access and gain from it.

> "The value of data is very regional, and is largely focused on who benefits from it as much as who owns it."
>
> San Francisco workshop

Several best practices for operating approaches and processes for data owners and custodians were also introduced into the discussions. These focused on areas such as payment for access to the data, and how ownership rights are transferred among the various stakeholders. Each of these models is different to those of today, where most of this activity is done by the processor.

Problems and Dilemmas:

- Is 'ownership' a useful/practical concept when it comes to certain types of data such as personal data?
- If not, what alternative concepts can we use to replace it?
- What other ways can we use to allocate rights, benefits, and responsibilities relating to data across stakeholders, including governments, technology companies, multinational corporations and individuals?
- In what circumstances does 'ownership' remain a valid notion?

What We Heard

In Frankfurt, the view was that in order to understand the value of our personal data, there must be a *"shift from a world where we have unclear views on data, lots of confusion, panic, and uncertainty, and no real alternative options for what to do with our data than what is provided by a few tech giants, towards a world with universal clarity of data value, ownership, and rights."*

Distributing Value

Type "who owns your data" into Google and you'll get dozens of interesting papers and articles – all with different opinions. But does it really matter? Many in our workshops thought not, and agreed with this perspective from Bangalore; *"data is not created by an individual, it's a joint effort; but it's not realistic to think that ownership is the proper debate to be having. There are multiple owners of data: think of bank transactions. Individuals interact with banks, creating at least a two-way process. Ownership is an inaccurate term; it's too loose to frame the question."* One way that this could be addressed is that individuals retain full ownership of their personal data in machine-readable format, but outsource its management and distribution to professional custodians, curators, or data brokers.

Managing Value

One way to manage the value of data is through personal data stores. These could allow individuals greater transparency on just how their data is being used. Essentially, this is a *"central repository for personal data, where individuals can access and control the access of others to their data."*[5] The creation of a new profession, privacy agents or data brokers, was also explored. In London, they were compared to the role played by asset managers, where *"in the main, we trust others to do it on our behalf – and can choose how (e.g. active, passive, ethical). The same may emerge in this space by trusted third parties (TTP), making it easy for the customer."*

Participants in our Kenya workshop built on the idea. In Nairobi, it was suggested that if there were a central repository for data, *"…allowing business and government to access personal information, but individuals to maintain control of their data and benefit from it,"* then *"… there will be wider access to information, without jeopardising personal privacy."*

Ownership to Custodian

There was general agreement that we will have to move on from 'ownership' to 'custodianship' within a decade. In Bogota, the suggestion was, although *"those who own data will continue to exploit its value…more data will be used for public benefit."* In Washington DC, they suggested that it would lead to *"better use of data from larger and more aggregated data sets"* that can have greater impact. Finally, in Sydney, it was suggested that we may well see more collaborative use with *"data being used to optimise social good – "data commons for social good," for example, focused on fewer car accidents, less teenage suicide, the ability to crowdsource health solutions, enhanced social belonging, more inclusive/less isolation and marginalisation – so data can make life better."*

This means there is a need for greater transparency, more information, better action, and a more widely shared informed view on data ownership and its implications. In a culture where everyone starts with trust as a default, the Danish view was that *"we can move on to community ownership of data – via cooperatives within society – that then provide the trusted platforms that can scale into broader ecosystems."* In San Francisco, a reflection was that *"the value of data is very regional, and is largely focused on who benefits from it, as much as who owns it."*

3.3 Power and Influence

Data is a means of exercising power, as well as a focus for multiple struggles for power. Regulation focuses on rebalancing influence between companies, government, and society.

Workshop participants around the world were acutely aware that with data comes power; that the more data an organisation can collect, use, or control, the more power it has at its disposal. This power can come in many forms. It could be the power to make decisions that affect peoples' lives by, for example, giving or withholding their access to services. Some organisations' use of data gives them the power to act as 'choice architects', deciding what information is to be presented to people and how. Concentrations of data can create concentrations of economic power, which in turn could affect the distribution of available benefits.

Given the many and varied ways in which data is collected and used by all the different parties, we found scope for multiple different power relationships, for example, between:

- Policy makers/regulators and large data-driven companies;
- Governments and their citizens;
- Companies and their customers;
- Different/overlapping political jurisdiction

> "When companies mess with complexity too great to monitor or understand, disclosure becomes an empty gesture."
>
> London workshop

There were also many different suggested ways of addressing unhealthy imbalances of power. The following generated particular interest:

Transparency: Many workshop participants were particularly concerned by what they saw as the unaccountable power of proprietary algorithms that are effectively immune from scrutiny, and give the organisations which develop them huge influence. The lack of transparency makes it almost impossible for anyone else to understand the economic, political, and cultural agendas behind their creation.

Accountability: There was also much concern about the ability of search engines and social networks to influence the information individuals are presented with. The power to include, exclude, and order the presentation of information, allows these companies to ensure that certain public impressions become permanent, while others disappear. Without knowing what a search engine actually does when it ranks sites, we cannot assess when it is acting in good faith to help users, and when it is biasing results to favour its own commercial, cultural, or political interests.

Ways of rebalancing power: Debate focused particularly on whether global technology companies have accrued too much power. Questions were asked as to whether they exercise this power responsibly, and what (if any) safeguards, regulations, and reforms are needed to create a healthier, fairer, safer, more innovative or resilient data ecosystem. Some workshop participants felt that the activities of those wielding disproportionate data power should be restricted by increased regulation. Others sought more radical responses by dispersing power more equally (via competition rules and anti-trust legislation, for example).

Problems and Dilemmas:

- Organisations collecting and using large quantities of data can generate significant value for individuals, society, the economy, and for themselves. At the same time, however, they may create excessive concentrations of power, and/or use the power they do have unfairly or inappropriately. How should these dangers best be addressed? By who?
- Moreover, by what criteria should we judge whether an organisation has accrued too much power, or is using this power unfairly or inappropriately? Who should be responsible for making such judgements?
- if a corporate entity is deemed to have too much power or to be exercising its power irresponsibly, what are the appropriate mechanisms for effective action?
- How should these decisions be implemented and enforced?
- How can/should disputes between different entities and jurisdictions (local, regional, global) relating to the collection and use of data be handled?

> “Whatever happens, people still need to be at the centre of the system, not the machines. This will be difficult, because artificial intelligence is becoming more and more dominant.”
>
> Dakar workshop.”

What We Heard

Questions relating to the exercise of power cropped up in most of our discussions. To provide a flavour of the discussions, we provide some examples here.

There is a growing sense that some companies are benefitting disproportionately from the collection, use, and frequently the sale of personal information. The Bangalore workshop pointed this out by saying, *“the consumers’ rights are always fringe; they don’t have the power of the likes of Google or Facebook.”* This is driving a public desire to give individuals greater control over their data. It was recognised, however, that doing this could create a new dilemma; how to maintain control of our data without losing the benefits and conveniences that exchanging personal information for digital services undoubtedly provides.

Transparency: We heard many calls for more effective legislative frameworks to help shape the emerging data economy in a more equitable way, to increase transparency, and make technology companies more accountable. Many in Africa and Asia, inspired by the EU’s stance on GDPR, were keen take up the challenge. In Mexico City, the view was that *“the biggest change will be in the way governments control data.”*

> “No one has yet worked out the extent to which patient data can compromise government security.”
>
> Singapore workshop

In Dakar, it was observed, *"as the power of data increases, it can be used to warp our sense of reality. Fake news is only an early sign of things to come…"* Across our workshops there were multiple calls for the need for greater digital literacy, so that individuals can choose what products and services they use, and have better control over their own personal data. Many argued for greater transparency and intelligibility around the use of data. They pointed out that if it is too difficult to understand what is being done with our data, it is impossible for individuals (or organisations) to have an equal relationship with the companies that exploit it. Some suggested that increased transparency would go a long way to addressing this, but it is not a solution on its own. One comment made in London was that *"when companies mess with complexity too great to monitor or understand, disclosure becomes an empty gesture."* For the power of data to be more equally spread, there needs to be greater public understanding about how data is being used. Some in London even suggested that transactions that *"are too complex to explain to outsiders, may well be too complex to be allowed to exist."*

Accountability: Across Africa and India, there was a strong sense of frustration about the dominance of primarily Silicon Valley American companies. Many saw this as a new form of colonialism, with personal data becoming the latest raw material exploited by the west. Participants in Singapore and Australia felt that managing the flow of national data was an issue of national security. In a workshop in Singapore, specifically focussed on patient data, we were told that the law restricts the sharing of health data beyond national boundaries because *"no one has yet worked out the extent to which patient data can compromise government security"*

In Bangalore, participants felt that the lack of transparency about how data is used and manipulated has led to a growing *"digital gap, both at country level and also for individuals."* This was also echoed in Madrid, where it was felt that this data divide will continue to grow, and will *"continue to be dominated by issues around transparency, ubiquity, and control."* Others reiterated the need for greater transparency about how data is managed and shared, in order to allow individuals to have greater control of their data.

Regulation: A number of mechanisms to ensure a more even distribution of power were discussed. This included greater interoperability and portability (spreading access), and the possibility of breaking up those organisations which have themselves become monopolies. In Bogota, it was suggested that public private partnerships could be the best way to create and implement better governance. Many advocated the establishment of a "Global Data Vision"[6], and a global body to develop and oversee the implementation of regulation. Sounds great - but when pressed, no one was really able to suggest how this should operate in practice, and where the ultimate responsibility should lie.

Finally, in Asia and the US in particular, we had conversations around geopolitics and how different ideologies might influence the use of data. In Hong Kong, the question was asked, *"what would be the implication of China winning the debate around data, and what would happen if it exports its values around the world?"* In Washington DC, the comment was, *"if you see this as competing modes, then it matters, because as China grows, more people/nations will try to emulate it."* Prosaically in Dakar, the view was, *"we don't mind if it's noodles in the morning or burgers in the afternoon; we need to create our own solutions."*

3.4 Global vs Regional vs Local

While many support further globalisation of data, others seek to assert stronger regional and national control to protect citizens and strengthen economies.

In many circles, there is a strong assumption that global 'Big Tech' firms can and will continue 'doing what they like'. But there is powerful sentiment, especially in fast-growing regions such as Africa and India, that governments should assert more control over data, to protect citizens' rights, develop the economy, and maintain a sense of cultural identity. This is creating potential conflict with those seeing global data flows as key to economic growth.

If the world was ruled by a single authority making wise, legitimate decisions and capable of implementing them efficiently and effectively, life would be simple. But it isn't. Instead, our reality is extremely complex. We are governed by a myriad of different authorities with overlapping jurisdictions and widely varying histories and culture, definitions of who 'we' are, interests, incentive and priorities, and powers. The overlapping nature of these jurisdictions means there is often confusion or conflict about who should have, or who has the right to deal with specific issues, so that multiple parties all feel they should be the ones in charge. While on the other hand, some issues fall between multiple stools with no one taking responsibility.

The data revolution is unfolding in this context. It is creating an urgency for new understandings, rules of conduct, and so on, but confusion as to who is best to lead in their creation; triggering 'turf wars' as different parties seek power and influence, creating new arenas and flashpoints of conflict as well as new requirements and opportunities

> "There needs to be a framework of common principles allowing public and private use of data across multiple jurisdictions. To achieve this, first there has to be global collaboration around a universally agreed set of standards."
>
> Hong Kong workshop.

Problems and Dilemmas:

- When is it necessary/desirable for data to flow across national borders?
- What different rules should be applied to different types of data (e.g. personal, non-personal), different circumstances and use cases?
- Which bodies, at what level (local, regional, global), are best placed to take the lead on this?
- How to ensure a) their legitimacy in the eyes of key stakeholders, and b) their effectiveness?
- How to address key stakeholders' concerns (e.g. the dangers of a new 'data imperialism', the risks that constrained data flows could undermine innovation and economic prosperity)?
- How can countries ensure that they benefit from the data they produce?
- Do new innovations around AI and Machine Learning need a different form of governance and regulatory approach?

What We Heard

In workshops around the world, we heard the same basic refrain. Data has thrown up many new issues, and policy makers and regulators need to catch up. We heard calls for more regulatory action wherever we went. Likewise, the need for greater collaboration and coordination between government and industry. But there was no clear consensus as to who should, or is best placed to, address these challenges, and at what level: 'local' (i.e. national), regional (e.g. EU), or via some global body?

Various solutions were explored. They fell broadly into three different options:

- Global regulatory body
- Regional regulatory bodies: America, the European Union and a China-centric Asia
- National regulation

In a world of multiple overlapping jurisdictions, a common feeling was that: first, the management of data throws up issues that are so universal in their significance, for example around privacy, ownership, ethics, and 'fair shares' of value, that common solutions need to be found; and second, that no existing organisation is currently able to take this role. As a result, many suggested that we need a higher-level body which could set things straight, for example in terms of creating an ethical framework to establish principles and practices common to all.

The idea first came up in Bangalore, which suggested that *"the creation of a World Data Council may well facilitate international negotiations."* Such a Council could help develop consensus around issues such as *"data sovereignty, and to negotiate cultural differences around privacy, for example."* Some drew comparisons to the efforts made around establishing a collective approach to climate change. In Hong Kong, the suggestion was that there should be *"a framework of common principles allowing public and private use of data across multiple jurisdictions. To achieve this, first there has to be global collaboration around a universally agreed set of standards."* Workshops in Jakarta, Bangkok, Singapore, Mexico, and London all called for *"an independent global data regulation framework (maybe like the G20)."*[7] In Dakar, the call was for *"governments and nations (and perhaps even organisations) to start thinking seriously about the construction of a Data Vision… a strategic template for the use of data and data-driven technologies."* Whichever the favoured approach, it was clear that there is a common appetite for a higher, independent authority to set the standards, define the common ground, and ensure balance and independence.

But who, or which organisations, will be trusted, and able to take the lead on this? While across the discussions, there was a universal desire for ‘someone else’ to come and sort out how to regulate data, many in our workshops were aware that global alignment may be too hard to achieve, not just because of the scale of the challenge and the agreements required, but also because of mistrust between some governments and multinational corporations. This was particularly evident across Africa, India, and in some parts of Asia, but was also recognised in mainland Europe.

The World Economic Forum is just one of several major organisations trying to develop an international, collaborative, global approach, however, few in our workshops felt it would be effective.[8] In Madrid, for example, opinion was that *“dominant Western services, built by Western engineers, reflecting Western values, and built on Western data, will increasingly be seen as either imperialist, irrelevant, or inappropriate in different cultural regions.”* Overcoming conflicting political imperatives and competing commercial interests will therefore remain extremely challenging.

Regional Regulation

A more practical option, perhaps, is a regional approach to data regulation. Regional bodies can deal with these complex issues more easily in a local cultural and political context. In Europe, the EU is already supporting new doctrines that are producing regional rules on privacy, data, and espionage. In Pretoria, it was suggested that a pan-African solution to data regulation could work; *“ideally this should emerge as a regional set of standards rather than just a local one, as this would both help to improve impact and prevent individual governments from increasingly using data regulation to drive top down state control of very powerful individual data sets.”*[9]

Many we spoke to are keen to learn from others. For example, participants in both Asia and Africa are watching the progress of the EU’s GDPR regulation with interest, and may well support similar measures. *“GDPR will change the data landscape in Nigeria, and bring in new standards”* It is not only Europe that is showing leadership here. China’s economic clout and growing influence across Asia and Africa may mean that there is a swing towards their walled garden strategy. It will be interesting to see which will ultimately dominate.

`

Again and again across Africa, we heard that *“the liberal economy or capitalist / Western society currently has a stranglehold on the poorest countries,”*[10] and that *“African data should stay on African servers.”*[11] The rationale behind this is so that local data can be more easily accessed and used to benefit the local economy, but also to prevent (largely US) multinationals from extracting the value of African data for themselves. Preserving cultural data was specifically prioritised in Kenya and Nigeria - *“cultural data is an asset store, and this should be licensed – it should be seen as intellectual property.”* [12] In Dakar, there was a call for *“data to be used in the national interest, not simply for the benefit of international companies.”* In a fast-growing continent, which has already had bitter experience of exploitation by the West, there is little appetite to allow data to become yet another resource which is extracted for another country’s profit.

National Regulation

The pros and cons of national regulation were widely discussed and often seen through the lenses of data sovereignty and data localisation, both of which restrict the flow of data across borders. Data sovereignty makes data subject to the laws and governance structures within the nation it is collected, and data localisation restricts data flows across borders by either mandating companies to keep data within a certain jurisdiction, or by imposing additional requirements before it can be transferred abroad. The objectives behind these restrictions can be diverse, and include privacy, cybersecurity, national security, public order, law enforcement, taxation, and industrial development, amongst others. Both approaches appeal to a growing sense of national identity, and support for them is gaining traction in a number of markets we visited, particularly in Africa and Asia.

In highly populated nations such as China and India, there was a view that confining access to national data will facilitate economic growth, build or protect political power, and increase local innovation. In Africa, this view was combined with a strong sense that there is a need to stop *"expatriate organisations grabbing the opportunity" and protect citizens from "data colonisation."*[13] Coincidently, in Europe, although there is a general desire for open data flows, there is also a sense that this has to be carefully balanced against the principle of privacy as a human right.

Proponents of cross-border data flows argue that local legislation undermines free trade by adding onerous and expensive obligations for businesses. These include building, operating, and maintaining data centres in multiple countries, as well as creating and updating separate data sets – even if they are a mirror of those held elsewhere. Add to that the inconvenience of having to go through a number of regulatory approvals to either operate in a market or comply with specific sector rules, and it's clear, they argue, that this restricts opportunity.[14] A 2016 report suggested that the effects of liberalising existing measures could add an estimated 8 billion euros per year to the European economy alone.[15] In emerging economies, some felt that the continued imposition of localisation measures will not only impact economic growth, but they will also have a negative impact on social development. In Dakar, it was observed that *"protectionism and boarded approaches to data could lead to a stifling of innovation, social uprising, mistrust in the potential for data to do good, suppression of whole segments of the world population, and large-scale state corruption."* Others pointed out that localisation potentially weakens national security – the more data centres there are, the more opportunities hackers have to target.

Keeping up with and capitalising on the growth and use of data will not be possible without the growing pains of adjusting regulation to account for this expansion. Looking ahead, it is clear that new techniques and legal constructs must be devised to ensure that we are able to extract value from data, while continuing to protect individuals' rights and acknowledging cultural differences. Quite how to achieve this in an effective and beneficial way is not quite so obvious.

3.5 Trust and Trustworthiness

Organisations seek to build trust in data use. This is increasingly about being more 'trustworthy', which is focused on being truthful and more transparent.

In the workshops around the world, there was a widespread sense that very few organisations, if any, can be trusted with data. Indeed, just as increasing levels of trust are needed, apart from some nations where trust in government remains high, the sense from most discussions was that levels of trust are in decline. The emerging challenge for organisations, policy makers, and regulators is, what does it take to demonstrate trustworthiness? On what basis can/should organisations be trusted with data?

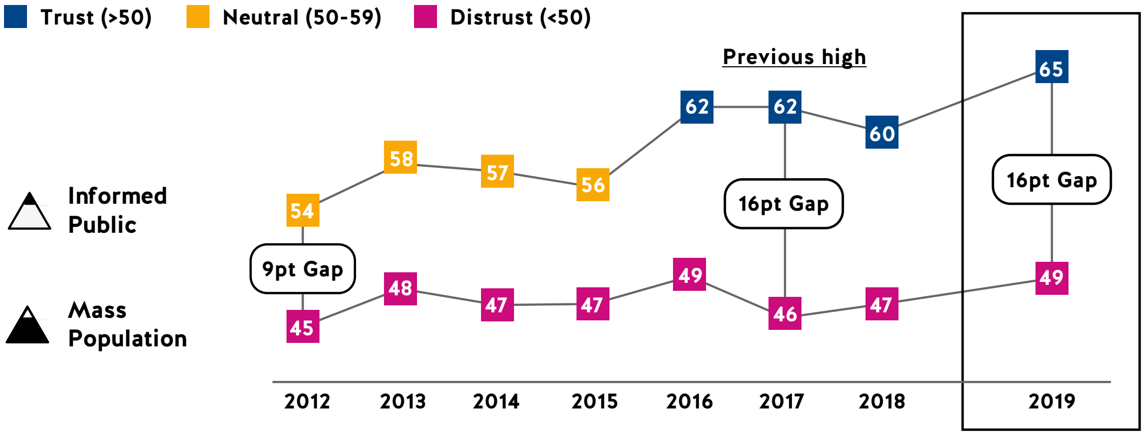
Context

Trust is an economically potent force. When people trust each other, the costs of doing business fall (as less time and effort is spent negotiating, manoeuvring, strategising, monitoring, policing, and enforcing), while opportunities open up, because people are more willing to work and co-operate with one another, including sharing data. Likewise, low trust environments tend to create high operating costs (because of all that time effort invested in negotiating, manoeuvring, strategising, monitoring, policing, and enforcing), while opportunities close down as fewer people are prepared to risk working with others, or for example, to share data with them.

> "As concern around security continues and the confidence of African developers increases, there is growing appetite for Ivorians to look after the data they produce and become less dependent on western (or other) nations."
>
> Abidjan workshop

Globally, our workshops took place at a highly particular time: the Cambridge Analytica scandal was unfolding with clear impacts on the degree to which users trusted, not only Facebook with their personal data, but also organisations more widely, as questions were raised about the tech sector as whole. As one student put it, Facebook, Amazon, and Uber are all *"brands that we trust less than we used to."*[16]

The 2019 Edelman report found there is a wide gap between the more trusting informed public and the far-more-sceptical mass population, marking a return to record highs of trust inequality. The phenomenon fuelling this divide was a pronounced rise in trust among the informed public. Markets such as the U.S., UK, Canada, South Korea and Hong Kong saw trust gains of 12 points or more among the informed public. In 18 markets, there is now a double-digit trust gap between the informed public and the mass population



Edelman Trust Barometer (2019)

The Trust Divide: There is a 16-point gap between the more trusting informed public and the far more sceptical mass population, marking a return to record highs of trust inequality

This specific context added another layer of controversy to an issue which is already extremely complex. When it comes to trust, there are many dimensions to consider, such as:

- **Trust in who?** Are we talking about trusting big businesses, small businesses, national governments, supra-national organisations, or citizens? Each of these has different relationships with each other. Whether or not customers trust companies, or citizens trust governments may throw up very different issues and dynamics to regulators trusting/not trusting global companies, or global companies trusting/not trusting politicians.
- **Trust to do what?** We may have 100% trust in someone's capability and competence, but 0% in their motives, or vice versa. There may be multiple boundaries, where we trust a party within a certain range of constraints, but not beyond them.

Within this context, what it takes to earn and keep trust can differ greatly from situation to situation. Further complications arise from the dynamics of how trust works.

One of these complications is the relationship between trust and transparency. If one party isn't aware of another party's actions, their trust levels may be high, but misplaced. In such cases of 'ignorance is bliss', trust levels can fall precipitately as people are shocked to discover the truth. A climate of mistrust and suspicion can then set in, as the pendulum swings the other way, so that even good, trustworthy actors are not given the benefit of the doubt.

A common, but mistaken, assumption is that changing levels of trust translates directly into changing degrees of behaviour - for example, willingness to share information. However, multiple factors can intervene to break this connection. For example, one party might not trust another, but still feel they have to share information, because otherwise they would forfeit access to a service. In such circumstances, actors that are not trusted (and who may indeed be untrustworthy) are not directly 'punished' for not being trusted. The proliferation of new technologies such as the Internet of Things (IOT) and Artificial Intelligence (AI), may mean that for simple operational reasons, whether they like it or not, citizens will be obliged to 'trust' more.

> "We need to recognise that data is not truth; it just presents information in different ways. We must learn to recognise bias or lose our freedom of choice."
>
> Madrid workshop

Sometimes trust is bilateral: it's all to do with whether Party A trusts Party B to do something specific. But sometimes it's general: for example, a sense that 'no one out there can be trusted'. These different dynamics generate different behaviours. Levels of bilateral trust can influence whether and how two parties deal with each other. A general sense that 'no one can be trusted' is more likely to increase pressure for 'system-wide' political or regulatory interventions.

Issues and questions such as these came up time and time again in our workshops around the world. For example, there was widespread suspicion of the motives of some Big Tech companies and their desires to monetise data (Edelman's 2019 Trust Barometer shows that more than 60 percent of respondents, globally, believe "tech companies have too much power and won't prioritise our welfare over their profits").

There were also strong differences of opinion as to who is trustworthy: some cultures trust 'government', but not 'big business'; in other cultures, most noticeably in the US, it is the opposite. This is in stark contrast to attitudes in some parts of Asia, particularly Japan and Singapore, where there is confidence that the majority of government operates in the best interest of its citizens - but less confidence in business to behave in a similar way. The same is true in Canada and Scandinavia. Across Africa there was widespread acceptance that corruption is rife – both in government and in the private sector; trust there is effectively absent. (One issue this throws up, as we'll see later, is that in Africa, it's common for many individuals to lie when asked for data, creating a significant knock-on effect relating to the trustworthiness of data that is collected.)

What We Heard

Although there was widespread excitement about the way data is transforming society, and recognition of the multiple benefits this brings, some in our workshops expressed caution. There were fears that the mere fact that data is becoming so ubiquitous means that we will trust it too much and fail to question its accuracy or its provenance. *"We need to recognise that data is not truth; it just presents information in different ways. We must learn to recognise bias or lose our freedom of choice."* This was the view in Madrid, where participants argued that the issue is, in a way, "over-trust," as there is a growing disconnect between our dependence on data to manage our lives and our understanding of the ways it can be interpreted. They suggested that the public risks becoming increasingly vulnerable to exploitation both by political and commercial actors; *"increasingly data will be used to control emotions, particularly amongst the young and the susceptible. Brands and governments will be keen to exploit this, to exercise new ways of influencing consumers."*[17]

"Low levels of trust in government, institutions, and Big Tech, devalues data by making databases unreliable. Citizens are choosing not to share accurate information."

Washing DC workshop

This is all well and good if organisations behave responsibly, but if there is a *"trust Chernobyl"* [18] trust between is broken , the consequence may well mean that people are no longer prepared to share their personal data and are less likely to believe the information they receive from government or other organisations. *"It will be interesting to see to what extent we allow our intimacy to be breached (health, financial, personal information)."* [19]

In general, our conversations around trust were divided in two ways; trust in the management and control of data, and trust in the accuracy of data.

Who can we Trust?

In Madrid, it was observed that our increasing familiarity with technology and growing confidence in our ability to access data is re-shaping how we trust – rather than refer to an expert, for example, we use crowd-sourced data to make a broad range of decisions, from where to eat, to treatment recommendations. At the same time, the popularity of social networks has changed who we trust. *"We have seen the transition of power from nations to corporates, and now it is from corporates to the people."* Certainly, throughout our conversations there was a sense that trust has shifted to greater confidence in peer groups or communities, rather than in traditional institutions or in those of a supposedly superior status. Many who are searching for reliable alternatives to traditional trusted sources of news and information are going online to use social media and a network of "friends" or opinion-sharing communities to find what they believe to be true.

Cultural differences are also important when considering who to trust. In Abidjan, lack of trust in the intentions of Western organisations is galvanising support for the Communauté Economique des États de l'Afrique de l'Ouest (CEDEAO). This has coincided with increased confidence in the ability of African technology skills, *"As concern around security continues and confidence of African developers increases, there is growing appetite for Ivorians to look after the data they produce and become less dependent on Western (or other) nations."*

> "Big data and AI provide a huge opportunity for intended and unintended discrimination."
>
> Bangalore workshop

Inaccurate Data

Many workshops felt that trust in data that is publicly available and free to use is declining, because it is increasingly difficult to discern if the information that we are presented with is, in fact, accurate. This is true both for government data and also for information received on social media. There was acknowledgement that distinguishing truth on social media channels is particularly challenging, as it is often difficult to identify the original source for a post or news item. Given this, the recommendation from Hong Kong was that citizens need to become more adept at understanding what is factual and what is not; *"there is a need to recognise that data is not truth, it just presents information in different ways and we must learn to recognise the bias, or lose our freedom of choice."* Failure to ensure citizens have the sufficient skills to distinguish fact from fiction has the potential to lead to a breakdown in trust, and could potentially lead to disturbance and even civil unrest; *"there is a feedback loop – fake data leads to low trust leads to fake data. There are diminishing returns, and trust needs to be maintained in order to ensure a safe and successful society."*

The potential negative feedback between lack of trust in government and government's subsequent ability to provide trustworthy data was highlighted in a Washington DC discussion of people deliberately providing false information. The example given was about research into US Census data, which suggests that around 20% of the information given is false, because citizens do not trust government not to use the data against them. A comment was: *"low levels of trust in government, institutions, and Big Tech, devalues data by making databases unreliable. Citizens are choosing not to share accurate information."*

We heard the same in Lagos, where we were told that such is the level of distrust in both the national government and the private sector, that citizens are unwilling to share their personal data with anyone – this in turn renders government statistics so inaccurate that they are rendered almost useless for meaningful analysis. One suggested solution to this was to implement robust regulation around the collection and use of public data. *"Improved data policies will improve trust in government – currently there is limited trust, because there is limited accountability."* [20] However, certainly in Nigeria, there was little hope that this could be implemented any time soon.

Concern was also expressed on growing reliance on AI, especially relating to the delivery of government services. Workshop participants were particularly concerned about programmers' ability to exclude bias in the selection of data used to train AI, or indeed identify it quickly should it occur. In Bangalore for example, it was felt that *"Big data and AI provide a huge opportunity for intended and unintended discrimination."* In Johannesburg the view was that if pubic concerns around data bias grows, there is a chance that they will no longer trust the products and services that are delivered, and certainly would not wish to participate in sharing their personal data. To address this, it was suggested that data should be labelled with, *"data dignity metrics," which could be used to measure and monitor the use of data for the common good, while maintaining the "dignity" (appropriate levels of privacy, for example) of individuals."*

Irresponsible Use of Data

The main actors in the data-driven economy, large tech firms and governments, were both widely criticised in our workshops. Time and again we heard discussions on the way that the many technology firms, particularly social media companies, exploit the data that we share, with little regard to personal safety or privacy. Few believed that lessons had been learned from the Cambridge Analytica scandal and that in the future we could be more confident in the organisations which have control over our personal data. In Hong Kong, it was observed that *"as understanding of the current Big Tech companies grows, expect more disagreement about their current business models."* In Bogota, they said, *"manipulation of the people will continue."*[21]

In addition, there was clear frustration with what was seen as a lack of leadership within the technology sector. In London, the perception was that it is this that has generated the real crisis in trust; *"it's not a crisis of trust – more a crisis of leadership. We can't impose trust downwards."*[22] The conversation went on to focus on the importance of trustworthy behaviour – and the need to make it accountable, *"…it's about confirmation, not trust."* Similar views were expressed in Singapore and Toronto.

Sometimes we heard debates about national security and the need to protect citizens from bad actors. This mistrust can seep into many, perhaps unexpected, areas. For example, participants in Singapore and South Africa both stated that one reason why DNA data is not shared with the US, is national security.

A number of different alternatives were identified, which could help rebuild trust in the use of data and data organisations. These are some of the solutions:

- **Greater transparency.** In Dakar, it was agreed that the public revelations around data lapses and the exploitation of personal data by some technology companies, have demonstrated a failure of self-regulation. The consequence of this is that *"tech companies will be obliged to be transparent about the data they collect, and the uses they make of it. This will be driven by increasing consumer pressure, and a competitive environment in which transparency and responsible data use become a point of differentiation."* Others agreed; from Madrid to Hong Kong, Singapore to Bogota, it was felt that social media companies in particular, should be more proactive in helping to distinguish between truth and inaccuracies on their platforms. There were numerous examples about how misinformation has influenced behaviours in both rich and poor countries, including overt bias in elections and online scams.

Full transparency may not, however, be a silver bullet; too much information can also be confusing. In London, it was observed that *"full transparency is only really needed if trust is absent. It certainly does not mean a requirement to share mountains of information as a means to ensure ethical behaviour."* In Frankfurt, the view was that as private citizens become aware of just how much personal data about them is being accumulated and traded, the demand for greater transparency will grow, and regulation will likely follow; *"if there is no transparency, it will block acceptance of online services."*

- **More Accountability:** There was universal consensus that greater accountability could increase trust, but there were differences in opinion about how this could be achieved. Ensuring that government data is accurate was of particular importance in the Washington DC, Tokyo, Singapore, Lagos, and Copenhagen workshops. In Lagos, the view was that the only way to achieve this is through open multi-party collaboration. *"Improved data policies agreed by multiple stakeholders will improve trust in government - currently there is little trust in government use of data, because there is limited accountability."* A suggestion from Denmark was that there would be greater public confidence in public institutions if there was *"Data NATO or a UN organisation, which could develop and oversee guidelines, codes of conduct, and shared standards."*

- **Technical solutions:** Some in the workshops suggested that new technologies such as blockchain may go part of the way to providing a reliable safeguard against abuse, and therefore help rebuild trust. In Tokyo, the view was that it *"spreads responsibility and increases trust in the system."* Creating a distributed, immutable record of information — which can never be deleted or modified — would at least provide a degree of transparency. Data could be recorded and distributed in a more transparent fashion, and could not be changed without amending all records across most users. Content creators could use distribution channels that guarantee that their content does not get altered, filtered, or blocked by a third party. Equally, a distribution channel leveraging blockchain could make it more difficult to censor and limit access to information.

> "Improved data policies agreed by multiple stakeholders will improve trust in government - currently there is little trust in government use of data, because there is limited accountability."
>
> Lagos workshop

- **Consumer influence:** In Bangalore, it was felt that regulation has been too slow to control the behaviour of some of the technology companies in their exploitation of data. Therefore, they suggested that public opinion is more likely to drive change in advance of any regulatory response to the decline in trust. *"The public response to unethical behaviour often happens before the law is enforced, or indeed appropriate regulation created,"* and from a personal data perspective, *"growing public understanding of potential harm to the individual will lead to increasing demands for better rights and greater accountability."* Those in Copenhagen built on this idea, suggesting that encouraging greater citizen involvement in monitoring the use and accuracy of data might help build trust. *"Is there, maybe, a role for something like Wikipedia in the mix here?"*[23]

- **Digital literacy:** Many felt that greater public education around the use of personal data would both help to build public trust in open data for public services, and give citizens sufficient skills to be able to identify when that trust could be misplaced. In Santiago, the hope was that recognition of this, alongside some hefty fines, would moderate corporate behaviour; *"when the public is more involved, accountability becomes "horizontal" rather than vertical."* As awareness grows, the ability to *"watch the watcher" and "critically understand"* will mean that large organisations of all kinds will temper their actions and take greater account of what is considered to be acceptable – both off and online.

- **Generational shift:** There was a recognition that trust in technology and the data that it delivers, is dependent on generational expectation. Some, for example, suggest that millennials are much more likely to be data-savvy around security and privacy and so on, than older generations, and may be less likely to be concerned about it. *"Gen X is the last pre-digital generation – the generation after will have better understanding of the importance of management and control."*[24] However, some fear that the next generation will be so dependent on technology, that "any data will be believed to be fact, and its veracity will not be questioned.[25] But we also heard the opposite view; *"in ten years' time, things will be more balanced. We are currently in a transitional phase and in a state of flux – people were scared of the car when it was first invented."*

Looking Forward

Throughout all our workshop discussions, it is clear that we are at a point of transition. Technology innovations, powered in the main by a select number of hugely powerful global organisations, several of which are not widely known, are triggering dramatic changes across all sectors of society, and influencing how millions of people live their lives. Often these changes are for the better, but not always. Such is the momentum, that many citizens feel that these changes are being 'done to them', whether they like it or not. This makes trust all the more important.

Building trust is not one, single challenge. It is multi-faceted. Data-based systems rely on the accuracy of the data that is fed into them. They only work effectively if enough people trust them to share accurate data, and believe in the accuracy of the information that they get in return. When incidents occur which reveal irregularities, corruption, or incompetence, trust is damaged, making individuals less confident about the benefits of participation. The risk is that growing numbers decide to scale back participation, or provide inaccurate data. If enough people do this, the system fails.

This aspect of trust is primarily technical - of creating systems that are fit for purpose. A second, more complicated and controversial dimension of trust relates to the motives and intentions of different stakeholders. The challenge for those organisations and institutions leading the transition to a data-driven economy and society, is to demonstrate that they are *trustworthy.*

Being trustworthy is not the same as being trusted. It means that organisations accept they should be held to account; that they demonstrate that they have 'good intentions'; that ethics are not something to talk about for PR purposes, but actually shape what decisions are made and how they are implemented. Greater transparency helps, but is not the only answer, particularly when trust in corporates is at a low ebb. A robust regulatory framework, either developed globally or regionally, would do much to create standards, along with checks and balances to curtail the power of the large corporates, which many we spoke to felt are still largely unaccountable. Individuals also have a role to play by becoming more aware of their rights and responsibilities online. If successful, and we create a lasting, robust, and trustworthy system, then the next generation can only benefit from its potent force.

3.6 Shared Language

People are unclear on where the value in data comes from or what form it takes. A key step is a common language about data that provides clarity of terms

Mounting discussion in the media and politics about data, its ownership, use, and its value, highlights a lack of consensus around how to describe fundamental concepts. In government, business, and civil society, this undermines the ability to build alignment and develop robust ways forward. A simple, shared, accessible terminology is increasingly being called for, in order to establish a common understanding of what the key issues are, and what options are available to address them. This lack of a common language and understanding is a major impediment to attempts to build cooperative or regulatory endeavours. Without it, the possibility of reaching an agreement or deciding on an appropriate course of action is limited, if not impossible. Given this, there was widespread consensus in our workshops that time and energy must be spent to define and agree terms around the use and value of data.

Problems and Dilemmas:

- Is it possible to create a 'common language' where, across the world, key stakeholders all use the same terms and definitions to describe what is happening with data?
- Is it possible to create a shared understanding of what the issues and options are, even if there are disagreements as to how important these issues are, or what the most desirable courses of action are?
- If it is not possible to create such a common language and shared understanding, how to advance debate and understanding of the multiple issues being raised by the emergence of a data-driven economy?
- If it is possible to create this common language and shared understanding, what is the best means of doing so, and who should lead/take responsibility for this quest?

What We Heard

Beyond the varied metaphors for data (sunshine in Tokyo, the periodic table in Singapore, religion in Madrid), myriad views on the definition of key issues, such as informed consent or digital literacy, were expressed everywhere. In the vast majority of workshops, the lack of agreement around precise, common terms for the key elements of the digital world was highlighted as a major concern. These were not just at a holistic cross-society and cross-industry level, but also within individual sectors. For example, our preceding 12 discussions on the future of patient data in 2017/18 highlighted how little is understood by professionals within healthcare on the differences between aggregated and anonymised data, ownership and control; machine learning and artificial intelligence (AI), and artificial general intelligence (AGI); as well as between data bias and data quality. Other sector-based discussions on automotive data in the UK, US, and Germany showed similar different interpretations.

In our workshops, examples such as these were all repeated in varied locations. Different definitions were used for data sovereignty and data localisation, between a data tax and digital taxation, and between data literacy and digital literacy – even by regulators. There was widespread acknowledgement of this and resounding support for the need to develop a global, cross-sector agreement for the terminology of data in multiple locations around, including Jakarta, Bangkok, Dakar, Mexico City, Toronto, and even Washington DC. Those in Singapore voiced the view of many, when they suggested that the rationale for this is to deliver *"a more clearly articulated government data strategy to enable community-driven initiatives which have wide public benefit."*

Language is not only about policy, however. It is about understanding. Without an agreed language around data use, it is difficult to see how populations can become digitally literate. Concerns about this sparked a total of nineteen separate discussions on Digital Literacy during the programme. Irrespective of geography, age, employment, or method, the message is clear; *"the divide between the technology literate and the technology illiterate will be a huge challenge, and will have grave consequences if not addressed."*[26] The reasons for this are not hard to uncover. As access to connectivity increases apace, and governments increasingly rely on data to connect with their citizens, managing cyber risks, ensuring individuals have the skills necessary to engage with the state, and building a workforce fit for a digital economy, are all priority areas. Failure to address digital literacy will have consequences, not least widening the digital divide, creating skills shortages, and extracting value from data. But, how will governments be able to extend a digital literacy programme if the lack of clarity around the language of data remains unresolved?

> "The divide between the technology literate and the technology illiterate will be a huge challenge, and will have grave consequences if not addressed."
>
> Tokyo workshop

4.0 Key Future Shifts

In addition to the cross-cutting themes, a number of future shifts were identified during our workshops. Their impact varies dependent on geography and sector; however, they were all considered of significance in multiple different discussions. These are reflected in the graphic below and were:

1. **Data and Digital Literacy:** An informed perspective around data, how it is acquired and used, increases public confidence, overcomes misunderstanding, and aids better decision-making.

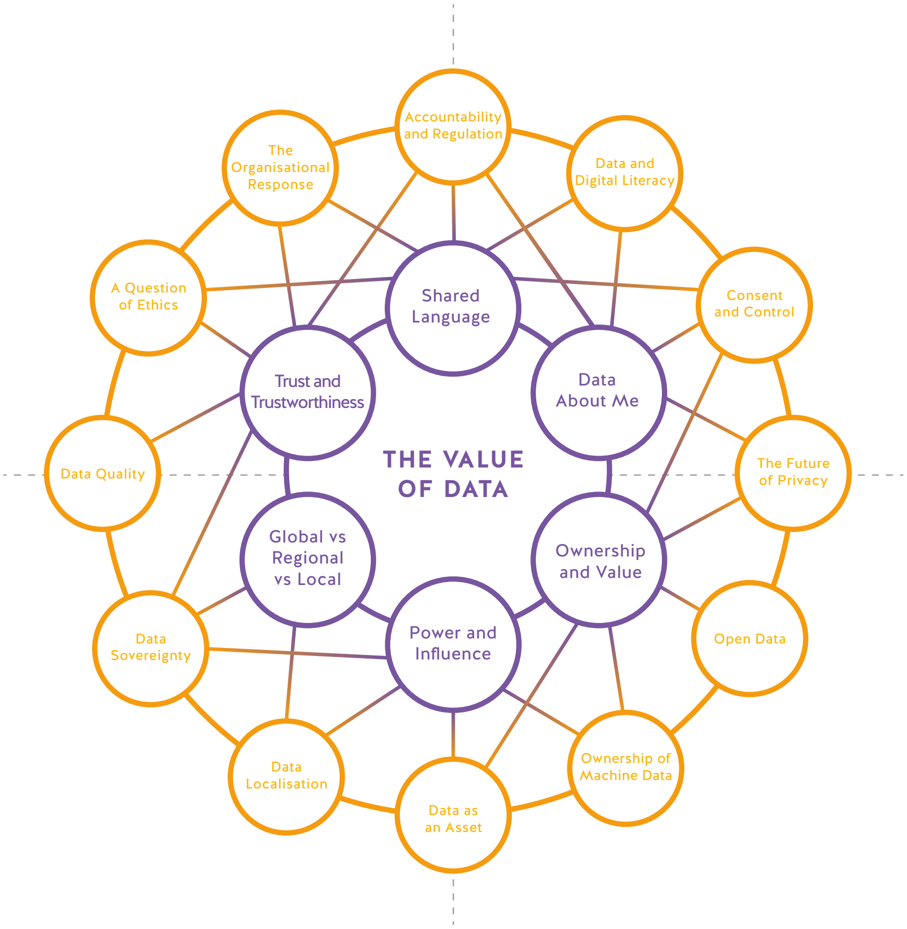
2. **The Future of Privacy:** There is a strong belief by some, in the right to data privacy. But others see that in an era of heightened security, this is a contradictory and outdated concept.

3. **Consent and Control:** Depending on informed consent as the basis for processing data is unworkable. Rethinking our view of what it is designed to achieve drives a new approach.

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 for good.

5. **Ownership of Machine Data:** Debates on who has what rights to what IoT data escalate. Questions around title, control, and usage of data lead to many sectors taking different views.

6. **Data as an Asset:** Organisations are obliged to account for what data they own or access. They are required to report their full data portfolio and are taxed on this.



7. Data Localisation: Nations see benefit in copies of all citizen and machine data in regional centres. Government and local companies seek access data held by foreign corporations.

8. Data Sovereignty: More governments see control of national data as a means to protect citizens' rights, develop the economy, and maintain a sense of cultural identity.

9. Data Quality: As we seek better insight, concern about biased, poor, and false data grows. Cleaning and validating data is a social, political, and commercial battleground.

10. A Question of Ethics: Ethical data use grows as a concern, but we struggle to agree a global approach. Sectors set their own standards and try to align on some common principles.

11. The Organisational Response: The management of data requires a 21st, not a 19th century approach to business. With digital becoming the norm we move on from principles based on physical products

12. Governance and Regulation: Rising concern about the use of data influences public opinion. Policy makers seek a more joined-up approach to regulation, governance, and accountability

4.1 Data and Digital Literacy

An informed perspective around data, and how it is acquired and used, increases public confidence, overcomes misunderstanding, and aids better decision-making.

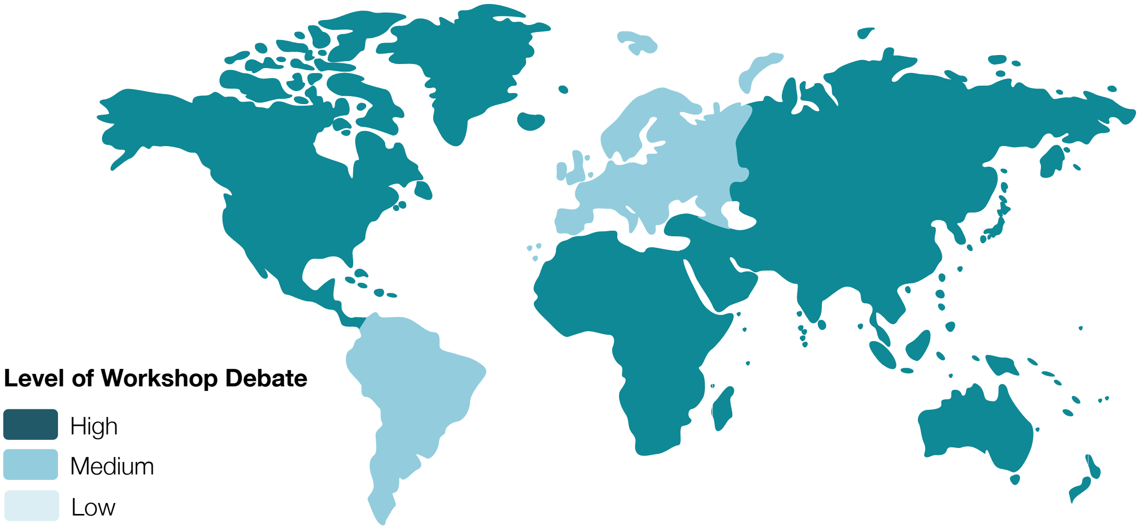
Context

At a time when a plethora of technologies are both augmenting and replacing human capabilities, many in our workshops believe there is a pressing need to ensure greater public, political, and organisational understanding of the value and use of data. Regulators need to be more informed; workers need better technical skills; and citizens need to be equipped to manage their digital footprints to better engage with public services and protect themselves from possible abuse. How to address this and counter what was seen to be an increasing digital divide, sparked nineteen separate discussions on Digital Literacy during the Future Value of Data project.

These discussions focussed on three different debates around data literacy:

- **Regulatory preparedness:** Is there sufficient understanding amongst policy makers to manage the transition to and the impact of digital technologies successfully? Can regulators better support digital literacy?

  In the main, it was acknowledged that regulation will probably always trail technology, and therefore in order to be as prepared for the expected transition to a more automated working environment, closer collaboration between business and policy makers is essential. In Lagos, Nairobi, and Bangkok in particular, there were concerns that, without greater technical understanding, policy makers will find it difficult to truly comprehend and manage the social and economic changes ahead. To address this one option, which was given widespread support, was the idea of greater collaboration between national regulators; many suggested that a global, or more likely, regional body could establish an education framework, set clear literacy standards, and share best practice.



Most in our workshops felt that greater understanding of the potential that data has to drive economic growth will shape what and how we learn. In London, it was observed that teaching basic logic and reasoning and providing a backbone for training computer-literate adults is already priority for a number of governments.[27] Indeed, a commonly held view is that, such is its significance, a basic understanding of coding will soon become a part of the core curriculum, like maths and languages. In Madrid, the recommendation was that alongside practical skills, better understanding of ethics, control, and privacy is also important. They observed that the millennial generation is likely to be the first to benefit from policy changes, and given this, we may face a generational divide, as there will be those who are unable to adjust to the changes that technology will bring. Governments will have to prepare for this.

- **Active workers:** Does our economy/society/workforce have the skills needed for a digital age? Do we need to train or retrain workers so they can actively participate in the digital economy?

Having and maintaining the right skills is critical to deal with technological change.[28] As technology is very adaptable, the ability of machines to see patterns and outperform humans at recognising images is expected to affect high and low skilled employees alike. As a result, the workers of tomorrow, including the most educated elite, may need to ensure the skills they learn complement those that are easily replicated by a machine, and remain flexible and open to learning new skills.[29] Many in our workshops felt that there is insufficient public awareness of how quickly change is coming upon us, and therefore little understanding of the new skills which will soon be needed.

A number of corporations already have their own learning platforms to keep staff up to date; IBM, for example, has an AI Academy that recommends courses from a curriculum provided by Coursera. However, some felt that, although useful, this form of “up-skilling” will merely increase the divide between those already in the professional elite, and those with fewer opportunities. The real need, they argued, is a “re-skilling” of the wider workforce. Lack of digital literacy may mean that unskilled workers may find themselves locked out of the workplace completely, with their roles performed more efficiently and cost effectively by machines. Given this, there was widespread support for corporates to get more actively involved in training programmes.

- **Informed citizens:** How best to ensure citizens can understand and manage the benefits and risks of using and sharing data? How can education help them to navigate the internet and digital platforms, and engage with social media?

In Madrid, Copenhagen, San Francisco, and Singapore, it was felt that the priority for any public digital literacy programme should not be about enabling individuals to master a particular skill or to become proficient in a certain technology platform, but rather it should be about equipping them to thrive in an increasingly digital society. Teaching citizens to manage their digital shadow, and helping them to better understand how to protect themselves from fraud, they argued, should be a national priority.

What We Heard

The Digital Divide

From Washington DC to Tokyo, Bangkok to Sydney, and Manila to Johannesburg, concerns about those who will not have access to digital education were raised. In Tokyo, the perspective was that *"the divide between the technology literate and the technology illiterate will be a huge challenge, and will have grave consequences if not addressed."*[30] Similarly in Washington DC, they said, *"in 10 years, society will be more digitally literate overall, but adoption will be lumpy – in part because of public appetite, and in part because of lack of opportunity. Consequently, the threat of increasing inequality remains a strong possibility."* [31] There is already a significant literacy gap to address. A number of countries we visited still cannot guarantee even a basic education for everyone. This was observed in Pretoria, where they pointed out that, although there is a huge need for digital literacy, the priority in some areas should be to begin with the roll-out and mainstreaming of Early Childhood Development programmes. Only once young people can read and write, can digital literacy be addressed; *"a computer is just a box if you don't know how to use it."* In India, it was observed that technology can also help to reach those who were previously cut off from education, and that more should be done to introduce mobile literacy programmes.

A Global Approach

Most agreed that there is a need to establish some common global standards; *"we need harmonised regulation,"* or, at least, best practice around data literacy, but there is little expectation that this will happen any time soon.[32] Some think time will sort this out. First articulated in Bangkok, but echoed in other markets, there was an assumption that *"we will eventually figure out the educational requirements necessary to deal with a data-driven world, and go on to build ethical education platforms which will be accessible to all."* Not everyone shared this view. Instead they argued that it will be difficult for citizens to truly understand how best to manage their personal data without a change in the way data is managed. They called for regulation to clarify how personal data is used.

Corporate (In)action

In Washington DC, the judgement was that *"we need to find ways to connect data literacy to people, in real terms. Business needs to understand this too, and Big Tech in particular may need to take some responsibility. Without a universal approach to this, there is a risk that inequality will increase."* In Copenhagen, they pointed out that *"there must be various entry points to digital education, both through schools and also available to those returning to the education system."* To address this, future policy should *"enable lifelong learning (covering more technical skills, interdisciplinary, improved research methodology, and better networks), and then fuller integration of digital cross-domain knowledge."* Failure to address the problem risks the damaging scenario of suffering higher unemployment and a skills shortage at the same time.

> "There must be various entry points to digital education, both through schools and also available to those returning to the education system."
>
> Copenhagen workshop

In Bogota, it was observed that, as jobs of the future are going to change, so too will our educational needs. Given technology will likely replace many of the traditional jobs, rather than focus on purely academic achievements, they recommended that there should also be a focus on the skills we will need to work in the future; *"the way that we educate our children will have to change to adapt to the needs of a more technical society where skills such as collaboration, and softer qualities such as integrity and compassion, not just better maths and coding skills, will have greater value. As yet, there is little understanding of this in the public sector, so it is difficult for regulators to develop appropriate policies that will offer long-term benefits."*

In Dakar, the outlook was optimistic. They felt that with the right kind of political support, investment in data literacy presents an opportunity for African economies to catch up with the likes of China, Korea, and Singapore, which have already had great success in data innovation. *"We must be ready to build a generation of digital culture. Our young people should start learning to code. They must learn to work digitally and more effectively."* This perspective was echoed in Lagos; such is the pressing need for development, they argued that the priority should be *"to teach Nigerians how to use, access, and navigate the Internet. Education about safety and security is less important."* Conversely, in Washington DC, there was concern that policy makers do not currently see digital literacy as a priority; *"support for greater digital literacy would benefit from a "moment" which demonstrates how it can be a vehicle for social change."*

The Generation Game

Looking ahead, some suggested that greater digital literacy will simply come with time. *"The next generation is inherently more sophisticated. They understand a data-driven society implicitly, and know how to protect themselves. Similarly, next generation policy makers will be more sophisticated."*[33] However, in Madrid, it was felt that, although technically able, young people may not have the emotional maturity to deal with the social implications of new technologies. To address this, they suggested that *"young people should have to prove their emotional maturity before being allowed to participate in social media sites."*[34] They argued that public education, therefore, should have a stronger emphasis on philosophy, critical theory, ethics, and anthropology, in order to provide students with the necessary skills to participate in a new social contract.

> "Support for greater digital literacy would benefit from a "moment" which demonstrates how it can be a vehicle for social change."
>
> Washington DC workshop

Truth and Illusion

Provenance and authenticity of data were major concerns in our discussions, and the debate on who has liability and is accountable for ensuring truth and accuracy was often raised. Some argued that it already threatens democratic values and confidence in government, and therefore there should be increased public awareness about it. Initiatives to address this include digital literacy programmes, the creation of safe spaces online, and controversially, as in Uganda, taxing social media use – although in the same Nairobi workshop, this was also described as a way to limit free speech.[36] The Madrid workshop proposed *"clearer labelling and better terms and conditions, to help people understand how their personal data is used and managed. We could even consider labelling content by using colour-coded schemes, as found in the food/energy sectors."* Those in Singapore agreed in principle with this, but pointed out that *"labelling helps to identify truth, and perhaps branded news is a way to help the public identify responsible channels. However, all of this is dependent on maintaining public trust in the established media."*

Awareness and understanding

The hope is that growing data literacy will mean greater public engagement online, which will in turn give citizens greater access to a range of public services, such as health and social care, education, and transport. In Santiago, it was also argued that higher transparency, greater accountability, and public awareness about the importance of data and government use of it will act as a way of monitoring corporate behaviour, particularly around the use of AI; *"when the public is more involved, accountability becomes "horizontal" rather than vertical."* As awareness grows, the ability to *"watch the watcher"* and *"critically understand"* will mean that large organisations of all kinds will be obliged to temper their actions and be more considerate of what is considered to be acceptable – both off and online.

Implications For Data Value

None of the issues highlighted by our research - the need for policy makers and regulators to better understand new technologies and their implications, for workers to improve their digital skills, and for citizens to better understand the potential consequences of how their data is collected and used - can be addressed by 'a quick fix'. They need time to develop and mature. But growing recognition of their importance represents a step forward. The triple agenda for improved digital literacy represents an important plan for action and improvement; necessary pre-requisites of a healthy data-driven economy - essential underpinnings of effective functioning - just as the '3Rs' became an essential underpinning of the industrial age.

> "When the public is more involved accountability becomes 'horizontal' rather than 'vertical'."
>
> Santiago workshop

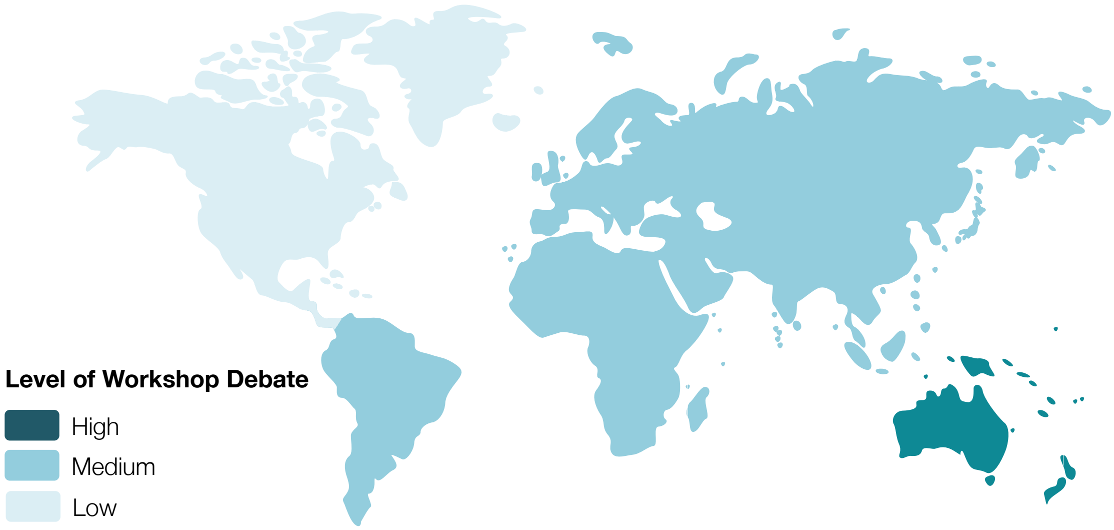
4.2 Culture, Governance and Privacy

Differences in culture and governance drive different attitudes towards privacy. Some believe in the right to data privacy; others see this is a contradictory and outdated concept.

Context

Greater availability and access to data is changing attitudes to data privacy and security. Our workshops revealed a diversity of opinion about this, depending on geography, culture, and age. There were a wide range of views about the definition of privacy itself. Is it about unedifying and unjustified snooping? Keeping potentially embarrassing information private? Threats to civil liberties? Risks arising from the ability to use private information to harm an individual? Our discussions divided broadly between those who felt that privacy is a hard-won human right and should be protected, and those who argued that, in our data driven world, guaranteeing privacy is impractical and may even compromise national security.

The European and international institutions such as the EU and the UN, as well as several governments, are firm believers in privacy as a human right. But not everyone agrees. Conversations in Abuja and Dakar, Tokyo, Jakarta, and Singapore, revealed an ambivalence about the issue. In both the US workshops, there was support for the “third-party doctrine,” which has long governed privacy law and holds that there can be no privacy expectation on data that is shared with a third party. In Shanghai, we were told that, although views are changing, privacy is not considered important in China; indeed, there is no direct Mandarin translation - the Chinese word for privacy, yinsi, is mainly associated with secrecy and poor mental health.



As with so many of our discussions, building consensus was complicated by a lack of clarity around language and what privacy really means in practice. The concept is abstract and touches multiple issues, including the implications for national security, the protection of minors, consideration around what are the legitimate boundaries over who has access to and benefits from data, and many highly specific areas about, for example, IOT data or facial recognition. Furthermore, generalisations are unhelpful because privacy is defined by its context. It does not mean absolute secrecy - we share sensitive information with doctors, friends, families - but when we reveal information in one situation, we trust that it won't surprise us in another.

To privacy advocates, there is a growing personalisation-privacy paradox: we want to have products and services that are customised to our needs and actions, but also want our data to be private, shared when we want and only to the actors we authorise for its use. Some people - those who are not privacy advocates - saw 'privacy' as an anachronism – an issue which has been overtaken by events and which maybe didn't matter very much in the first place. Others see it as pivotally important, defining the shape and future of the entire Internet age. Although recent data breaches and the consequent news headlines have raised public awareness around the issue, this has yet to significantly influence behaviour. So, policymakers are faced with a dilemma; should they legislate on the basis of how people actually behave online, apply a set of idealised archetypes, or suggest how they *ought* to behave? The view from our workshops was that, as understanding of just how much of our personal data is traded online increases, there will be greater clarity about what information people are prepared to share, and who to share it with, in exchange for better service or an improved quality of life.

To date, the primary focus for the privacy agenda has been around the exploitation of personal data – the collection, use, and value extraction of data by companies. However, the collection and use of data by governments is a growing issue, particularly as data-driven decision-making, including AI, is being more widely adopted. For governments, provided the right checks and balances are in place, there are huge benefits; it can help to address financial shortfalls and investment needs aimed at improved healthcare, transport systems, and public services, for example. Such is its transformation, some in our workshops argued, that democracies will not only have to collect data for the improvement of public services, they need it to remain competitive. If the West enacts too stringent privacy laws, it will have less data — a key raw material for artificial intelligence — and as a result, will put itself at a competitive disadvantage to the likes of China, where surveillance is becoming pervasive.

"People are prepared to exchange information about themselves for a better life. At worst, they are indifferent. As we share more data, in ten years' time, concerns about privacy will reduce still further."

Tokyo workshop

In some instances, differences in privacy laws are acting as an unintended trade barrier, and restricting innovation. The recent roll-out of GDPR across the EU was, in part, designed to address this. Compliance is not easy. However, it is clear that, for the first time, the hefty fines and associated publicity which is generated from a failure to comply, gives regulators sharper teeth than they have had in the past, and provides companies a compelling reason to assert more control over digital supply chains to better control data flows.[36] Many regulators are keen to learn from the successes and failures of GDPR, and are watching its roll-out with interest.

Generational Shift

Whatever the view today, attitudes to on-line privacy are changing, as the next generation, which has not known life before the internet, matures. This does not mean that we will find alignment. Again, we saw diversity in opinion about how this would play out, as everyone struggles to find a balance between privacy, convenience, and security. In London, it was suggested that, because of the compelling nature of new and enticing data services, there is a strong chance that privacy, as we know it, even in Europe, will no longer be an issue. The workshop in Johannesburg took the opposite approach, arguing that rising data literacy among both citizens and states will lead to greater understanding of the negative consequences of oversharing, and therefore sensitivities about privacy are likely to increase.[37] There was divergence as to how to manage this. Some see that technical solutions such as encryption will ensure that the right to privacy is maintained, but others advocated the need for more transparency so that individuals are more informed, and therefore better able to control how their data is used.

A Global Approach?

The big challenge ahead is whether or not privacy can be addressed via global agreements. There is general acceptance that there is a need for it. As different regions all seek to progress data regulation via the likes of APEC and GDPR, the emergence of a global privacy framework is championed by those looking for better control and greater transparency. The World Economic Forum is just one of several major organisations trying to develop an international, collaborative, global, approach.[38] Key focus areas are on delivering meaningful transparency, strengthening accountability, and empowering individuals. The inventor of the web, Sir Tim Berners Lee, is also working on the issue. He advocates a new “Contract for the Web,” which aims to protect people’s rights and freedoms. It states that governments must ensure that its citizens have access to all of the internet, all of the time, and that their privacy is respected, so they can be online “freely, safely, and without fear.” As Sir Tim himself observes, “no one group should do this alone, and all input will be appreciated.”

> “The massive increase in data will enable massive personalisation. There will be no privacy, because of the compelling nature of the services available.”
>
> London workshop

Inevitably, not all countries or even states are moving at the same speed and in the same direction, so it is likely that regional regulation will continue for some time. In America, for example, the U.S. Constitution does not contain any explicit protection of privacy, so the judiciary has been searching for ways of connecting existing constitutional protections with the privacy issues of the day, such as the Fourth Amendment's protection against unreasonable search and seizure. Despite calls from a range of CEOs for better policy legislation, the US at a federal level has lagged behind other regions. This might be addressed if other states follow the example set by the recent California Consumer Privacy Act (CCPA). However, the appetite for change may be low; privacy was not seen as a priority for discussion at either of our workshops in San Francisco and Washington DC. This is despite research from the likes of Pew suggesting that US citizens *do* care about privacy, but don't know how to address it.[39]

China and India, each of which have more people online than either Europe or America have citizens, have diverging and contradictory approaches to privacy. Interestingly, India, one of the world's most populous countries, has taken a somewhat contradictory approach to privacy legislation. It recently announced a draft data protection bill. Companies and the government must generally abide by legal principles similar to the EU, and as with GDPR, this law would apply to all entities, everywhere, that process Indians' data. At the same time, it is also supportive of data localisation, and mandates that Indians' data should remain within national boundaries. It has also proposed Chinese-style rules to extend the state's surveillance powers. In March 2019, the government put out a draft ecommerce policy, arguing that the personal data of Indians should be treated as a 'national' asset.[40]

In China, although the law did not even define what counts as personal information until 2018, there is increasing clarity around security obligations and responsibilities, due to public concern about the impact of data theft, and the ambition of Chinese companies such as TenCent and Alibaba to enter Western markets. This sits uncomfortably beside the government's appetite for surveillance, which has led to a tightening of data protection rules for companies, while making it easier for the state to capture more private information.

Given these complexities, it is unsurprising that some see that companies are using privacy issues for competitive advantage. Apple's 2019 marketing campaign launched at CED in Las Vegas, includes a major privacy pitch, "What happens on your iPhone, stays on your iPhone." Recently, Facebook promised that the content of all messages will be encrypted, regardless of the platform they are on.

> "Nigerians are not confident about privacy, which is why many protect themselves by having an online alias - this guards them from interest groups and government surveillance."
>
> Abuja workshop

What We Heard

Changing Attitudes to Privacy

Our workshops revealed that national attitudes towards privacy varied dependent on the levels of trust . In Tokyo, we were told that *"people are prepared to exchange information about themselves for a better life. At worst, they are indifferent. As we share more data, in ten years' time, concerns about privacy will reduce still further."* In Jakarta, they said, *"Indonesia is a very sharing country – across all cultures and all demographics, and also culturally, Indonesians are inclined to overshare."* In Africa, there was a similar response. In Dakar, for example, it was noted that *"in Europe, privacy is a big concern. There are historical reasons for this. We are a more open society.*" In contrast, in Lagos, we heard that *"Nigerians are not confident about privacy, which is why many protect themselves by having an online alias - this guards them from interest groups and government surveillance."*

Some suggest the concept of privacy is losing its appeal. In London, one suggestion was that that *"the massive increase in data will enable massive personalisation. There will be no privacy, because of the compelling nature of the services available without it."* It was also pointed out that accepting this will take time to become culturally acceptable; *"change will be slower than expected. We are high on the hype cycle for data. Some realism around its limitations will emerge."* In Manila, it was observed that this sort of behaviour by corporates and the very wealthy could *"lead to an economy of scarcity around data. How we manage privacy in the digital age, therefore, will be a key determinant of the future value of data."*

Whatever the view today, attitudes to on-line privacy are changing, as the next generation, which has not known life before the internet, matures. Again, we saw diversity in opinion about how this would play out. In London, it was suggested that privacy as we know it will no longer be an issue. Because of the compelling nature of the services provided, there is a strong chance that *"society will have ownership of everyone's data."* They disagreed in Bangalore, where it was said that *"privacy will become more of a public issue. There will be growing concern around state surveillance and how to minimise the harm of governments having access to "all" data."*

"How we manage privacy in the digital age will be a key determinant of the future value of data."

Manila workshop

Regulatory Choices

There are huge benefits of sharing data to improve the workings of financial shortfalls and investment needs aimed at transport systems and public services. But still, the danger of excessive surveillance is worrying for many. Although technology itself is agnostic, without the right checks and balances, it can still be used to cause harm. In Dakar, it was said, *"there should be clear rules on which data is collected and for which reasons. We need ways to protect vulnerable people."* For example, although law enforcement officials around the world can use AI to identify criminals, it can also mean that they (or others) are able to eavesdrop on ordinary citizens. Both the China and the US governments are introducing facial recognition to track their citizens. Some consider this to be a step too far.[41] Many argued that new, globally agreed principles will be needed to ensure consensus on what degree of monitoring is reasonable. In Jakarta, the suggestion was that *"if we have or hold data, we can't shy away from responsibility, but we need a globalisation of data framework."*

The big challenge ahead is whether or not privacy can be addressed via global agreements. There is general acceptance that there is a need for them. In London, the assessment was, *"today we have a patchwork of data privacy laws, but data flows globally. We will need to see global privacy principles."* As different regions all seek to progress data regulation, the emergence of a global privacy framework is championed by those looking for better control and greater transparency. In Bangalore, it was observed that *"the creation of a world data council may facilitate international negotiations. Currently, there is little consensus around data sovereignty – cultural differences around privacy, just one example."* But who, or which organisation, will be trusted, and able to take the lead on this? As attempts at Internet governance have shown, creating a supranational entity is challenging, owing to conflicting political imperatives and competing commercial interests.

Many within our workshops believe that GDPR has set the standard which others should follow.[42] In Mexico City, the view was that *"there are already some global standards, and some nations are already acting transnationally. GDPR is having impact beyond European boundaries."* In Nigeria, just one of many cases, it is seen that *"GDPR will change the data landscape and bring in new standards. It offers a template for localised legislation, and has highlighted some of the key issues around data that are not yet a priority in Nigeria, but will increase in significance over the next decade."*

> "Today we have a patchwork of data privacy laws, but data flows globally. We will need to see global privacy principles."
>
> London workshop

Across Australia, Asia, Africa, and South America, we consistently heard 'GDPR-lite' as the shorthand for what was needed locally as well as globally. Similarly, in Jakarta, the perspective was that *"there will be an Asian alternative to GDPR, driven by Asian ethics and principles."* These may, for example, be less focused on the individual. Across Africa, there was also interest in developing locally relevant regulation. In Lagos, a thought was that, with slow progress to date, moving ahead, *"the private sector will put pressure on government to ensure that there is clear legislation around accountability, and demand the creation of a Nigerian Data Protection Policy that reflects the same principles as those articulated in GDPR."*

Implications for Data Value

Global consensus on what are appropriate levels of privacy is still out of reach – and current views are often defined by culture. However, with common frameworks now being adapted and adopted for several different regions, the potential for some alignment is emerging. While several believe that privacy will not be an issue in the longer term, most agree that for the next decade, particularly for multinationals and many of the more democratic governments, it will continue to be a primary concern. With privacy also now being used as a source of competitive advantage, and used as a mechanism to build trust and credibility, several companies are trying to use it as a point of differentiation.[43]

"There will be an Asian alternative to GDPR, driven by Asian ethics and principles."

Jakarta workshop

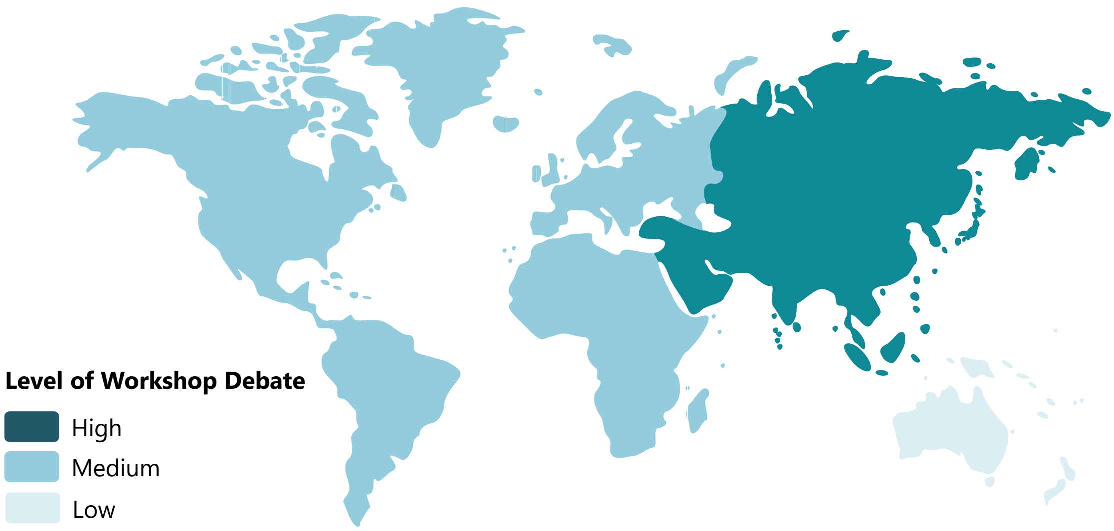
4.3 Consent and Control

Depending on informed consent, as the basis for processing data, is unworkable. Rethinking our view of what it is designed to achieve, drives a new approach.

Context

Collecting, sharing, and trading personal data is the bread and butter for many online companies, and constitutes an important source of revenue. The general public is only gradually becoming aware of this, and some are now beginning to question whether they are comfortable with this model, particularly in the light of revelations of the misuse of data which took place during our research period.

Our discussions confirmed that various dilemmas must be acknowledged in addressing informed consent. The first involves the evergreen tension between data collected for use in "marketing," and the data required for "operations." Some felt that only data which would be of benefit to the user should be collected and processed, whilst acknowledging that this would necessarily restrict the operations of the processor, and the ability to create value for themselves. The second dilemma is a recognition of the need to balance the demands for personalised products and services, with the necessity of data privacy. Given this, there was a strong view from our workshops that personal data should be considered to be a personal or corporate asset, and that as such, customers should have access to sufficient information so that they can make informed decisions about the extent to which they are prepared to trade it in return for products and services.



But quite how to do this is complicated. Too much information in the form of small print about terms and conditions may put people off – it's not only difficult for people to digest, but the amount of personal information currently being gathered can be shocking. Service providers therefore fear that revealing the full extent that data is collected and monetised might risk their current business models, as customers become unwilling to continue to share their data. Consequently, some in our workshops argued that, rather than grapple with how to deliver "informed consent," it would be more sensible to identify new ways in which individuals can maintain control over their data. This could include for example, more rigorous industry regulation, increased government regulation, or the adoption of intermediaries who can better represent consumer needs and control access to personal data, based on pre-agreed principals. Finding the right balance between these solutions was discussed in 11 workshops during the Future Value of Data project.

Many in our workshops argued that, although well intentioned, the current process of achieving consent is unfit for purpose. The European Union's General Data Protection Regulation (GDPR) states that informed consent must be freely given, specific, informed, and unambiguous, but this is very difficult to achieve. The current approach is for customers to tick a box online that confirms they have read and agreed to a contract that allows service providers to collect, share, or trade their personal data, in exchange for various online services. This is impractical, as the majority of customers are disinclined to spend time reading the small print – indeed, they find it irritating to be constantly asked to do so. As a result, most of us only have a hazy appreciation of the potential consequences of disclosing personal information – when, how, and why our data is going to be collected, and with whom this data is going to be traded or shared.[44] In fairness, expecting providers to be able to articulate the nuances of consent in a digestible form doesn't work either. If companies have short and simple privacy policies, they are criticised for not providing enough detail; if they are too long, no one will read them.

Finally, consent only works when customers have the option to use a different service. Given the size and scale of the main digital platforms, some suggest that providing consent, informed or otherwise, is a pointless exercise, as users feel obliged to use the service, and have to accept the terms and conditions, simply because there is no meaningful alternative. Germany's antitrust watchdog has recently ruled against Facebook to this effect. Facebook is appealing the decision.

> "We are not sure if the whole population en-masse will be able to deal with consent, despite improved literacy."
>
> San Francisco workshop

It's not only service providers who are gathering information. Governments also have to wrestle with what limits should be placed to balance public-interest data collection, with individual rights to privacy. For example, a smart city operated or commissioned by a local council has the ability to collect a great deal of personal data about citizens in the course of their daily lives, with the promise of delivering better public services and more efficient interaction with government and local authorities online. But at what point does this become intrusive? Added to this, managing informed consent will get even more complicated as new technologies, such as facial recognition, Internet of Things, quantum computing, and AI emerge, not to mention the growth in the availability of complex pricing models, such as the bundling of different products and services. All of this suggests the need for alternative ways to ensure that those who provide data can exert better control of where and how it is used. Possible solutions discussed during our workshops include greater digital literacy, increased regulation, the adoption of data managers or personal data stores to represent individuals, and potentially a payment to users by service providers in return for access to data.

What We Heard

In Bangalore, the conversation began with a discussion around the taxonomy of data. *"Consent needs to be defined differently. Legitimacy and reasonableness need to be clearly articulated."* This was taken up in Singapore, where the view was that *"there are conflicts between what consumers understand as ownership and consent, and what companies see as access. This shows that there is a need for clearer definitions, articulating new terms. We don't have a clear language."* A recent report by the University of Southampton concurs with this need. *"This is non-trivial, given the rate of change in ICT and the very broad set of purposes to which data could be put."*[45]

In San Francisco, it was observed that, although there are short-term incentives against ensuring greater transparency around the use of personal data by service providers, longer term, there are also clear economic, business-model, and regulatory pressures that should encourage organisations to put greater emphasis on ensuring better public understanding around consent. However, *"the tech is ahead of the regulation here – and that is how, why, and where unscrupulous methods can be used."* They also pointed out that, although greater digital literacy may *"deliver greater self-empowerment,"* the availability of information does not necessarily translate into individuals making informed decisions; *"we are not sure if the whole population en-masse will be able to deal with consent, despite improved literacy."* Given this, they argued that more innovation is needed to find ways to both engage users in better ways of managing how data is being used, and ensuring that products and services are designed so that consent is an integral part of their development. Suggestions included adapting existing technology to include bite-sized explanations, and the ability to more easily review the options around consent.[46] Participants also suggested that data should only be shared if it delivers value to the person from whom it is harvested, but acknowledged that, if online companies are obliged to limit data harvesting to that which has specific benefits to its users, significant changes in current business models may ensue.

"There are conflicts between what consumers understand as ownership and consent, and what companies see as access. This shows that there is a need for clearer definitions, articulating new terms. We don't have a clear language."

Singapore workshop

In Madrid, it was felt that informed consent should be dismissed in preference for establishing agreed standards of behaviour; *"… what we need is a clear set of principles."* This view was endorsed in Jakarta, where it was stated that companies, rather than individuals, should take on a greater burden of responsibility for the management of personal data; *"we have consent fatigue. Organisations need to take this responsibility away from the individual and place a greater onus on the company to ensure that there is no risk or harm."* Those in the Hong Kong workshop suggested that regulators and corporates should work in partnership, and that stakeholder engagement and collaboration is the most sensible approach - albeit one that would take time to achieve. However, the worry was that debates about who should take the lead in this process may mean that, *"without consensus and engagement, the private sector will self-regulate, developing a 'this size fits us' approach, which will not offer an equal platform."*[47]

Some believed that government-led regulation is the only effective way to address the problem, and felt that Europe's GDPR has opened the door to new possibilities for policy makers in other markets and is *"raising the bar for transparency globally."*[48] In Nigeria, it was stated that *"GDPR will bring in new standards,"* and in Santiago, *"Chile will look to other countries as benchmarks for good and bad references."* In San Francisco, there was also general support for greater regulation, in particular a new CCPA law, which comes into force in 2020, that makes California-based companies follow stronger data protection rules, including giving the state's consumers more insight and power over how their data is used, and imposing fines when online companies don't comply. In Johannesburg, it was suggested that increased regulation is the most likely approach, because it is driven by *"consumer pressure and a rising demand for data transparency."* A key driver of this will be the rising digital literacy, which leads consumers to *"wake up and care about the use of their personal data."*

Alternative models were also discussed. In Toronto, it was suggested that, rather than fight for informed consent, which in their opinion is impossible to deliver, it would be more practical to acknowledge that personal data is a necessary raw material for the service providers, and therefore, individuals should be compensated for its use. They therefore suggested that a *'data dividend'*, which could be paid to all citizens by the service providers in return for allowing their data to be collected and monetised by service providers. This would mean citizens could be reimbursed annually for the use of their data by the companies which intend to use it. It follows a similar model to that implemented by the oil companies, which paid a dividend to Alaskan citizens for the extraction of the state's oil resources.

"We have consent fatigue. Organisations need to take this responsibility away from the individual and place a greater onus on the company to ensure that there is no risk or harm."

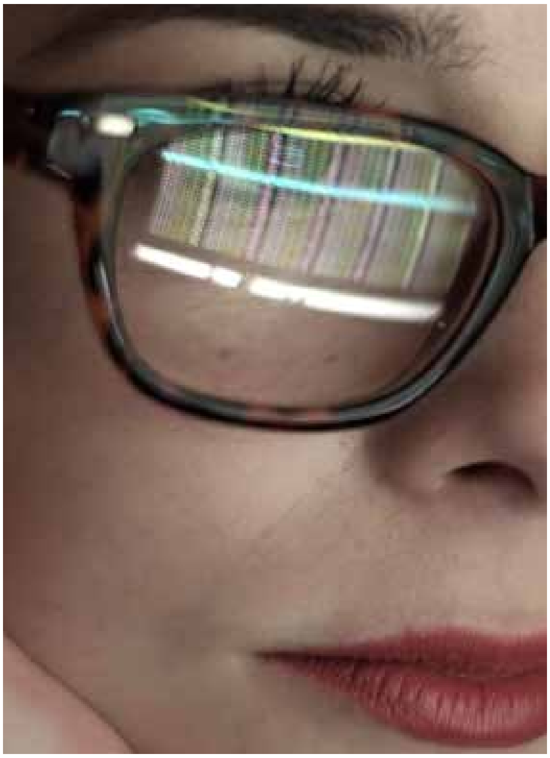
Jakarta workshop

Some, in London, Tokyo, Singapore, and Johannesburg, argued that rather than force consumers to make decisions that they are simply unable to manage, greater focus should be put on the role of data managers who, as trusted third party intermediaries, could better represent consumer rights and enable *"selective and contextual data sharing in context and for the right reasons."* This would give consumers greater control of the principles around which their data can be used, but spare them the drudgery of having to check this every time they sign up to a new service. Regulation, they argued, would therefore be better placed focusing on responsible sharing rather than increasing transparency.

Implications for Data Value

The concept of 'consent' has revealed a fault line that exposes assumptions that lie at the heart of all policy making and regulation, reaching all the way back to the legal myths that form a foundation of contract law - the assumption that all contracts are made between free and equal parties who are fully informed of the nature and consequences of what they are agreeing to (and behind that, the assumption that human beings are first and foremost 'rational' decision makers, always in the business of making 'rational' choices).

The big question is what to replace it with, and in the meantime, what reforms to make to its operations. Many suggestions for more practical, realistic, and workable alternatives have been put forward, including the involvement of trusted third-party intermediaries. Progress on this front will be key, if safe, efficient, and trusted relationships between organisations and individuals are to be established and maintained.



"Without consensus and engagement, the private sector will self-regulate, developing a 'this size fits us' approach, which will not offer an equal platform."

Hong Kong workshop

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.

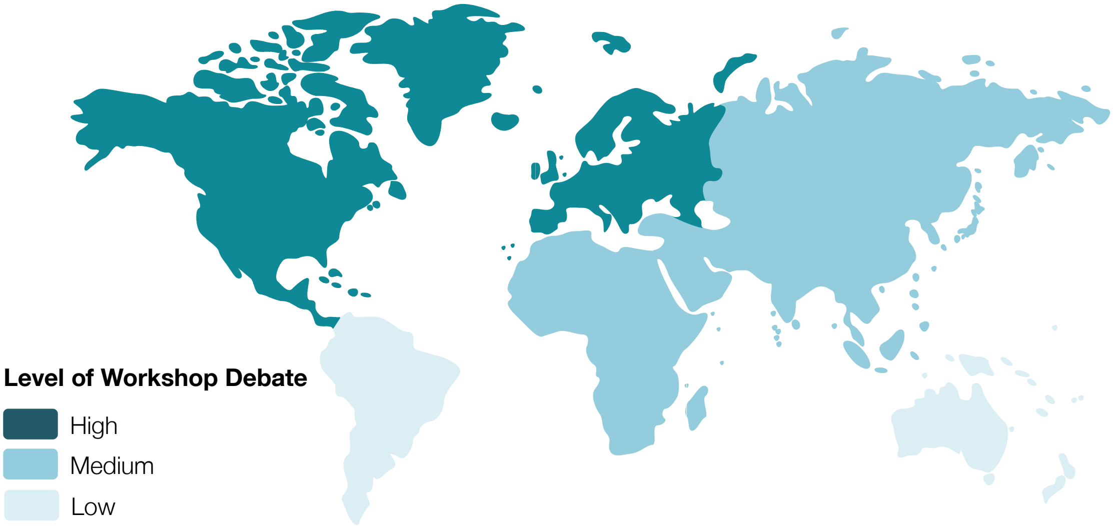
Context

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]



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]

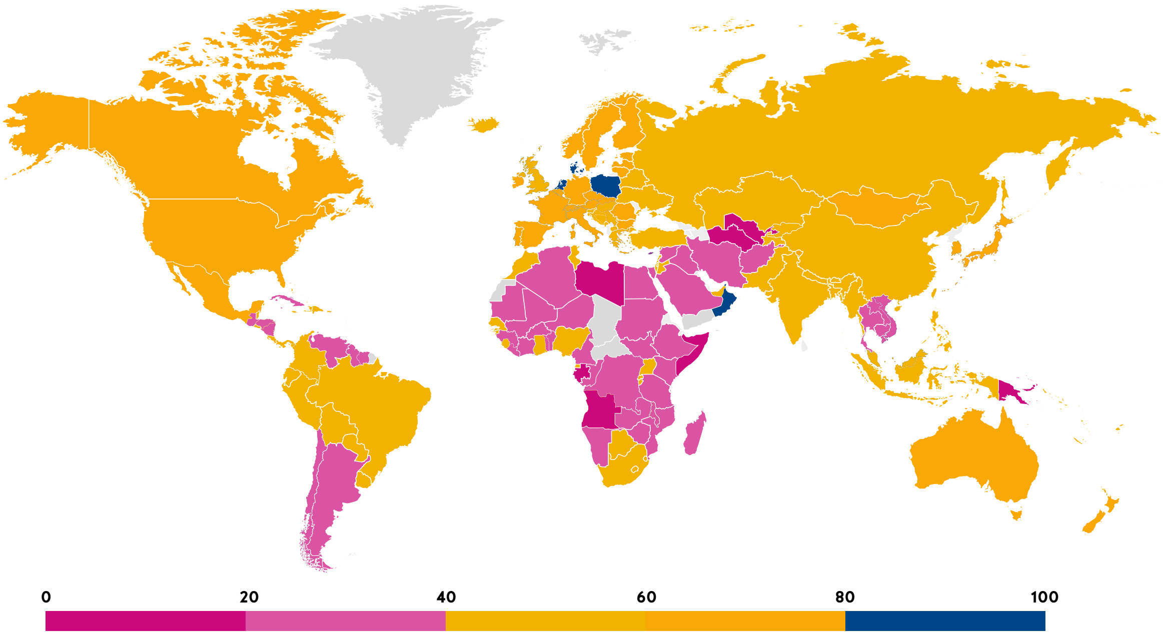
A number of studies suggest that less than a third of the data that is being made available is actually being used.[61] There are many reasons for this, not least a lack of data-handling skills among officials, activists, and journalists. Also, to be truly effective, open data needs to be accessible 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

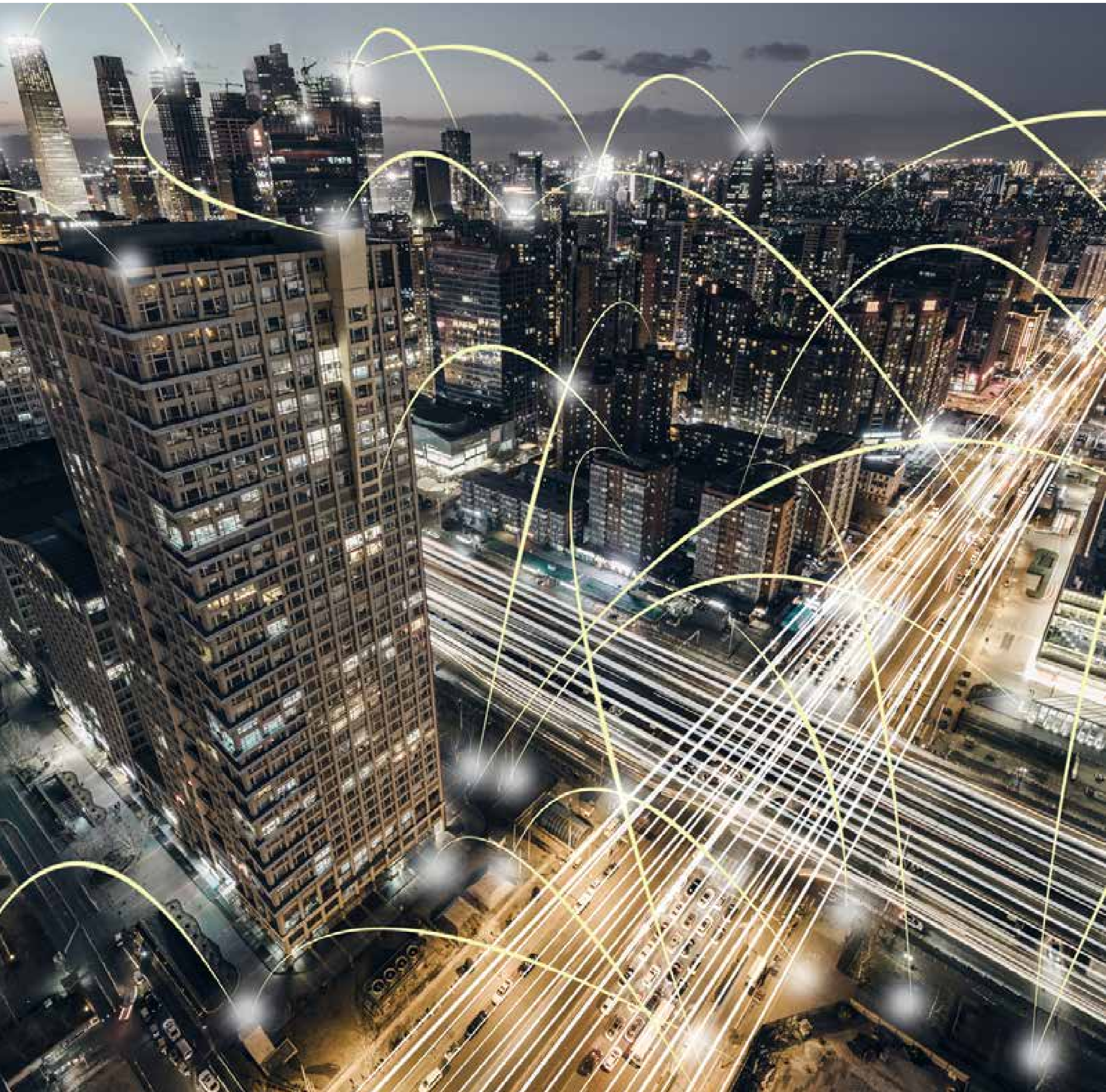
In Copenhagen, it was suggested that we need to define what we mean by the open use of commercial, sensitive, and non-sensitive data:

- 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.

4.5 Ownership of Machine Data

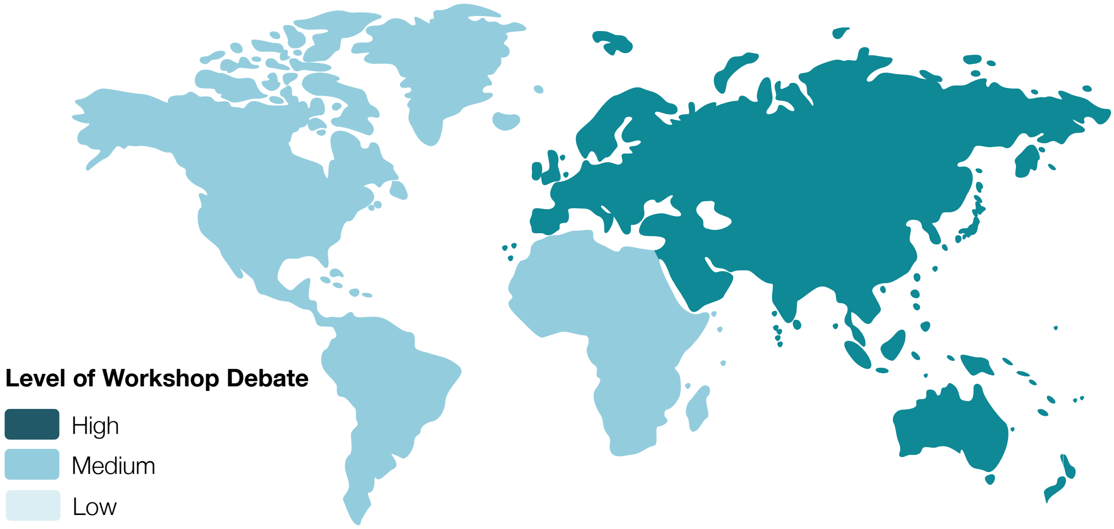


Debates on who has what rights to what IoT data escalate. Questions around title, control, and usage of data lead to many sectors taking different views.

Context

So far, most attention on data has focused on personal data. But looking forward, attention could shift to the increasingly vast quantities of information generated by machines – over 50bn connected devices are forecast by just 2020.

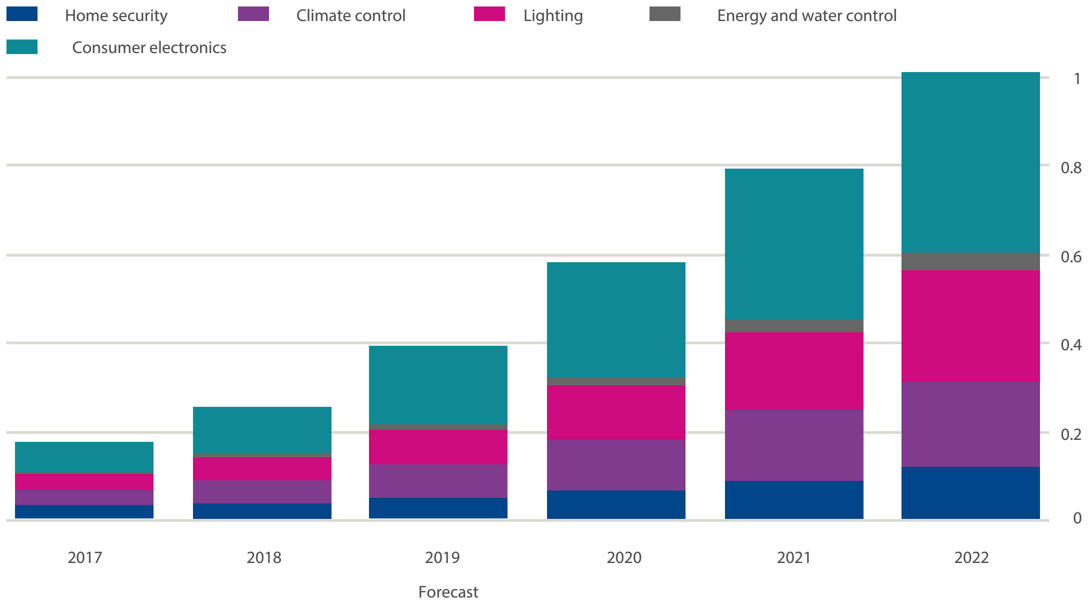
Machine to machine (M2M) data and the broader Internet of Things (IoT) is growing rapidly, having a huge impact on the way we live and how society operates. While many sensors broadcast data, some connected devices act like digital hoovers, sucking in all kinds of information which can be analysed by others and shared and shared again. Without even the click of a button, vital and mundane data is spreading across supply chains, between cars, within buildings, and beyond. Indeed, such is the expected growth trajectory of this type of technology, that some think that by 2030, every device will automatically have a built-in sensor and internet connection.[75] Estimates of just how many connected devices will be in operation vary. By 2030, there will be 200bn of them, says Intel.[76] Cisco reckons around 500bn.[77] China will soon generate 20% of all the data from connected devices.[78] EMC forecasts that the IoT will soon need up to 40tn GB of data storage, while IDC sees 175 zettabytes of data by 2025.[79]



As yet, there are no well-developed principles around the value of data extractions, but the likely financial impact of IoT is high. Bain predicts that by 2020, business-to-business IoT applications will generate $300bn a year.[80] One estimate suggests that a 10% increase in machine-to-machine (M2M) connections will generate more than $2tn in the US over the next decade.[81] PwC predicts there will be $6tn of investment in the US alone.[82] Whatever the actual numbers, one thing is clear; as the Frankfurt workshop put it, whoever owns all the IoT data is about "to become a very big deal."[83] It will also therefore be a source of intense controversy.

Ownership Uncertainty

The key question to ask is whether IoT data will have greater value if it is proprietary or open to all. Certainly, maximising the opportunities presented by the IoT is not as straightforward. A core issue is that in many sectors, there is as yet no agreed approach for machine data ownership, and many grey areas over control, beneficial use, and access. While there has been steady progress on the complex debate on personal data, for machine data there is little clear headway on whether, for example, ownership aligns with a device manufacturer or the device user.



SOURCE: IHS Markit

Rise of Machine Data: IOT devices growth (2017 to 2022)

In the increasingly automated agricultural sector of the 21st century, for example, the farmer may or may not own the data produced by the machinery in the field; the farm equipment manufacturer often has the right to take that data and use it across a wider system.[84] Across the food supply chain, just as wheat is harvested, processed into flour and used to bake bread that is sold on to an end user, so too is data. As was suggested in a parallel Future Agenda discussion on future land use, *"for a supply-tracked beef burger, the debate on who owns the provenance data about the cow it came from – the farmer, the meat processors, or McDonalds – is just one simple example about which there are alternative views."* The McDonalds supply chain is famously efficient and collaborative, but with millions of farms involved, who actually owns what data is not clear.[85]

In the automotive arena, many are excited about the potential and roll-out of increasingly connected autonomous vehicles – all generating and sharing huge volumes of data. Toyota estimates that the data volume between vehicles and the cloud will reach 10 exabytes (1018) per month around 2025.[86] Many owners or leasers of a car may believe that the data it produces, and so at least a good proportion of the value, does, or should, belong to them. But others across the sector have different views, and answers might vary according to the nature of the data.[87] For example, location, speed, destination, outside temperature, and emissions data, may well be made open for all to use, while more specific information on, for example, road condition, fuel levels, driver tiredness, brake and tyre condition, as well as even accident data, may be held by several interested parties, including the car manufacturers, insurance companies, repair services, government agencies, and fuel brands. *"Very little automotive data, other than detailed engine performance information, may be proprietary. As such, there is likely to be little value in the data itself, but rather the impact shifts to the outcomes of its use."*[84]

Overall, given all the activity, investment, and strategy development by a host of major governments and companies, from our discussions, there is no universal answer to the question of machine data ownership on the horizon. Many different parties with varied vested interests are keen to at least agree some ground rules, if not come to a global protocol, but it may be years before significant progress is made. Gaining clarity on who owns machine data and who is legally entitled to use it for analysis and additional value creation, is a key priority for many.

"The provision of leading-edge analytics will help maximise the potential value extracted from data, and provide a more level playfield for SMEs."

Jakarta workshop

What We Heard

As mentioned previously, many believe that data should not be subject to the laws around property. However, in the West, the owner of the data is often considered to be the organisation that holds legal title to the device that recorded or generated the data – be that a streetlight, a tractor, a doorbell, or a high-speed train. As long as there is no other agreement in place, then perhaps the only entity that has the right to use or dispose of that data is the one that actually produced it in the first place. So, data title is like a deed to a property. However, as was highlighted in our parallel conversations on the value of automotive data, *"the organisation who has possession of a machine is not necessarily the owner of it; things can get rather uncertain when for example equipment is being leased from one organisation to another."*[89] As leasing is now the preferred approach for many sectors, from agriculture and transport to healthcare and building management, this matters.

Some experts feel that whoever generates the data owns it, and it can then be sold on. But others suggest that in the increasingly complex ecosystems and decentralised supply chains and webs now operating across many sectors, the source of any data may be from multiple parties, plus a host of those involved in the product delivery think that they own the data. Not surprisingly, therefore, several in Tokyo suggested that *"we need a fundamental rethink about who owns the data."* In fact, *"there are no general laws about information property, other than some regulatory rules in vertical industries."*

Going forward, workshop participants feel that the most significant change will come in two main areas – the role of AI and access to analytics.

- In Frankfurt, some felt that *"data will increasingly be created and used by machines, and never be touched by humans."* As such, *"machines will make automated decisions, as M2M and AI authority takes over,"* and so control moves to the algorithms, or whoever owns them. A linked proposal in San Francisco was that *"we will see algorithmic regulation to address machine data that is beyond human governance."* So, as machines create and use more data, maybe AI will be needed to police this, and included here will be the questions of ownership and value. An additional view from Japan was that *"in the future, metadata will be built by AI"* and *"the ownership of metadata will be challenged."*

> "We will see algorithmic regulation to address machine data that is beyond human governance."
>
> San Francisco workshop

- In Jakarta, there was a strong view that new data analytics capabilities from the Internet of Things should be made more accessible to wider industry rather than just Big Tech. *"The provision of leading-edge analytics will help maximise the potential value extracted from data, and provide a more level playfield for SMEs."* Indeed, several felt that, if not appropriately regulated, this imbalance of capability between the few leaders and the mass of industry could lead to significant inequality at both a sector and a national level. A related view in Germany is that over the next decade, *"mid-sized businesses will struggle, as large corporations benefit because they have the resources and the data."*

More generally, the consensus in a Stockholm discussion was that we need to move towards a more *"heterogeneous understanding of IoT," and* potentially require some sort of *"quality of assurance for IoT data."*

Implications for Data Value

In a field where trillions are normal day-to-day statistics, it is increasingly apparent that the ownership of machine data is already a big issue. Given the uncertainty, and who has rights in what circumstances , some see it as surprising that so many major companies and VC funds are making huge investments in smart cities, connected cars, and digital trade, and most view the potential value of the machine data as a central part of the business case. However, despite the lack of clarity, interest from cities, governments, and wider society will undoubtedly grow. The provenance, ownership of, and access to machine data is a mounting debate across many industries. The value of that data and to who is set to become pivotal.

> "Data will increasingly be created and used by machines, and never be touched by humans."
>
> Frankfurt workshop

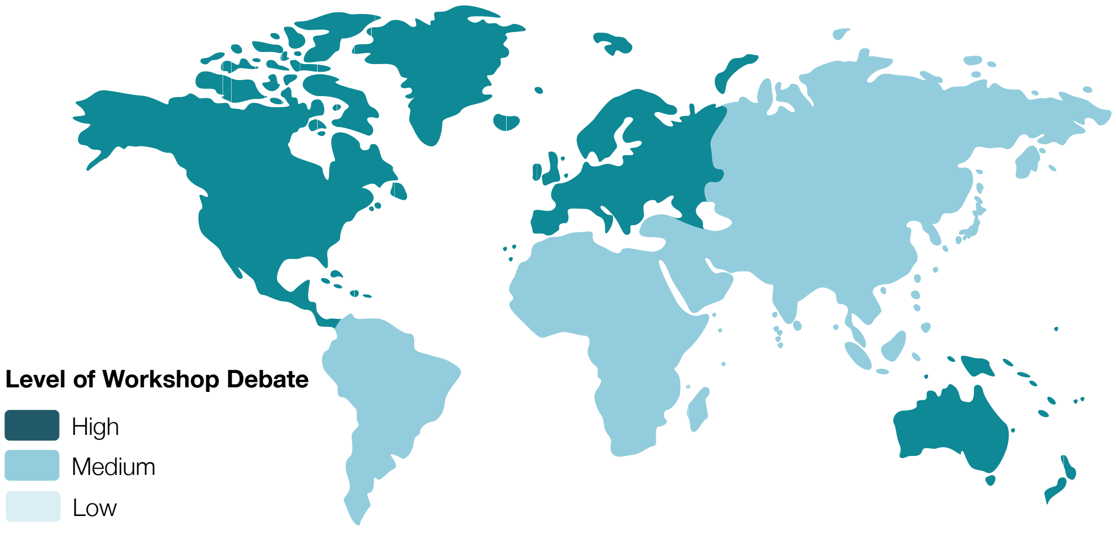
4.6 Data as an Asset

Organisations are obliged to account for what data they own or access. They are required to report their full data portfolio, and are taxed on this.

Context

It is increasingly recognised that data is a valuable asset to the organisations that collect it. But so far, data-driven businesses have not always aligned well with existing business concepts or taxation mechanisms. A company which owns lots of property or other physical assets clearly has a lot of 'capital'. But can or should data be seen as an asset and even as 'capital' - especially when it is either personal or machine data that is not owned by the organisation concerned. Moreover, what is the value that is being taxed?

If data is officially recognised as a corporate asset, significant organisational, industry, and trade implications could follow. As first articulated in a workshop in Jakarta, if a company's future value includes an assessment of the data that it owns, manages, analyses, or accesses, then the way data-based businesses are valued, and perhaps taxed, will be transformed. Data may itself be measured as an asset. The possible implications of this, for business, for economic growth, and indeed how national GDP is measured, are considerable.



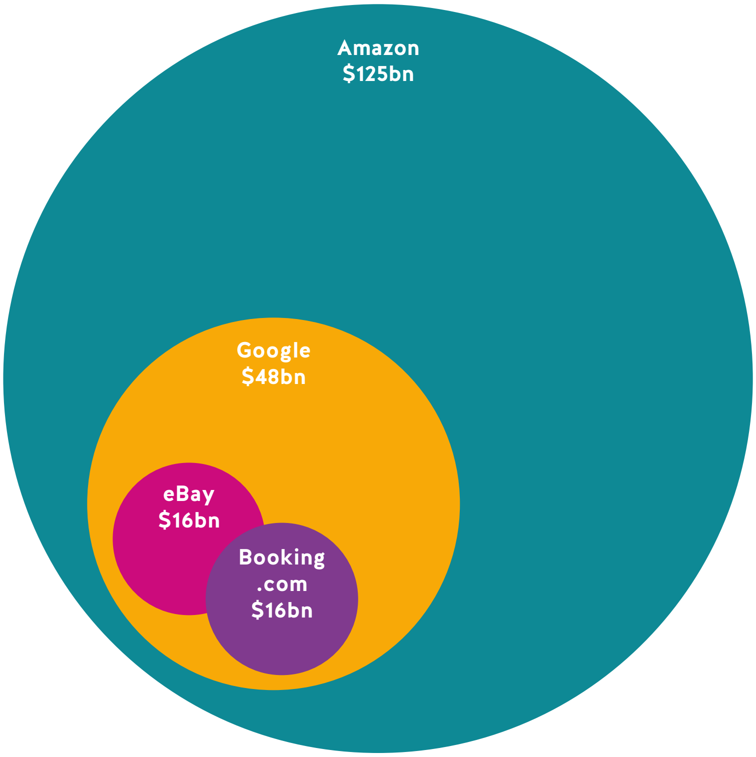
Data as an Asset

Many experts suggested that if data is considered to be an independent asset, then it will be more rigorously monitored and tracked, and potentially regulated. Increasing numbers of academic researchers are investigating this scenario.[90] If data is officially recognised as a corporate asset, in the future, organisations may well be obliged to account more clearly for the data they control and use. Every major company, government, and NGO may legally be required to declare the value of its data assets on a regular basis. This could involve formal accounting valuations of some data sets, but it could also include assessments of the value generated by these assets.

The pivotal challenge here is how to value one entity's data so that it can be compared against another's, or a wider benchmark. Flows of data are not a commodity: each stream of information is different, in terms of timeliness, or how complete it may be. This lack of 'fungibility' makes it difficult to define a specific set of data, and to put a price on it so that the value of one data set can be determined.

> "EU taxing commercial activity of digital firms is not taxing data – it is about closing taxation loopholes."
>
> San Francisco workshop

SOURCE: https://www.imf.org/en/News/Seminars/Conferences/2018/04/06/6th-statistics-forum



Who Has What: Estimated Value of Data (2017)

Although the current focus for many in business and government is on personal data, different sectors are trying to come up with an agreed way to value their own specific data sets. The oil industry, for example, is beginning to align around its seismic analysis used to map reserves; in the automotive sector, efforts are underway to find a way to value the data generated by connected and autonomous vehicles; and the value of IoT data within smart cities is a mounting area of attention. Governments are also keen to understand the value of their data assets and are trying to establish common standards. In 2018, for example, a UK Parliament Select Committee[91] discussion suggested that the value of the aggregated NHS patient data set could be around £10bn.[92] The UK Government is sounding out options.

To provide some rigour, the IMF, among others, is trying to help define an approach to calculating data assets; researchers at a November 2018 conference explored how measuring economic value needs to recognise the impact of data. One paper estimated that in 2017, Amazon's data was worth $125bn and was growing at 35% per annum – so data accounted for 16% of the total market value of the company.[93] Google's data was worth $48bn at the time.[94]

Some consider that those with the data assets are already making plans for calculating their value. For those interested in buying information on the dark web, for example, the relative value of personal health data is around ten times the value of an individual's credit card information.[95] Experian, for one, has detailed what common pieces of personal information are currently sold for.[96] The FT also has a personal data calculator.[97] More legitimately, a host of investment banks, economists, and consultants are doing their own analysis on the leading tech companies, as a means of better rating them and predicting future stock values.[98]

Data as a Liability

Once data is seen as an asset, it can also become a liability. It certainly has to be stored and properly maintained – both of which incur costs. Businesses have to allow for this. Accountants will still have to balance books and calculate data equity, so having data liabilities to offset against data assets will be important; after all, assets provide a future economic benefit, while liabilities present a future obligation or risk. Storing some kinds of data could, for instance, be seen to erode user trust and therefore become a liability. It may also mean that costs of securing data will outweigh the costs associated with losing it. Data security experts argue that it would be more appropriate to consider the vast amount of the data organisations hold as a liability, since the value they can extract from it is minimal in comparison to the costs of preventing it from being stolen or misused, or paying the price when it eventually is

> "If we actually did have a more formal system for measuring the value of data as a capital, we might be better able to use it, since 'how to use it' would be factored into this value."
>
> Madrid workshop

Some markets such as the UK are already charging significant fines to companies that fail to protect the data in their care. Increasingly, this, combined with the ingenuity of today's hackers, has meant that corporates must set aside capital to account for this. An unintended consequence may be that competition is stifled, as the barriers to entry for new business becomes simply too high.

Digital Taxation

Controversial in the US, but more widely accepted elsewhere, is the idea that governments could (and should) exact a tax on an organisation's digital activities. The EC has proposed a so-called digital service tax of 3% on the local activities of Big Tech firms such as Google, Facebook, and Apple.[99] The UK has set a precedent by announcing its intention to introduce a digital services tax by 2020, so that multinationals "with profitable UK businesses pay their fair share."[100] Other member states in the EU have put forward proposals at a national level. Recently, the OECD also announced a target of 2020 to agree similar rules.[101] To date, all these focus on taxing revenues from activities.

Data Tax

What is being discussed so far is not a tax on data, but on digitally-related income. However, this could be a precursor to a wider tax on data – and in particular on an organisation's data assets. Just as several European countries and the likes of British Colombia in Canada apply an annual personal wealth tax, based on the market value of assets that are individually owned, so if a company's data has an agreed value, then, it is argued, governments could exact an annual data asset tax on top of, or as part of, corporation tax.

For organisations, there is a clear downside to a data tax. Many see that it could stifle innovation, as information is dumped in order to minimise costs. On the other hand, some think that, from a social impact perspective, this could be a significant leveller, and would herald the end of the data land-grab of recent years. They argue that if it happens, this is simply a sign of a growing maturity in the data sector, and a realignment of power and money.[102] Whichever view is taken, researchers are now looking at the broader implications of the extra value creation and the impact on national and global GDP, if digital revenues, data taxes, and other data assets were included in calculations. As one US workshop participant stated, *"when data capital gets combined with digital tax, then it will become really interesting."*

> "It is more likely that a common approach to certifying data for valuation will evolve from the bottom up, via an industry, regional, or even community approach."
>
> Tokyo Workshop

What We Heard

Data Assets

There was general agreement that, rather than being *"initiated at a global level from the top down, it is more likely that a common approach to certifying data for valuation will evolve from the bottom up, via an industry, regional, or even community approach."*[103] Ways in which to "justify how to put a value on something that may not belong to you" were discussed in Hong Kong. In San Francisco, the view was that this would best be undertaken by an independent governing body, in order to ensure transparency and credibility. This idea was also explored in Toronto, where it was proposed that *"we need a common framework that is agreed (per industry)."* Many around the world concurred with this; however, there was no consensus around which global organisations would be capable of taking it forward.

Data Liability

In Europe, existing liability laws are based on the concept of physical products, so there were a number of discussions around whether these could be adapted and applied to data-based products.[104] In Sydney, it was proposed that the idea of data liability should be extended to include data negligence, and one suggestion was that there is *"responsibility to share and use data for the common good,"* while another was "failure to use data appropriately for both private and public benefit will be seen as negligent."

Data as a Capital

Another suggestion originally coming out of Sydney, and supported in London, San Francisco, and Toronto, was to add data as a 7th capital in the multi-capital model that currently underpins integrated reporting. A number of organisations are already moving from simply reporting on their financial impact, to include social, environmental, natural, and human capital in their annual reports.[105] Led by the likes of AXA, Puma, and Unilever, a growing portfolio of major companies are involved in these discussions, and are preparing to disclose the wider impact of their business outcomes. They are trialling and agreeing standardised approaches for measuring and reporting the impact and value of what they envisage is the full range of activities, so including data capital in the mix could be a timely evolution. In Manila, it was felt that *"if we actually did have a more formal system for measuring the value of data as a capital, we might be better able to use it, since 'how to use it' would be factored into this value."*

"When data capital gets combined with digital tax, then it will become really interesting."

San Francisco workshop

Others disagree, pointing out that, unlike other intangibles such as R&D assets (e.g., patents), which may well depreciate in value over time, the aggregation and recombination of data can create new value, and therefore data capital may well grow faster than the other six and so skew future views of an organisation's impact. Some think data is already being accounted for through R&D. In London, the view was that *"data capital reporting is happening and here, already baked into much R&D valuation, especially in terms of IP,"* while in Toronto, one comment was that *"this is just like IP capital (but broader)."* However, in San Francisco, a challenge to this was *"does data itself count as IP or do you have to do something with it to make it valuable?"* If it does, then a separate tangible value on data capital, at least in business terms, may emerge.

Data Taxation

While many companies are lobbying for a global agreement on data taxes (via the OECD), several US firms and political leaders are arguing strongly against this move. The view in the San Francisco workshop was that this is *"governments fishing for ways to generate income from data, and does not feel right,"* and that *"EU taxing commercial activity of digital firms is not taxing data – it is about closing taxation loopholes."* Others see that these initiatives give licence for other countries to follow suit.[106]

South African opinion was that, in general, *"African governments don't have the capacity to tax the digital economy – they don't even tax the oil industry properly."* Several expressed doubt about the ability of regulators to address the problem *"… governments [in Africa] face significant challenges if they want to tax digital transactions. There needs to be a better understanding of the data value chain; where data is created, the value it produces, and who benefits from this."* They also noted that, although in theory, social media is already being taxed in some locations, the reason why Ugandans may have to pay the equivalent of five cents a day to connect to any of their preferred social networking sites is more about curbing freedom of speech rather than redirecting revenues.[107] In Jakarta, the perspective on this was that *"the issue is very politically dependent – it is driven by the individual finance minister – and how he wants to raise income."*

> "There needs to be a better understanding of the data value chain; where data is created, the value it produces, and who benefits from this."

Implications for Data Value

Although several in the digital economy dislike the idea that data can be considered as an asset, many others, including governments, inter-governmental organisations, and consultancies, are very keen to push the concept forward. As yet, it may not be coherent in terms of the mechanics, but if an industry or region can agree fundamental principles, a whole raft of change will be set in motion. The challenge is to create a regulatory environment which encourages competition, while making information-intensive organisations more accountable for the data in their care.

Some initial discussions about the value of Amazon's and Google's data over and above its financial wealth, suggests that either this is not currently being factored in. If, within the next decade, analysts and economists come to some shared understandings, seeing data as an asset could be one of the biggest influences on how we see the value of data, and may well determine how responsible organisations are seen to act.

"We need a common framework that is agreed (per industry)."

Toronto workshop

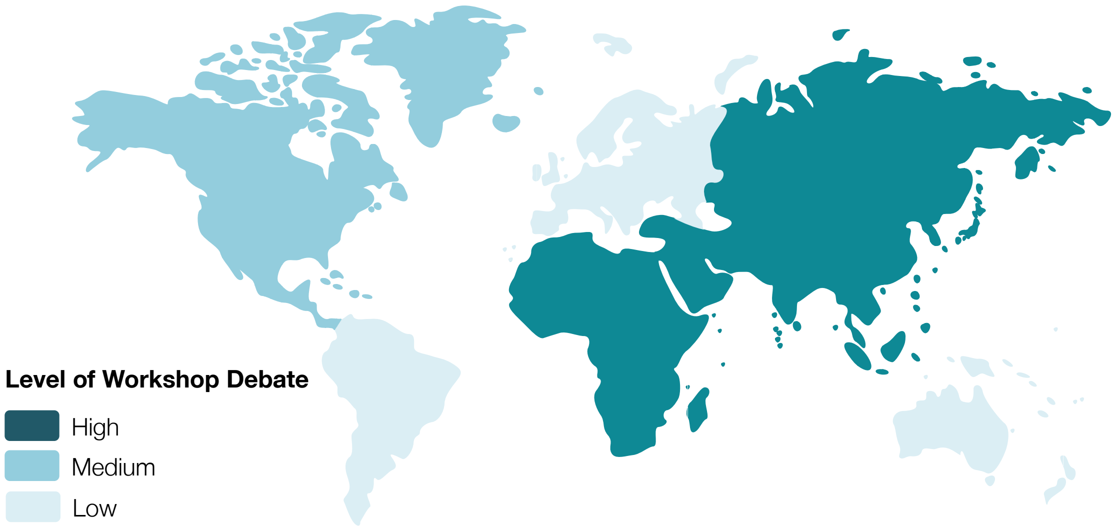
4.7 Data Localisation

Nations see benefit in copies of all citizen and machine data in regional centres. Government and local companies seek access to data held by foreign corporations.

Context

Data localisation aims to ensure that a copy of all nationally-generated data remains stored and accessible in the country of origin. It attempts to restrict data flows across borders by either mandating companies to keep data within a certain jurisdiction, or by imposing additional requirements before it can be transferred abroad. The objectives behind these restrictions are diverse, including privacy, cybersecurity, public order, law enforcement, taxation, and economic development.

Support for localisation is growing in a number of countries. In highly populated Asian nations, such as China and India, many think curbing access to national data will facilitate economic growth locally, and build or protect political power. This is prompting many new measures. In India, for example, in 2018, the Reserve Bank of India prohibited companies from sending financial data abroad, and a draft government policy envisages a ban on the international transfer of data generated by Indian ecommerce users. The number of restrictions on cross-border flows has tripled over the last decade, with over 80 in place at the time of writing.[108]

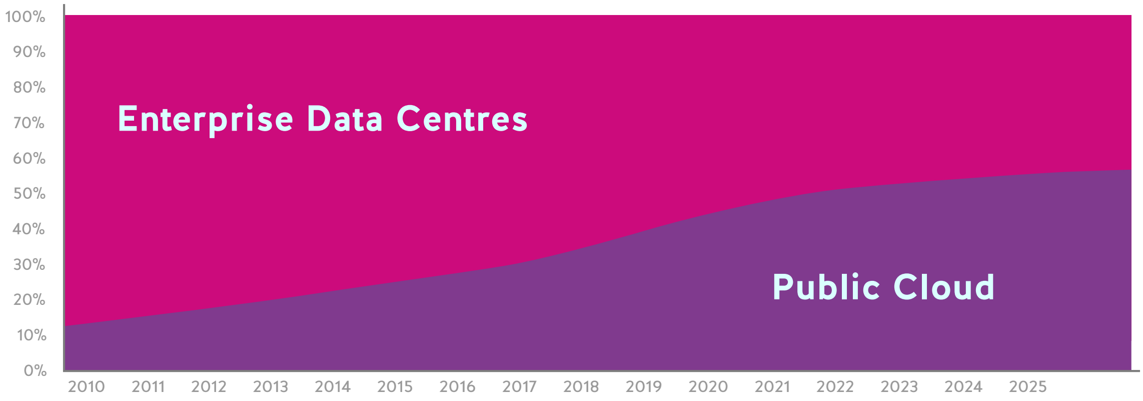


Opponents of data localisation argue that it restricts, rather than stimulates growth, with consultants such as Deloitte suggesting it will have negative economic consequences.[109] Proponents of cross-border data flows argue that local legislation undermines free trade by adding onerous and expensive obligations for businesses, including building, operating, and maintaining data centres in multiple countries, as well as creating and updating separate data sets – even if they are a mirror of those held elsewhere. Add to that the inconvenience of having to go through a number of regulatory approvals to either operate in a market or comply with specific sector rules, and it's clear, they argue, that this restricts opportunity.[110] Opponents of data localisation therefore argue that it is counterproductive for emerging economies, constraining economic growth and with a negative impact on social development.

What We Heard

In the discussions, those in favour of data localisation focused on three main areas:

1. **Economic Development –** Encouraging investment in and the development of national data centres that drive, and are linked to, foreign direct **investment.**
2. **Technology Ecosystems -** Seeding growth of local centres of data expertise and access, that encourage regional company innovation and growth.
3. **Market Access –** Using data regulations as a political lever, where multinationals cede control of data sets in return for market access.



Source: Data Age 2025, sponsored by Seagate with data from IDC Global DataSphere, Nov 2018

Global Data: Data Stored in Public Clouds vs Corporate Data Centres

Economic Development

A constant thread throughout many discussions was that, despite the increase in global GDP, the real value of data trade to date has been largely ringfenced and retained by multinationals. In Hong Kong, opinion was that *"there are some companies whose profits exceed the GDP of many nations, and which wield extraordinary power. This power is in private hands and not accountable to democratic processes, which is potentially very dangerous."* There was a sense in some workshops that participants, several of whom were policy makers, wanted to push back against this. In Bangalore, for example, the perspective was that *"companies don't respect governments, unless they have a workforce on the ground."* India's richest man and Chairman of Reliance Group has been quoted as saying, "India's data must be controlled and owned by Indian people and not by corporates, especially global corporations."[111] The national government is keen to address this, and sees the potential to both curb the power of large foreign companies and also boost local industries through localisation legislation. China is adopting a similar approach, and other nations are watching with interest. In our Jakarta workshop, it was observed that *"there is a risk of an increasing digital divide... so the role of government in relation to the management of data could be transformative."*

However, in Sydney, it was observed that localisation laws are only really beneficial for countries with large populations; *"a few mega-countries like India can have their own independent system, but most others know that they do not have the influence to restrict sharing."*

Technology Ecosystems

The other, connected, argument in favour of localisation is that it can boost the local tech sector. This was proposed in Nairobi, where it was felt it would *"drive locally-driven tech innovation"* and *"facilitate the development and enactment of legislation to support growth in IT service consumption – as an engine to spur data centre growth."*[112] On the face of it, this might seem true, as more data centres will have to be developed locally. However, others argued that a boost for the data centre business will be outweighed by lower efficiency from using relatively expensive domestic data storage, and by the loss of foreign processing trade. They also pointed out that, increasingly, goods supply chains have an associated data stream feeding information back and forth between the manufacturer and the user. Growth will be therefore restricted if data cannot be aggregated internationally.[113]

"There are some companies whose profits exceed the GDP of many nations, and which wield extraordinary power. This power is in private hands and not accountable to democratic processes, which is potentially very dangerous."

Hong Kong workshop

Building on this, in Manilla it was felt that the existing Philippines data protection laws are suitably robust, and provide effective controls around the potential misuse of data. Therefore, rather than close its doors to data, it was suggested that the opportunity is to position the country as a *"centre of excellence when it comes to processing data from other regions and countries."*

In India, localisation legislation is setting precedents, and is supported by a powerful combination of tech leaders, and state and national politicians, not to mention the Reserve Bank of India.[114],[115],[116] The current proposals cover national security, economic development, and the desire to build local technology-enabled innovation ecosystems. Multinationals, including those from India itself, such as TCS, Infosys, and Wipro, that are dependent on operating within agreed international frameworks, however see this policy as short-sighted.[117] In the Bangalore workshop, one prognosis was that *"a new compromise may well be developed, based around international standards…..however, the situation is likely to get worse before it improves, as there is currently little consensus around data localisation."*

Market Access

With its Great Firewall, China successfully controls its own internet. Although many outside China agree with the principle of sector-focused data localisation for the likes of health and financial services data, some see numerous contradictions in the Chinese Cyber Security Law, which came into effect in June 2017 and was fully enforced in early 2019.[118] This includes controversial provisions affecting transfers of personal data out of the country, and prevents firms unwilling to comply with these rules from operating there.[119]

One important issue is the extent to which the Chinese government has access to data stored within its boundaries. Microsoft's Azure cloud service in China claims to be in an independent third-party data centre, and the AWS infrastructure is privately owned. However, few in any of our discussions on this believe that they are beyond the reach of the Chinese state. Apple, by contrast, has chosen to use the Guizhou-Cloud (GCBD) – a government-owned data centre. This was questioned in our Bangkok discussion, where there was scepticism about the real depth of the company's stance on privacy. In the West, Apple has positioned itself as an organisation that defends privacy as a civil right.[120] However, some, particularly those we spoke to in Asia, now see that these principles have been compromised in order to access the significant Chinese market.[121] Certainly, the view in Bangkok was that *"Apple has caved in."* Furthermore, concern was expressed about the independence of the global Chinese technology companies which store data from other countries on their servers. Many believed that they are also obliged to give the Chinese government access to their records.[122]

> "Data differences are one aspect of a large systemic conflict... but this matters, because as China grows, more people/nations will try to emulate it."
>
> Washington DC

In Hong Kong, the perspective was that we are witnessing a cultural challenge to the way the internet will be managed in the future; *"what would be the implication of China winning the debate about data, and what would happen if it exported its values around the world?"* As this battle continues, there may well be one set of internet standards for the West, and another for key parts of Asia, they argued.

Implications for Data Value

Several nations are now pushing back against localisation regulation, most significantly the US and the EU. In Washington DC, this was framed as part of a broader geopolitical change; *"data differences are one aspect of a large systemic conflict... but this matters, because as China grows, more people/nations will try to emulate it."* There is also significant action across SE Asia. In Thailand and the Philippines, both of which have separate data privacy legislation that could be applied to data localisation at some point, the general appetite was for the development of privacy frameworks that protect consumers, while also allowing data to flow across borders.[123]

Several put the rise of localisation regulation down to a lack of expertise amongst policy makers. In Bangkok, the suggestion was, "The quality of government officials' data knowledge needs to improve – and with it, the understanding of the potential benefits." In Bangalore, the view was that "we will see an increasing assertion of data localisation around the world, but at the same time there will be growing discontent as consumers complain of a slower Internet, and the delivery of goods and services being hampered. Potential investors may choose to go elsewhere."

Reasoning against localisation, Singapore is seeking to change the direction of travel, arguing that those that store data locally pose a risk to the growth of the region's digital economy. For example, the nation's central bank chief recently shared his view that "if data cannot cross borders, the digital economy cannot cross borders, and we will be poorer for it." Moreover, "a good part of data localisation that is happening in the world today is due to misguided notions of cyber security or data privacy, or worse still, old-fashioned protectionism."[124]

Data localisation is caught up in a pushback against globalisation, and there is a growing awareness of the divide between those who produce data and those who exploit it. Until recently, multinational organisations have profited from the lack of regulation, but many now see that, despite the cost and inconvenience, if they want to participate in the fast-growing, hugely populated markets of the new economies, there is a need for a stable and consistent regulatory environment. A suggestion first expressed in Bangalore that *"the creation of a World Data Council may well facilitate international negotiations,"* was widely supported.

Looking ahead, although there is interest in developing international principles, such is the dissonance between different nations, there is little expectation that it will happen any time soon. While multinational companies and inter-governmental bodies may increasingly lobby against localisation, in a world of increased patriotism and nationalism, they may well have to take more significant measures to address the very real concerns about cultural sensitivity, economic growth, and national security.

"A new compromise may well be developed, based around international standards…..however, the situation is likely to get worse before it improves."

Bangalore workshop

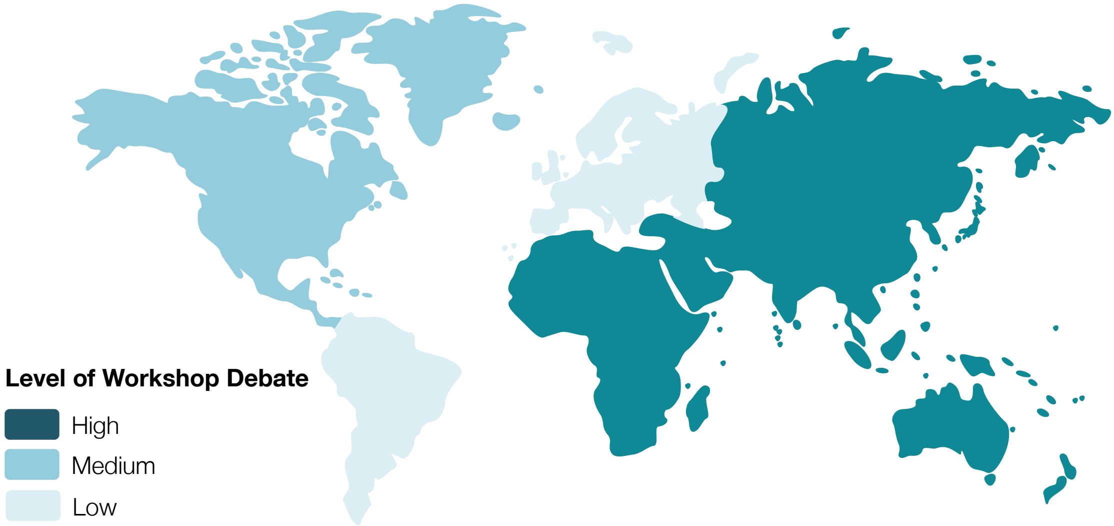
4.8 Data Sovereignty

More governments see control of national data as a means to protect citizens' rights, develop the economy, and maintain a sense of cultural identity.

Context

During the early days of the Internet, data flowed freely across national borders by default. The technology made it quick, easy, and cheap, and there were no rules, regulations, or public concern to stop it. Global corporations benefited particularly from this. But there is now a growing push-back.

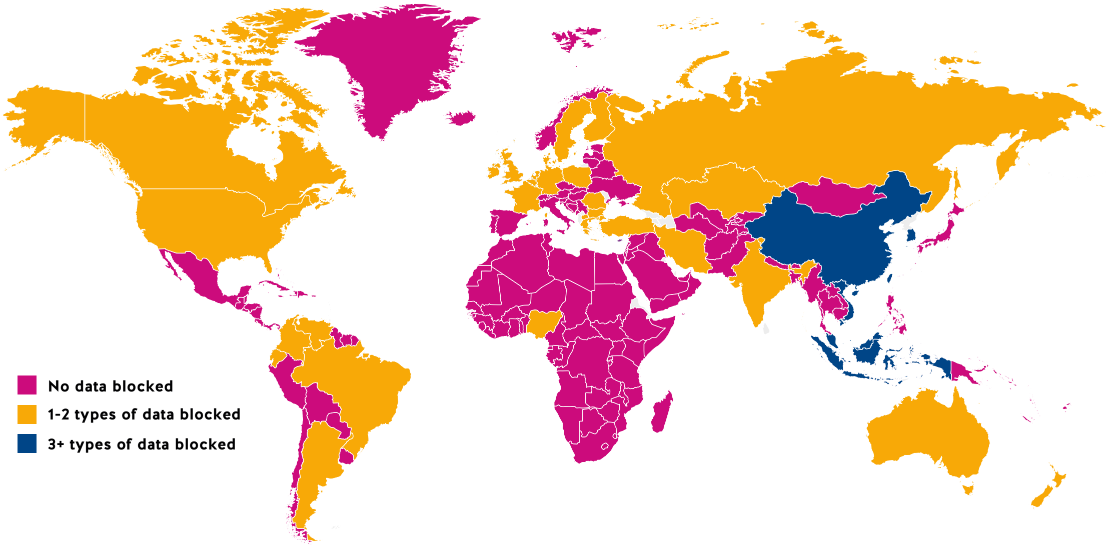
A rise in nationalist sentiment, mounting fears around privacy and data security, a determination by some to rein in 'surveillance capitalism', and demands that individuals and local economies should get a fairer share of the benefits of data, are all contributing to a worldwide trend to restrict or halt cross-border data flows. Today, over 60 countries are implementing policies designed to do this. Active discussions are underway between national and regional governments and the private sector to shape data sovereignty regulation across the Americas, Europe, and Asia Pacific.[125] Countries as diverse as Russia, Germany, France, Indonesia, and Vietnam have now mandated that their citizens' data is to be stored on physical servers within the country's physical borders; in the US, certain federal agencies require their data be stored exclusively within their national boundaries; Australia has a clearly defined legal framework for health data; Europe's General Data Protection Regulation (GDPR) also restricts organisations from transferring personal data that originated in Europe to any country without adequate data protection laws.



Those who are opposed to this rising trend argue that open data flows are fundamental to today's digital and physical commerce, and a vital catalyst for innovation. Therefore, the ongoing development of the digital economy and continued productivity growth across the more traditional industries, depend on the ability to transfer data, including consumers' personal data, within and between countries for efficient analysis, processing, and storage. Moreover, the freedom to move personal data without restriction between countries generates positive outcomes, not only for organisations, but for citizens and countries as well. This is particularly relevant in countries with an authoritarian government, or where there are restrictions around freedom of speech.

Why then is there still such support for data sovereignty? During our discussions, three primary reasons for its appeal were identified:

1. National Security
2. Citizen Surveillance
3. Data Imperialism



Countries Blocking the Global Flow of Data (2017)

What We Heard

National Security

In India, it was observed that in the future, *"the key players will be the data rich, not the richest - the amount and availability of data, rather than the size of the country, will define multinational treaties and data sovereignty power."* Many we spoke to agreed, and there were numerous discussions about how to protect access to sensitive national data of all types, particularly as advances in data technology has made rapid cross-border data sharing easier. In light of this, both American and Chinese surveillance techniques were a subject of intense debate, and our workshops looked at ways in which nations could enhance digital security by limiting cross border data flows and making investments in cloud computing.[126] A number of governments, including those in Brazil, India, and the European Union, have already sought to do this.[127] Elsewhere, conversations in Singapore, Jakarta, and Hong Kong highlighted the need for nations to retain control of their citizens' data, as a matter of national security. The concern in Jakarta was that currently *"all government, corporate, and personal email is largely dependent on western platforms,"* however, *"regulation is in development to address this."*

In Singapore, where trust in government is high, there were strong views about the importance of data sovereignty to ensure national security, particularly with regard to the sharing of health data: Although *"no-one has yet worked out the extent to which patient data can compromise government security……. our existing laws restrict the sharing of personal data (including health data) beyond the national boundary."*

Citizen Surveillance

Some argue that increasing state surveillance is necessary for national security, but it can also restrict individual rights. In Pretoria, there was recognition of the need to have a nuanced approach to balancing national security, with freedom to share and access personal data. The question was asked, *"how do we manage the legislation of personal communications in the name of national security – particularly in the fragile non-democratic states of Africa?"* They questioned the value of data sovereignty in countries where there is little or no trust in government, and pointed out that *"if there is an international shut down, there is no way of protesting, other than through the internet – data can be used where law can't go."*

> "The key players will be the data rich, not the richest - the amount and availability of data, rather than the size of the country, will define multinational treaties and data sovereignty power."
>
> Bangalore workshop

In Singapore, it was observed that *"the key question is how to establish the hierarchy of rights between individuals, citizens, corporates, and the government."* In China, maintaining control of all the data produced by its citizens enables the government to produce its social credit rating, and is used as a way for the state to maintain control. Every citizen has been given a score based on historical behaviour, and for those with low marks, this means restrictions on access to services and freedom of travel, with, at the extreme, passports being cancelled. This level of surveillance extends across all aspects of an individual's life - in Shanghai, we heard that *"all Chinese health data has to be on one of three government-backed Chinese companies' servers by 2020."* In another China discussion, we were informed about the rise of Internet hospitals, which are consolidating millions of health records and enabling the mass identification of individuals with specific characteristics of concern.

The Russian government is also demanding greater access to citizens' private data. Indeed, President Putin has recently introduced a law on "digital sovereignty," which in theory, will let the Kremlin censor or cut off the national internet. In practice, this would be difficult to achieve, as Russian internet companies have servers abroad and would need Western co-operation to do it. So far, Facebook and Google have resisted Russian requests to reveal their users' identities. But the pressure is mounting on them to comply.

"The key question is how to establish the hierarchy of rights between individuals, citizens, corporates, and the government"

Singapore workshop

Data Imperialism

Around the world, we heard concern that multinational companies, predominantly from the US, have built huge empires by treating data as a natural resource that can be extracted and exploited without fair recompense to those who generate it.

In Madrid, the consensus was that *"dominant Western services, built by Western engineers, reflecting Western values, and built on Western data, will increasingly be seen as either imperialist interlopers, irrelevant, or inappropriate in different cultural regions."* Elsewhere, there was widespread pushback against what was seen as Western greed. In conversations in both Nairobi and Johannesburg, the discussions focused on how to ensure that African data is not exploited by international companies as if they were just another natural resource. South Africa, for example, has restricted the sharing of blood samples with US-based companies, like ancestry.com and 23andme, for genetic profiling, because it *"does not want 'cheap' African data to be monetised by others."* In Nairobi, the conversation explored ways to protect African culture. Data sovereignty legislation, they felt, would ensure that *"in the future, we can respect the origins of African cultural data and monetise it ourselves."*[128] They also looked at ways in which to protect African data, by introducing *"appropriate [national] regulation and data transparency to move monetisation forward."* These should have "shared value models and clear reporting frameworks."

In Dakar, there was a call for *"the value of data to be used in the national interest, not only for the benefit of international companies."* Similar views were expressed in Abuja. *"Africa needs clearer policies around data – what is being gathered, why, and by whom."* In Abidjan, there were proposals about greater cooperation between African states: *"as concerns around security continue and the confidence of African developers increases, there is a growing appetite for Ivorians to look after the data that they produce, and become less dependent on Western nations."*

> "Dominant Western services, built by Western engineers, reflecting Western values, and built on Western data, will increasingly be seen as either imperialist interlopers, irrelevant, or inappropriate in different cultural regions."
>
> Madrid workshop

In Johannesburg, where the POPI (Protection of Personal Information) Act regulations came into force in December 2018, it was felt that a regional approach to protect citizens' data should be developed in order to boost the local economy. Students in Pretoria agreed, proposing that *"Africa needs its own servers and its own systems,"* as well as advocating *"'data decolonisation', so that Africa can establish control over the data that is generated within its borders."* Assuming government willingness to invest, they were strong supporters of *"the development of media and regionally specific content, using African data so that it would be more relevant to the local market, which in turn will lead to cheaper services and better products for consumers."*

There was some concern that, in reality, some authoritarian nation states would use demands for 'sovereignty' to enable them to peruse their own totalitarian ends, rather than to protect their citizens from 'foreign' intrusion and exploitation. To limit this risk, it was suggested in Johannesburg, that even if a country imposes data sovereignty legislation, there should be internationally agreed *"data dignity metrics,"* which will allow the monitoring and use of data for the common good, while maintaining the *"dignity of private citizens."* This, they felt, would have the advantage of limiting the potential abuse of power. Failure to achieve clarity around this, they feared, would not only restrict freedom of expression, but border protectionism would *"stifle innovation"* and may well, *"…lead to mistrust in the potential for data to do good, while increasing the risk of large-scale commercial and state corruption."*

The workshop in Sydney was sympathetic to the motivations for data sovereignty: *"you want to be manipulated by your own government – not another one."* However, many agreed that it *"depends on the type of data: Singapore may have tight control of health data, but it is open with commercial data. In Australia, we keep our financial service data sovereign."* Taking the long view, the conclusion was that *"a few mega countries can grow their own independent ecosystems, but most others know they are unable to restrict sharing."*

"It depends on the type of data: Singapore may have tight control of health data, but it is open with commercial data. In Australia, we keep our financial service data sovereign."

Sydney workshop

Elsewhere, although there was recognition that data sovereignty has the potential to have an impact, few in the European or US workshops felt that it would actually happen at scale. In London, which took place after the discussions in Africa, the workshop dismissed the idea of data imperialism as unfounded. Their perspective was that *"data sovereignty is not good, and data flows should be ensured."* Similarly, in a San Francisco discussion, data sovereignty was considered to be an over-reaction; one participant suggested, *"worrying about this is like moving the deckchairs on the Titanic – legislation is 5 years behind what is already happening."* The feeling was that, while other countries may be concerned about data sovereignty, *"in the US we are moving ahead and are more focused on making better use of data."* One comment was that *"it seems as though other countries are using data sovereignty as an excuse for not making progress,"* and *"we have bigger issues to address."*

Implications for Data Value

How data sovereignty is perceived is dependent on a number of different issues and motivations. It is easier to believe that sovereignty is a "good thing" if citizens trust their government to use it to protect their rights and promote their national interests. However, in countries where trust in government is low, data sovereignty regulation could be used to restrict free speech and contact with the outside world. In which case, many would consider it to be a "bad thing."

Size also matters. China, Russia, and India are "big" countries and, arguably, are in a better position to use data sovereignty to their advantage than 'small' ones. Their combined economic clout is certainly significant. Many established Western technology firms are keen to extend access to these profitable markets, as well as those in Africa, which boasts both a youthful population and a rising middle class, so oppose the idea of data sovereignty. Certainly, if the momentum towards data sovereignty continues, a good proportion of future data that is created, may be excluded from the global economy.

It may be that much can be done to limit the very real concerns we heard around the protection of citizen data. Greater trust, understanding, and collaboration between nations is certainly needed. Without this, we can expect even more states will act to constrain trans-national data flows. If this happens, the reaction to the calls for data sovereignty we heard in London and San Francisco seems like a somewhat short-sighted response to a changing political landscape.

> "Worrying about this is like moving the deckchairs on the Titanic – legislation is 5 years behind what is already happening."
>
> San Francisco workshop

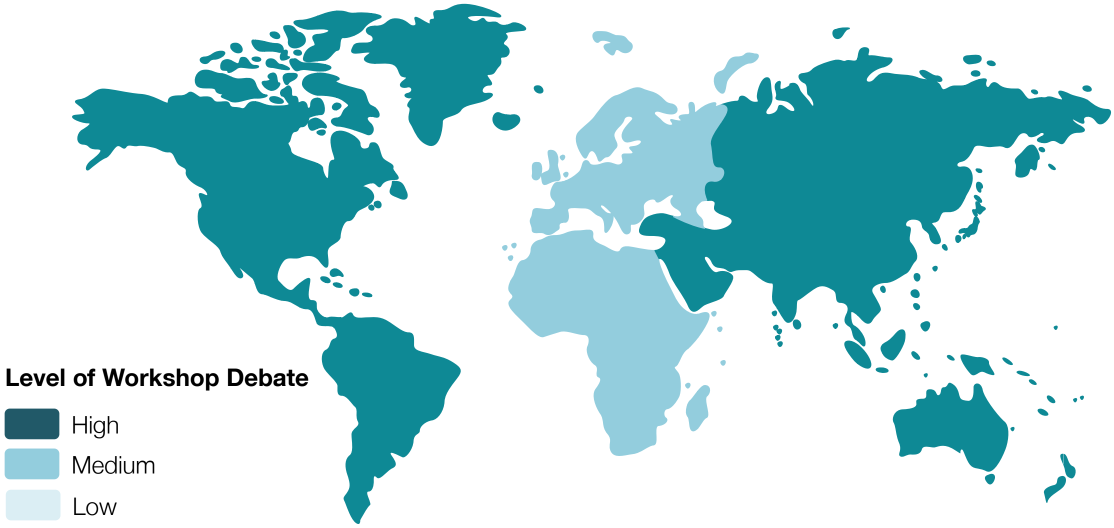
4.9 Data Quality

As we seek better insight, concern about biased, poor, and false data grows. Cleaning and validating data is a social, political, and commercial battleground.

Context

Whether it is basic administration, generating of new insights, making decisions, or organising their implementation, if the data that informs these activities is wrong, the outcome will almost certainly be sub-standard, inefficient, and potentially positively harmful.[129]

The most valuable data must be of good quality. Organisations clearly don't want bad quality data. Organisations that are in complete control of how their data is captured, indexed, and stored are in a better position to ensure quality, but for those that are seeking to combine information from external sources of varied quality and consistency, life can get tricky. That's why 'cleaning' data is big business. The question our workshops wanted to know is this: are we really rising to the challenge of poor data quality? If it is 'dirty', then all sorts of automated policies, investment, and even social decision-making may go astray; think of misaligned government funding due to inaccurate census data , children being wrongly removed from their parents because of an error in social service algorithms, or more mundanely, the duping of users on dating apps.



The Challenge

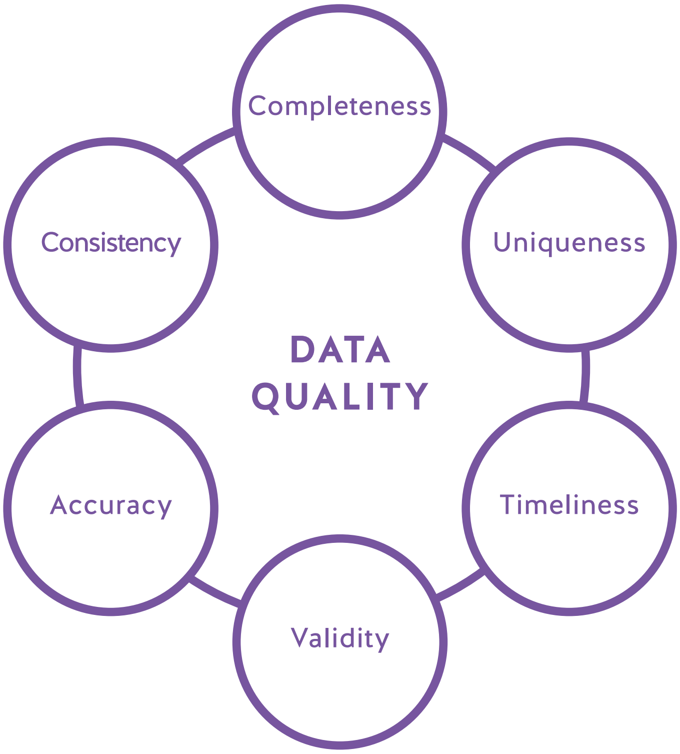
Our workshops distinguished between three types of low-quality data: poor, biased, and false.

- **Poor data** is incomplete, out of date, misattributed, misprocessed, or simply wrong. There are multiple reasons for this – from data entered into the wrong columns, to duplicate data or inconsistent entry, misspellings, and so on.

- **Biased data** refers more to sets of data that create a picture of something. This is now highly topical, as machine learning algorithms rely on these data sets to generate predictions and make decisions. A biased data set may simply reflect biases that already exist in society, such as the fact that most top jobs are held by middle-aged white men. But it can also reflect the values of coders, results from survey questions that are constructed with a particular slant, or arise from process/design issues such as data that is misreported in categorical groupings, non-random selections when sampling, or systematic measurement errors.

“We should focus on algorithmic awareness – NOT the elimination of bias, because we need to know why data was created.”

Toronto workshop



Data Quality: Key Dimensions

- **False data** is deliberately created to be inaccurate or misleading - although it may well seem to be high quality and from verified sources. This has also become highly topical when false information is deliberately shared on social media, but it's also generated when individuals deliberately input false data because they don't trust the organisations that they are sharing the data with.

All three of these are now escalating in both scale and impact. They can render some data sets hard or impossible to use, and if not identified, corrected, and isolated, they end up polluting good data sets and the decisions based on them.

Managing Poor Data

Clarifying whether or not information is accurate is as yet largely a human, lengthy, and expensive task, although AI and wider automation is beginning to help. It explains why, in 2018, the global pharmaceutical company Roche was prepared to pay $1.9bn for Flatiron Health, a start-up which can clean clinical information with a particular focus on cancer. The capability that Roche valued in particular here, was the 'human-mediated extraction.'[130]

Many companies are grappling with how best to achieve better quality data, quickly, and at low cost. Some are focusing on improving data capture, and others are looking at ways to correct the errors. One option is only to use the good data and remove the 'bad' - but within this, it is important to define what 'good data' is. From a health perspective, for example, there is an emergent perspective that just because data is not of medical quality, does not mean it has no value. It's a question of what information is appropriate. This is a time-consuming and expensive exercise - 80% of data scientists' time is spent cleaning data.[131]

Biased Data

Most concerns about biased data focus on the data sets used to train and refine automated algorithms. In Washington DC, the case of an Amazon recruitment programme was discussed. Amazon's computer models were trained to vet applicants, by observing patterns in resumés submitted to the company over a 10-year period. Most came from men, a reflection of male dominance across the tech industry. The result was that the self-learning system taught itself that male candidates were preferable. There is no guarantee that other ways of sorting candidates that could prove discriminatory might occur – indeed, the Amazon algorithms allegedly also favoured men who played lacrosse and were called Jared.[132],[133] Amazon has since scrapped the project, but it's a good example of how difficult bias is to manage. Considering the fact that around 55 percent of US human resources managers expect to use AI within the next five years, this is extremely concerning in just this limited arena of recruitment.[134]

> There is a feedback loop – false data leads to low trust leads to false data."
>
> Hong Kong workshop

Another example discussed by the workshops was the claim that the AI algorithms currently used to decide who goes to jail are getting it wrong, due to their dependence on historical data.[135] In 2016, courtrooms in the US adopted risk assessment tools to generate a "recidivism score." This is decided by machine learning algorithms which use historical data to pick out the patterns associated with crime, to produce a single number estimating the likelihood of a prisoner reoffending. A judge then factors this into a prisoner's rehabilitation, or the duration of their sentence. This means populations that have historically been targeted by law enforcement, such as low-income and minority communities, are at risk of being given high recidivism scores. In turn, this means the algorithm could amplify embedded biases and generate even more bias to continue the cycle. Because most risk assessment algorithms are proprietary, it's also impossible to interrogate their decisions or hold them accountable.

Some in our workshops worried about a lack of diversity in the technology industry, and how this is impacting the roll-out of AI. Only 22% of AI professionals globally are female, for example. The more algorithms determine social outcomes, the more software development teams need to ensure diversity, to spot when data biases are skewing the decisions. Although there are increasing calls for more female coders, inventors, and investors, so that technology companies can more accurately reflect society, change is taking some time to come into effect. Some suspect there is a negative network effect, that the small share of women in the field discourages others from choosing it as a course of study. Employers might not be able to undo societies' gender bias single-handedly, but they can take mitigating steps, for example, by building tech skills into schemes for women returning from career breaks, and providing greater transparency around pay and opportunity.

AI can help expose truth inside messy data sets, and will be used to great benefit in multiple different ways. But it poses potential risks as well as opportunities. A frequent topic of conversation in our workshops was the need for business leaders to establish a transparent process for monitoring the ethical behaviour of their AI systems. This could include common standards for training data for algorithm building and real-world applications. Part of the solution may also lie in regulation, including hefty fines for non-compliance, plus a concerted effort to ensure that there is greater public awareness of the potential issues.

"Labelling helps to identify truth, and perhaps branded news is a way to help the public identify responsible channels."

Mexico City workshop

False Data

‘Fake news’ is now big news, and a major headache for both tech companies and governments. There is a large and growing market for exploiting the vulnerabilities of the digital world, and some very smart, sometimes unscrupulous, players capable of supplying it. Such is the sophistication of some of the false information, that it can be almost impossible to identify it. Campaigners are pushing governments to develop tougher regulation to better protect civil society. Some are considering adopting tighter international protocols, such as those used to restrict the arms trade.

Much of this debate is beyond the scope of this report, but fake news is not the only form of false data. In our Washington DC discussion, for example, it was pointed out that around 20% of US Census data is thought to be inaccurate, mostly because citizens providing the information fear how government will react if they tell the truth. Officials for the US Census are not allowed to compensate for this, despite knowing that around 20% of the key data sets are wrong. Here, the inaccurate data is largely driven by public fear of government intervention, and some communities; often those in most need of support, such as the poor, recent immigrants, and the elderly, intentionally enter false data for personal information like income, health, and age. The unfortunate irony is, without census data to identify need, policy makers are unable to justify additional funds to support the very people who are not disclosing the correct data. When we discussed this a few days later in Toronto, there was an acknowledgement of similar statistical issues, but officials in Canada are allowed to ‘correct’ known data sets before they lead to ineffective policy and misguided activity.

It may not matter much if we give false email addresses to access public wi-fi, or when shopping for a new pair of shoes, but it does when there are important consequences. In Nigeria, such is the level of mistrust, that few give government agencies accurate information or correct emails. As was observed in Hong Kong, *“there is a feedback loop – false data leads to low trust leads to false data.”* The challenge comes when data has to be real enough to authenticate an individual, a machine, or a location. The principle of digital identity is important here, and has recently been explored in detail in another Future Agenda project.[136]

> We must be careful not to make the perfect the enemy of the good. Just because you identify bias, doesn’t mean it is inherently flawed.”
>
> Santiago workshop

What We Heard

Across the world, there was deep concern about the provenance and accuracy of information served up to individuals on social media, and the role of algorithms in this. In Bangalore, there was a *"growing concern how to monitor and control social media, to limit the manipulation of consumers by corporates and other organisations."* In Mexico City, a concern was that *"discrimination will be a big issue – particularly as facial recognition becomes more prevalent."* Singapore was more optimistic, *"AI will become more sophisticated around helping identify fraudsters, but we are not sure if it will be fast enough to identify fake news before it gets out …. Labelling helps to identify truth, and perhaps branded news is a way to help the public identify responsible channels."*

Many mentioned the seemingly blind confidence that there is in the accuracy of algorithms, and observed that even clean data can be biased. In Madrid, several highlighted that *"biased data is increasingly powering automated choices."*[137] In Canada, the suggestion was that bias should be managed through *"algorithmic awareness – NOT the elimination of bias, because we need to know why data was created."*

In Santiago, it was suggested that *"we need to work out if it is at all possible to measure bias."* Is it possible to develop a quality mark or traffic lights system for data, showing whether or not it is free from bias, moderately impacted, or severely compromised? However, in Hong Kong, the view was that *"we must be careful not to make the perfect the enemy of the good. Just because you identify bias, doesn't mean it is inherently flawed."* That said, in the same workshop, it was acknowledged that *"there is a risk that bias will be programmed into AI, which will lead to continuous marginalisation of individuals."* What is certainly the case is that, given machine learning is retrospective, the more we rely on machine learning, the more existing bias can potentially be entrenched.

One suggested solution was to *"consider developing strong regulation frameworks that require harm-based assessments of the application of data, and continues to monitor real-world harm."* Some in Hong Kong also wondered, *"should there be a world data organisation that can establish principles around bias?"* There, it was also proposed that *"the key question is which institution will be able to identify and exclude bias, both of input and output. Do facts need to be baked into this?"* Additionally, "*it is difficult at this point to identify whether the outcome will be positive or negative. There are plenty of examples of bias in China, around many issues – from mortgages and AIDS, to sentencing, diversity, and inclusion - and it is difficult to see how individuals have been categorised."*

Another thought in Sydney was that *"bias within data could lead to data inequality."* Looking forward to 2030 in London, some agreed, and saw that we will see *"more social exclusion in terms of in-built bias of automated process, networks, and creators."*

> "The challenge will be to extend legal protection over all aspects of life; for example, the wide range of potential cases which may have a discriminatory outcome that affect people or third parties."
>
> Santiago workshop

In Nigeria, the problem is more societal, as *"corruption and lack of trust in the system is driving the collection of inaccurate and fake data."* People intentionally give false information to the government and companies alike. This *"makes our databases unreliable, as citizens choose not to share accurate information."* Other than eliminating corruption, suggestions of how to overcome this focused on better public education to *"build a wider understanding of the benefits of data sharing."*

Looking Forward

The assessment from Copenhagen was that *"at the core, we need to have objective views of what is good data - but being clear on what is this 'objectivity' is a central question… a big issue for the future is who will decide."* They acknowledged that *"for public consensus, we may have to go through a period of more data anarchy and more fake data, before people change."*

The final workshop in Santiago agreed, *"between today and 2030, existing regulation needs to be updated. Policy makers need to be trained on this and so be able to agree on the appropriate use of algorithms, and to better identify instances of bias as a start."* We also need to consider taxonomy and how we classify algorithms; *"the challenge will be to extend legal protection over all aspects of life; for example, the wide range of potential cases which may have a discriminatory outcome that affect people or third parties."*

Some argued for a *"World Data Organisation, which can establish principles about quality and bias."*[138] However, controlling the spread of fake data is more challenging. It contaminates good data sets, distorts our perspective, and gradually misleads our actions.

Implications for Data Value

If our data in the future is to be useable, never mind of value to society and commerce alike, then it has to be reliable. The view from those who discussed this in our workshops was that society hasn't yet acknowledged either the scale or the complexity of this problem. Improved transparency and accountability processes can help, but it is also about underlying data quality. However, acknowledging and managing raw and contaminated data alongside cleaned data, is a necessary shift that many will need to accommodate. For most requirements, some inaccuracies can be managed, but certainly not all – think of clinical trials results, for example. Global consensus around acceptable levels of accuracy would help here, alongside an institution which can set standards and then arbitrate should disagreement arise.

It is clear that organisations that can efficiently, quickly, and accurately clean data are already adding value, and that high quality, structured data sets will continue to command a premium. As data becomes even more integrated into the operations of our economy and society, it is increasingly important to ensure and maintain its quality.

> "Corruption and lack of trust in the system is driving the collection of inaccurate and fake data."
>
> Abuja workshop

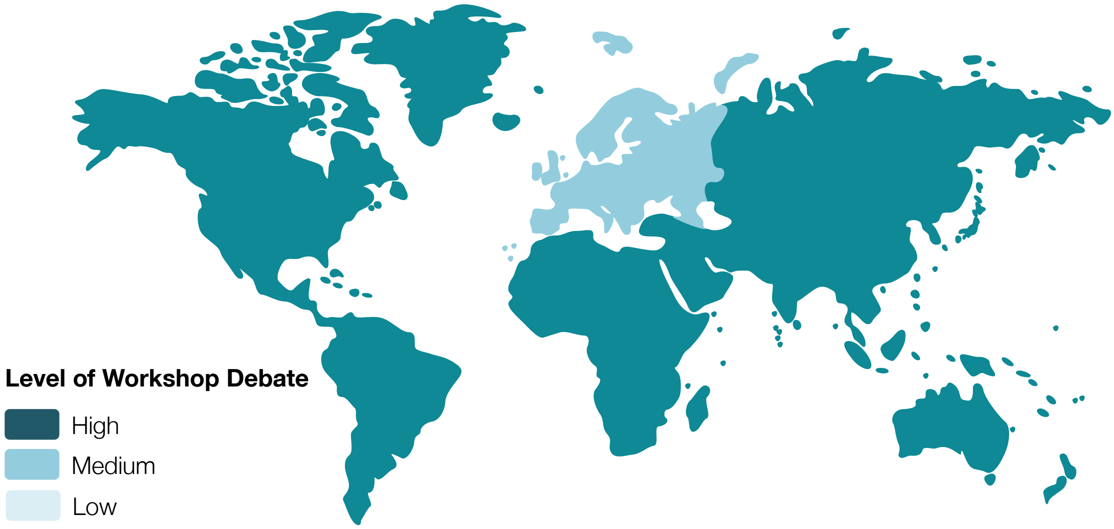
4.10 A Question of Ethics

Ethical data use grows as a concern, but we struggle to agree a global approach. Sectors set their own standards and try to align on some common principles.

Context

In the early days of the data revolution, it seems that many of those most deeply involved in data - and most at the forefront of how data is collected and used - gave the social implications of what they were working on very little thought. But how data is used and controlled raises many ethical concerns. Ethics is about the moral principles one adopts to guide one's actions and behaviours. It is about how people treat other people: whether their motives and intentions are benign, indifferent, or hostile; whether the effects they have on others is harmful or beneficial. Participants in our workshops often suggested that, in the race to collect, store, and use data, and the commercial opportunities that this creates, ethics have sometimes been sidelined.

Managing ethical complexities in an age of Big data can be tricky, given that little is covered by existing law, but there is growing recognition, particularly amongst governments and data organisations, but also more widely in civil society, that it is important. There is also growing recognition that a failure to rise to this challenge, risks undermining public trust, and confidence in the data industry as a whole.[139]



Recognising that there is an ethical dimension to data collection and use is one thing. Agreeing what the appropriate ethical code should be is another, especially given:

- The multiple different uses of data across multiple different industries (from medicine to finance, routine administration to decisions about entitlements, credit or benefits, as well as multiple applications of AI to generate insights and automate decision-making)

- The wide range of potential ethical impacts of data use (covering, for example, whether current shares of financial and other benefits are fair, the extent and implications of pervasive surveillance, or whether particular uses of data are creating or exacerbating unfair discrimination)

- The disparate nature of key stakeholders (for-profit corporations, governments, academic researchers, individuals as citizens and consumers)

- The different norms and values adopted by different cultures and societies

- The different circumstances, needs, and priorities of these different cultures and societies.

What We Heard

From what we heard in our workshops, there is little doubt that, in the broad sense data, ethics are becoming a key part of the data debate. The accelerating development and media coverage of AI is very much amplifying the challenge.[140] In 2018, Google - widely regarded as having the most advanced AI - published an ethical framework outlined by its AI principles, the first of which focuses on being socially beneficial.[141] Several workshops also highlighted Salesforce's appointment of its first Chief Ethical and Humane Use Officer, as a signal of wider change.[142] The company is striving to make the ethical use of technology a source of differentiation. Whether this can also be a source of competitive advantage, in a way similar to how some are positioning themselves around privacy, is not yet clear. But as more companies push data ethics forward in tandem with calls for action from wider society, momentum for action is clearly building. In the meantime, multiple companies are seeking to protect themselves from risk by setting up ethics committees to oversee best practice.[143]

> "This is about leverage – ethics don't win against market access. The reality is that commercial benefit wins over global ethics."
>
> Bangkok workshop

While for many, the ethics of the value of data and the ethics of data use become implicitly interlinked, key areas of debate in our workshops were:

- Ethics versus profit
- Cultural differences
- Ethics and regulation
- Respecting data rights
- Flexible framework

Ethics vs Profit

In Washington DC, we were reminded that *"ethics are how you behave when no one is looking: it's not what you can do, it's what you should do."* This is not always as easy as it sounds. While not always in conflict, companies are having to make difficult choices about their ethical and commercial approach. A balance needs to be struck so that they can demonstrate responsible and ethical behaviour, while protecting and promoting commercial or strategic interests with the potential for profit and other considerations to override ethics.[144] In our Bangkok workshop, for example, there was a notable anecdote about Apple, which now complies with China's requirements for data localisation.[145] The discussion concluded that *"this is about leverage – look at Apple's deal for China: Apple caved in – ethics don't win against market access. The reality is that commercial benefit wins over global ethics."*

Cultural Differences

We often heard that any ethical framework around the value of data must, like the wider ethics landscape, acknowledge significant cultural differences. Those in Johannesburg asked, *"how do we incorporate the enormous variety in moral and ethical beliefs between different cultures?"* Discussions in Manilla argued that *"ethics are inherently cultural and relative, and therefore inherently difficult to build into universal frameworks. If any universal framework were developed, it is highly likely to come from the West, where the data debates and infrastructure are more mature, and where the big data companies reside. This would be a new kind of cultural imposition on places like the Philippines."*

In Singapore, they said that there is *"a general assumption that we do not have a common language around data ethics. This is complicated by the richness of cultural differences, and diversity of legal traditions."* It also highlighted potential *"conflict between East and West philosophy,"* and questioned how things may change if, for instance, TenCent becomes as dominant as Google. Would a Chinese-driven view of ethics around data use and value be significantly different from the California perspective? Probably. In San Francisco, there was recognition that data ethics as a whole *"could well develop with alternative views globally – one driven by Western approaches and the other Chinese."*

> "Data ethics could well develop with alternative views globally – one driven by Western approaches and the other Chinese"
>
> San Francisco workshop

In Madrid, analogies were drawn with lessons from religion: *"any religion has a common set of values, but with a data religion (data-ism), the commonality is not there. There is a need to recognise that data is not truth - it just presents information in different ways, and we must learn to recognise the bias, or lose our freedom of choice."* Just as ethics generally vary across religions and cultures, so will views around ethical sharing of value.

Ethics and Regulation

The pros and cons of self-regulation vs government regulation were frequently discussed, particularly perhaps, due to the revelations around Cambridge Analytica and Facebook. Many were concerned that the current model, where individual companies self-manage their own behaviours, has failed, and that therefore, regulation is needed to limit the risk of unethical behaviours by some businesses.[146] The debate primarily focussed on whether regulation by industry sectors would be sufficient, or if central government regulation would be a better alternative. In Bangalore, it was observed that *"the law alone is not enough,"* and that even with regulation, there is a moral obligation for businesses and those who work within them, to behave with integrity. The Bangkok discussion looked at it from a different angle, suggesting national regulation, rather than corporate interest, was likely to have a stronger moral compass; *"ethics are inter-twined with regulation."* In order to balance the requirement to protect citizens and also maintain a competitive environment for business, they acknowledged that a range of regulatory approaches may need to be considered, including cross-sector collaborations, similar to the Partnership on AI.[147] One idea that was explored was the need for a 'Hippocratic Oath' for data scientists. Just as medical professionals pledge to "do no harm," individuals working with data should sign and abide by a set of common principles.

In Mexico City, the consensus was that *"we see that there will be two different approaches to the development of data ethics – public and private. It is the argument between regulation and self-regulation, and, between these, we may see different communities driving action."* In Sydney, it was felt that change is necessary, and *"some will be driven by company frameworks, some by self-regulation, and some by central regulation."* Looking ahead, one suggestion voiced in Washington DC was that the self-regulation route would only be effective if it followed *"a multi-stakeholder approach, which will establish principles and standards."*

Those in Bogota largely supported self-regulation. Although recognising the difficulties, there was optimism that *"with co-operation, there will be agreement about base standards, and self-regulation will then be able to establish an ethical framework which can be applied across all sectors."*

> "With co-operation, there will be agreement about base standards, and self-regulation will then be able to establish an ethical framework which can be applied across all sectors."
>
> Bogota workshop

Respecting Data Rights

There was much debate in our workshops about how to deal with ethics in markets in which there is little or no regulation, and where, for example, the concept of digital rights, which is well established in Europe, is poorly understood. Those in South Africa felt that in the first instance, as with human rights and cultural views of data value, acknowledgment of and respect for data rights *"are likely to be highly regionalised."* However, if we move in the direction of "data informing social development and public good," then we *"will need a mechanism by which the level of trust in the intention to use data for a common good, can be measured and monitored."*

Over in Manilla, they said that *"the public is moving from a position in which they are relatively unaware of their rights at all, let alone digital and data rights, to a more informed landscape."* The view was that, as data literacy increases and public understanding of the value of data grows, so too will their expectations that companies will be required to behave to prescribed ethical standards. Furthermore, *"we may need to consider completely new kinds of rights. Algorithms and AI will extend the need for rights to entirely new demands."*

This view was reflected in Mexico City, where they felt very strongly that *"over time, sufficient controls will be maintained to ensure that established ethical practices are not lost."* There will be *"legislation and increased governance to maintain innovation opportunities within the digital economy, without jeopardising human rights."*

Implications for Data Value

Where do all these views align? Despite the evident cultural differences, the common hope expressed in a number of workshops, is for some sort of global framework, or at least a set of principles for data ethics. If these are to be effective, then they will not only be designed to improve understanding, but they will also drive new behaviours. It's a good aspiration to have. However, given the cultural, political, and technological challenges, most recognised it is unlikely that a single global model will emerge any time soon.

> "We will need a mechanism by which the level of trust in the intention to use data for a common good, can be measured and monitored."
>
> Pretoria workshop

In Sydney, the call was to *"establish a framework and a set of principles. These need to be universal, flexible, and forward-looking. Individuals and organisations need to be able to assert and change their rights. They need to cover the collection, storage, and use of data – as well as the risks. They also need to cover the relationships (who, what, and how)."* In Singapore, the call was for a *"universal framework."* However, others in Manilla questioned *"the idea of any imminent universal standards."* Canadian experts agreed, and pointed out that *"there is no universal framework for this. But different systems/views have got to be on the same level, otherwise organisations will move to choose the best/easiest/most lenient/less enforced ethics jurisdiction, in the way they do for tax. So, there needs to be as much collaboration as possible; but this will not be possible globally."* In India, the view was that, *"the desired end state is an ethics framework …. But it should be based on existing cultural principles."*

Managing data is difficult, and developing practical solutions to ethical problems is also difficult. There is nothing easy about the interface between these two. Small surprise perhaps that our discussions did not reveal any magical solution to the challenges. There are none. However, there was widespread consensus in our workshops that the only way to ensure the sustainable value of the data that is generated, collated, processed, and monetised, is to work towards universal agreement around the ethical principles of its use.

How to achieve this is still under debate. Both "top-down" regulations, as well as "grassroots" efforts, seem to be raising more questions than answers about how we might define fairness, combat bias, and create ethics guidelines in data science and AI. Looking ahead, ensuring a proactive and meaningful approach to data ethics may well involve greater transparency than we see today, and greater expert engagement. For business, this may mean short-term compromises in efficiency and effectiveness, but few would disagree that in the long term, it is certainly worthwhile.

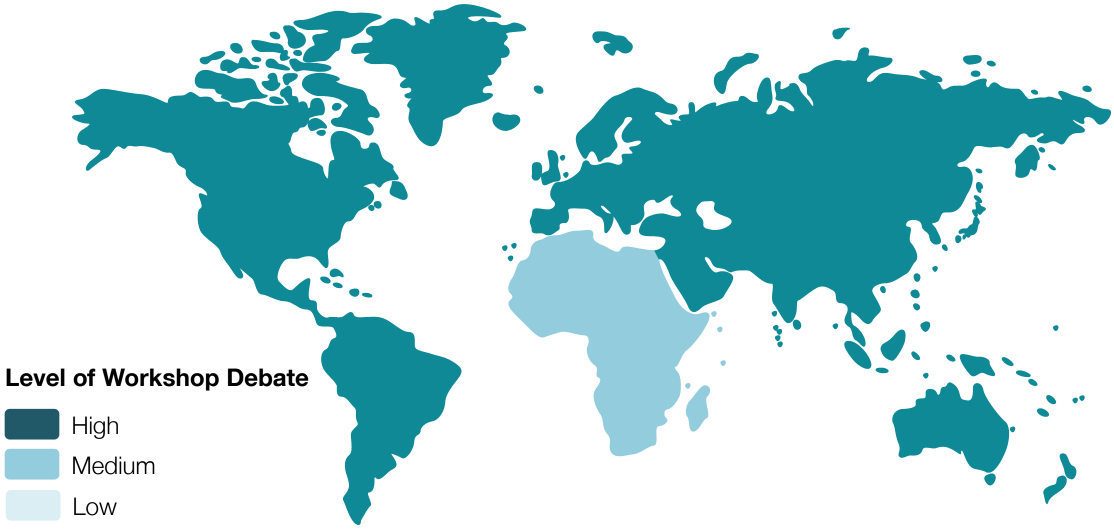
> "There is no universal framework for this. But different systems/views have got to be on the same level, otherwise organisations will move to choose the best/easiest/most lenient/less enforced ethics jurisdiction, in the way they do for tax. So, there needs to be as much collaboration as possible; but this will not be possible globally."
>
> Toronto workshop

4.11 The Organisational Response

The management of data requires a 21st not a 19th century approach to business. With digital as the norm, we move on from principles based on physical products.

Context

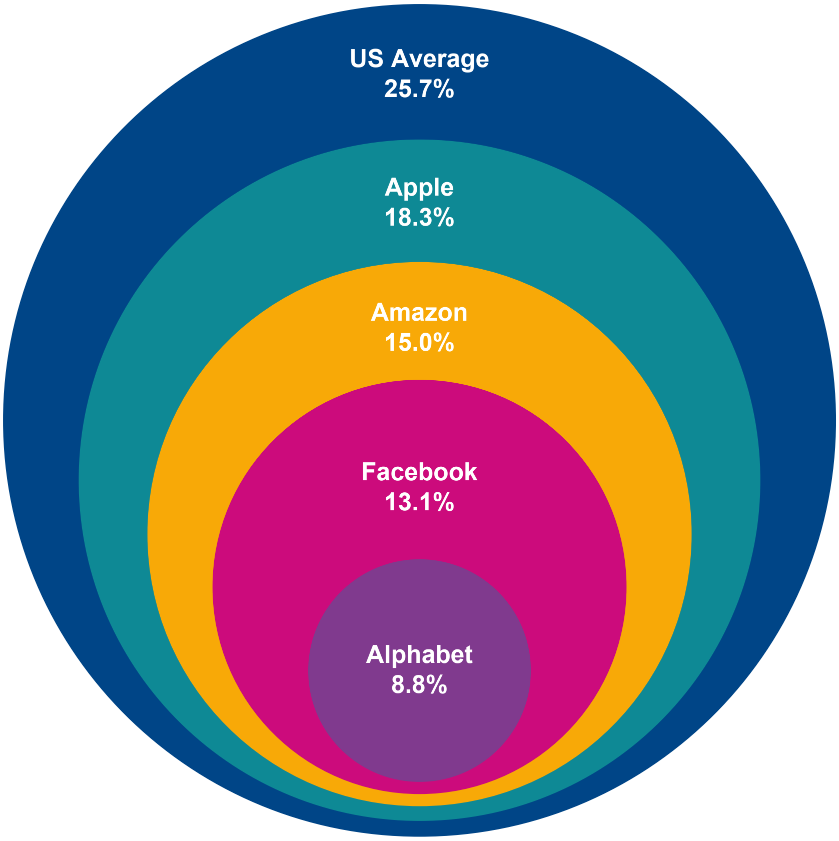
It is clear that many of today's digitally-driven organisations are significantly unlike traditional businesses. Multiple corporate leaders and a plethora of fast-growing unicorns are all seeking to deliver significant change, mostly via creating value from data. But questions are being raised about how these companies function, what their values are, and how their impact and influence is measured and held to account. Although Big Tech has replaced big oil, big steel, big banks, and the big 4 automotive firms as the world's most powerful companies, many see that the way they operate is not comparable. While Google and Amazon may have the same legal structure as other corporations, such as GM, Coca-Cola, and JP Morgan, the way they behave internally, and function externally, is meaningfully different. The growing perception is that existing regulatory tools and business norms are outdated, inadequate, or insufficient, in light of their changing business models.[148] Given that over the next decade, most organisations will gradually become data companies to a greater or lesser extent, many believe that new metrics are needed to manage them and judge their performance.[149]



A Different Set of Rules

Over the past ten years or so, the new data-rich organisations that have expanded, have done so in ways that companies in previous eras could not. Recent research has highlighted several reasons for this:[150]

- As software has replaced hardware, the cost of leading digital innovation has dramatically declined, allowing relatively small investments to yield large payoffs.
- Online platforms increasingly control vast amounts of valuable data, which they gather largely for free from their customers. The owners of these platforms enjoy substantial advantage from access to their customers' data, which is very difficult for others to replicate.
- The speed of change is now so fast that many regulators are behind the curve and unable to jump ahead of the innovators.

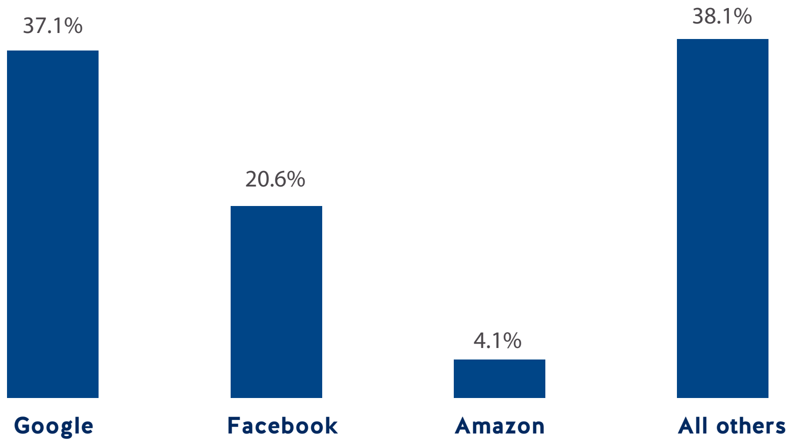


Effective US Corporate Tax Rates (2018)

One of the consequences of this, is that the core parameters – legal personality, limited liability, transferable shares, and even the concept of intellectual property - that have set the operating landscape for most companies for the last century, are no longer fully fit for purpose. For instance, many in our workshops argued that there is a fundamental difference between the economics of production of physical vs. digital products. Making things of value from resources and materials which have a finite supply, and therefore an implied cost, is completely different to making things from data, which is an almost unlimited raw material - the cost of creation and replication of which is fast falling to zero. Research by academics such as Mariana Mazzucato and Shoshana Zuboff, and the work of the Future of the Corporation project, are exploring potential new paradigms here, but as yet, there is no clear consensus on how this should be addressed.[151],[152]

Moving Goalposts

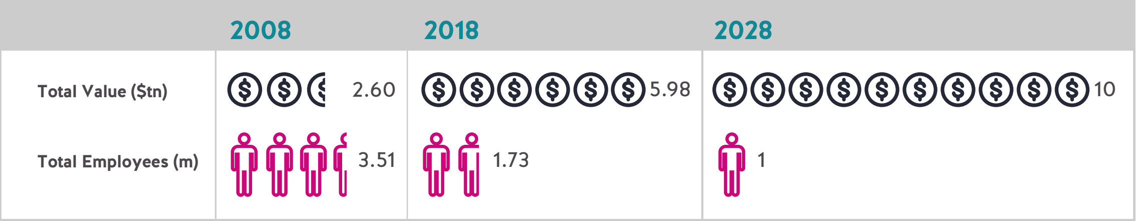
Meantime, the size and scale of the modern corporation is changing. In 1975, 17% of the market value of the S&P 500 was based on intangibles; by 2015, this had flipped to 84%. Many leading companies are now focussing on innovating to build IP, brand value, and other key assets, and up to 90% of the value of some firms is correspondingly assigned to intangible assets. Data is at the heart of this transformation. In 2008, the world's ten most valuable companies were worth a combined $3.5tn, and employed a total of over 3.5 million people. By 2018, the top ten companies were worth twice as much, but only had 50% of the number of total full-time employees. As new technology enables higher revenue per employee, then looking ahead another ten years, it is possible that the top ten companies will be worth over $10tn, but employ only 1m people. There are several key implications:



Share of US Digital Ad Revenues (2018)

- Economic power: There is an accelerating concentration of economic power within organisations whose core businesses are increasingly built on data. WEF analysis suggests that up to $2.3tn, or 40% of the total value of the top 20 global companies' current market capitalisation, could be associated with the data they own, access, and monetise. To give some context, that is more than the total GDP of Italy - the world's seventh largest economy. Furthermore, many in our workshops and beyond, considered that some digital firms "face no limits in ability to scale – the bigger they are, the bigger they are likely to grow."[153] This raises many questions around both the potential scale and influence of a corporation.

- Unequal wealth distribution: There is the associated issue of concentration of wealth for employees, and their potential disconnection from wider society. Although external shareholders clearly gain from a profitable organisation, many of the major digital companies have significant employee options and shareholdings, which have grown substantially. Moreover, the average income per employee of the top 5 companies (Apple, Amazon, Alphabet, Microsoft, and Facebook) in 2017/18 was $1.4m. With many employees now multi-millionaires, some question whether the majority understand what "normal" life is like for most citizens, and because of this, have less empathy with them. This is not just a West Coast issue. In the UK, Cambridge, the home of corporate research labs and multiple major start-ups, is now the city with the highest level of inequality – largely due to its success over the past 20 years, driving wealth into the hands of a few but not all.[154] There is a growing risk of those working for and running the world's most powerful organisations fast becoming disconnected from the society from which they earn their incomes.



	2008	2018	2028
Total Value ($tn)	2.60	5.98	10
Total Employees (m)	3.51	1.73	1

Smaller Big Companies - Value and Employees of Top 10 Companies Globally

- Low Tax: The way that many of the world's data rich companies are being managed, is frequently (and quite legally) minimising their tax liabilities. In previous generations, where manufacturing was the dominant industry, the production of goods, sales, and associated taxation was largely national. Even within the services sector, the co-location of human resources and much of the corporate activity, has supported regional tax income. In 2017, the UK Financial Services sector contributed £72bn, or 11% of total government receipts, with corporation taxes accounting for £12bn.[155] However, in 2018, compared to a standard US tax rate of 21%, Apple paid an effective tax rate of 18.3%, Amazon 15.0%, Facebook 13.1%, and Alphabet only 8.8%.[156] Many in our workshops felt that this was a poor reflection of their overall contribution to society.

As trust in Big Tech has declined, the structures and practices of several companies have come under particular scrutiny. As a result, their influence is clearly in the spotlight, and some face a regulatory effort to curb their dominance.[157] The EU has been leading here, but now India and some in the US are also calling for change.[158] There are a number of ways in which this can be addressed. Democrat and Presidential candidate Elizabeth Warren, for example, is calling to break-up Big Tech; others are seeking to curb their power by sharing data with other firms, and making it easier for users to switch to competitors.[159]

The Future of the Corporation

Looking ahead, many in our workshops felt that there is a need to consider how a future corporation, tech or otherwise, should function, not just economically, but how it can contribute to society and whether its role should go beyond that of a profit-making machine for its employees and shareholders. Fifty years after many Anglo-Saxon companies subscribed to the Milton Friedman view that the attention of a company should be to maximise shareholder returns, and that to pursue anything other than (legal) profit would be "pure and unadulterated socialism," there is change in the air.[160] Friedman's 1970 NYT article, arguing that the social responsibility of business is to increase profits, is now seen by many, but certainly not all, as setting a false direction that has led to the generation of wealth for investors and executives, but at a cost to employees, customers, the environment, and wider society.[161] Led by a number of high-profile pioneers such as companies including Patagonia, Unilever, and Virgin, a growing range of businesses are already adopting social purpose that complements their commercial purpose. Indeed, in August 2019, the largest US business group, the Business Roundtable, replaced its long-held view that maximising shareholder value is the defining corporate goal, with a more inclusive vision that takes account of other stakeholders.[162] It will be interesting to see how the data companies adapt to this.

> "We are close to a data oligopoly with too much control in the hands of the few."
>
> San Francisco workshop

Organisation 4.0

Several in our workshops suggested that there may be more viable alternatives to the corporate form within the next decade. We may well even see a different type of legal entity emerging for the data-driven organisation. New initiatives include hybrid forms, such as public benefit corporations; these are very much orientated towards having a strong social purpose. Others point to the previously controversial dual class share structure adopted by Google at IPO, and since used by many other tech companies. It allows entrepreneurs to control the corporation, without owning the majority of the cash flow rights. This is now so popular that stock exchanges have changed their listing rules to allow tech firms with differential voting structures to list their shares.

Looking ahead, we may well see the emergence of two separate systems for companies with different types of structure, governance, and regulation; one system for traditional product and service companies and the other for primarily data intensive firms. If there is a widening gap between two increasingly dissimilar and disconnected economies, governments and stock exchanges may need to set them apart from each other. This could, for example, be an evolution of the NASDAQ and Dow exchanges in the US. There may be different approaches for governance, for taxation, for research funding, for public support, and also for company valuation.

What We Heard

In our South African discussion, it was suggested that *“data will mean a whole new set of corporate metrics,”* while in Sydney, several felt that *“in the future, the Big Tech firms will have all the power.”* With data driving ever greater power and influence for those that control it, how companies are structured, focused, governed, and held accountable, may be about to change dramatically. In San Francisco, they said, *“we are close to a data oligopoly with too much control in the hands of the few.”*

Fundamentally, some see that there has been a power shift from government, society, and multinational corporations, to the transnational, global digital firms. From Jakarta and Bangkok, to Washington DC, Bogota, and Mexico City, we consistently heard that *“data is power,”* while in Frankfurt, the view was that *“those who hold the data hold the power.”* Our London discussion raised questions on power and agency, such as *“who has the power? How is it accountable?”* Moreover, it was suggested that *“data creates power, shapes the wielding of power, the balance of power, and the accountability of power.”* Many agree that this accountability has been sorely lacking over the past few years, and are supportive of greater regulatory action.

> “As we see new actors whose profits exceed the income of most nations, they will wield even greater power...this power may not be accountable and therefore is potentially very dangerous.”
>
> Hong Kong workshop

A forward-looking perspective from our discussions was that, *"as companies' awareness of their power changes, we will start to see increased leverage of power over data flows."* Potentially, as we *"move from self-regulation to trusted regulators, with clear demarcation of rights,"* questions will emerge around how power can be divested. In Mexico City, the expectation was that over the next few years, *"algorithms will become ubiquitous, and the companies that operate them will have little interest in the social impact that they may have."* As a response in Frankfurt, several proposed that *"we need more transparent algorithms,"* as *"we do not question the decisions that machines made for us."* Moreover, *"critical algorithms will be regulated."*

Some in San Francisco proposed that we may well see *"algorithmic regulation to address the issues that cannot be regulated by humans."* However, *"algorithmic governance may well enable the associated companies to generate more revenue with even less human capital."* The consensus in Copenhagen was that, for most companies, *"CXO understanding of data value will change,"* while in Hong Kong, it was added that many *"institutions are out of sync,"* and this has to change; *"as we see new actors whose profits exceed the income of most nations, they will wield even greater power."* Indeed, *"this power may not be accountable and therefore is potentially very dangerous."*

While some of the above shifts were in the background for our value of data discussions, there were multiple mentions of how, for digital companies, these may provide extra challenges. For instance, in Jakarta, it was suggested that *"we will need to look beyond the purpose of the company,"* as data can be shared and used for wider impact than many other assets.

In the San Francisco workshop, one proposal was that *"access to the truly valuable data is in the hands of a few companies,"* and so *"tech firms become the trusted source of data and services, including social services and healthcare."* Furthermore, we may soon see *"government ceding the running of many public services to more informed and capable private companies."*

> "We will need to look beyond the purpose of the company"
>
> Frankfurt workshop

Implications for Data Value

The whole basis upon how corporate entities behave, develop their cultures, are judged by society, and are rewarded by the markets, is evidently changing in some sectors. How, why, and where financial recompense is attributed, is being questioned equally by academics, government, and, in some areas, media. How one company can be worth $1tn and employ only 100,000 will be increasingly contrasted with those that are valued less financially, and yet employ more people. Data-driven companies and the digital economy are clearly different from the more tangible product and services economies, but they are currently being judged by the same parameters and have become uncomfortable bedfellows.

As power shifts, so does value – this is nothing new – but the norms by which one company and its performance are compared to another, are under stress. Monopolistic behaviour aside, traditional means of judging value for shareholders, against value for society's wider stakeholders, are changing: The current research on the Future of the Corporation is just one of several programmes seeking to propose new ways for firms to be managed, monitored, and valued.[163] There are significant implications for data-driven companies. Expect greater scrutiny of their corporate values, their behaviours, more transparent reporting, and changes in the way they are taxed. Some organisations will be proactive, acknowledge the need to change, and try to manage a more equitable distribution of profits and impact. Others may take a more defensive stance. Beware those who appear to support change, but do little to achieve it.

"We may soon see government ceding the running of many public services to more informed and capable private companies."

San Francisco workshop

2.1
6.13%
-1.12
+22.36
1.24
2.45
32.6
42.1
27.3
55.6
14.5
78.8
35.5
22.3

4.12 Accountability and Regulation

Rising concern about the use of data influences public opinion. Policy makers seek a more joined-up approach to regulation, governance, and accountability

Context

Ten years ago, such was the confidence and faith in the new technology companies, that many believed that the best approach was to allow the industry to self-regulate. It was certainly the cheaper and more time-efficient option. The view was that by creating an effective and credible self-regulation framework, companies would be able to react faster to the rapid pace of innovation. This was supported by an implicit trust that technology companies were acting for the good of society.

The message from our workshops was stark: today, that confidence has evaporated. Very few people now believe that a ‘data free-for-all’ will automatically produce the best of all possible worlds. Given the sweep of technology issues which are now shaping our economies, democracies, and personal lives, there is a need for governments to take a more active and assertive approach to regulation. The discussion has moved from whether tech companies should be regulated, to how.

This debate is both intense and complex. Issues and dilemmas discussed in the workshops included:

- How to marry effective regulation with the speed of technology change. In general, the policy regulatory cycle takes anything from 5 – 20 years, while a new digital service can sweep the world in just a few years; how can policy makers and regulators keep up?

- Given the pace of change, the lack of transparency of some organisations, and the consequent difficulties policy makers and regulators have to keep abreast of the new technologies and their implications, what is the best process to develop new rules and regulations? If rulemaking is to be a collaborative industry/governmental effort, how should this collaboration be organised?

- What is the best level to regulate? The digital revolution is a global phenomenon. Some in our workshops argued for an international body to create common rules and frameworks that can be applied globally. But is that practical? If not, is a regional approach better - or does that encourage the system to splinter? And do national regulators really have the clout to deal effectively with multinational corporations whose resources sometimes dwarf those of national states?

- What is the appropriate focus and scope of any new rules and regulations? For the past few decades, the regulatory priority has been to address real/potential consumer harms. But should this be broadened to include the health of data ecosystems and economies as a whole? If so, how?

- What are the best levers and frameworks by which to develop rules and regulations? Should they revolve around issues such as competition, or should they perhaps focus on more technical issues of financial reporting, accounting, and taxation?

- Who should we trust to develop policy? Can we trust national policy makers and regulators, as they may have a vested interest to install data capture and surveillance operations that potentially harm citizens as much as benefit them? Alternatively, can we trust technology companies which are founded to generate profit, not necessarily to protect the interests of citizens?

- How can we avoid the pitfalls of badly drafted regulation which has counter-productive effects, or stifles innovation?

"There is a need to co-design a regulatory framework for the digital age."

Frankfurt workshop

Keeping up with Change

Throughout our workshops, there was a strong sense that, over the last 20 years or so, the capacity of governments to deliver for their constituents is shrinking, at the same time as technology companies have emerged as a political force in their own right. Some, particularly those in the US, have been encouraged by a long period of laissez-faire government to innovate and disrupt at will. In so doing, they have created significant social benefits. But the perceived disregard by a select few, highly profitable technology firms for accepted standards of behaviour around issues such as privacy, security, and indeed tax, has caused widespread alarm. It is hardly surprising, therefore, that there is a regulatory and political backlash.

It is clear that technology companies and regulators must work more closely together and become more aligned to work out new ways to protect citizens' data. There are many good and thoughtful people in both camps who, if they use their combined expertise, are capable of building regulatory measures that protect users, without stifling innovation - perhaps considering incremental regulation, rather than waiting for an issue to mature. Whatever approach is ultimately decided, there was almost universal consensus during our workshops that this requires a change of mindset on both sides, and that the first step in this journey is the creation of a shared language about data, the establishment of common principles around data use, and common approaches to their implementation.

Common Purpose

Although we live in a time when the geopolitical landscape seems to be fracturing, governments need to cooperate more effectively with each other, given the way technology is oblivious to national borders. As with regulation around arms controls, the creation of international rules would help nations react and respond collectively, should they be violated. Work is already in progress in this regard; for example, the EU–U.S. Privacy Shield, acts as a framework for regulating transatlantic exchanges of personal data for commercial purposes, and, in 2018, President Emmanuel Macron launched the Paris Call for Trust and Security in Cyberspace. This high-level declaration in favour of the development of common principles for securing cyberspace has already received the widespread backing from state, international, and civil society organisations and the private sector. Although this doesn't require governments or corporations to legally adhere to any specific principles, it does act as a symbol of the need for diplomacy and cooperation in cyberspace, where it's hard to enforce any single country's laws.[164] At the same time, some countries have chosen to act unilaterally around issues such as online harm, content moderation, and malicious attacks.

> "As we move forward, we are likely to see more pockets of regulation that attempt to emulate or build on regulation elsewhere – such as GDPR."
>
> Johannesburg workshop

Many in our workshops voiced the view that, in order to develop robust international frameworks, it is vital to bring together those countries who are willing to address these new and fast-moving challenges, and to build momentum by leaving the door open for others to join the initiative. The US, for example, has yet to support the Paris Call, but building a coalition of those who are, is a good way to encourage its involvement and support.[165] In addition, the creation of international rules would certainly make it easier for countries to respond effectively, should those rules be violated.

We need to recognise that some issues may not lead to global consensus. Views around privacy, freedom of expression, and human rights, are viewed very differently in different countries. For example, China, which has the largest Internet user base in the world and ambitions to be the leading cyber superpower, and although it has moved to protect young people from online harm, it has little interest in empowering its citizens - rather it has followed its own distinct policy; shutting down websites and censoring content.

Despite this, the issues surrounding the development of new technologies require initiatives that are both multilateral and multi stakeholder in scope. In democracies, government policy makers, who have been given the authority to apply the laws under which we live, are uniquely placed to lead here.

Building or Constraining Monopolies

As John Naughton, for one, has summarised it, “one of the paradoxical things about digital technology is that, while in theory it fosters competition, in practice it leads to winner-takes-all outcomes. The reasons for this are complex – they include zero marginal costs, powerful network effects, power-law distributions, and technological lock-in.” The five biggest companies in the world are now all digital giants, each wielding huge power in their markets.[166] Just as with previous interventions into the oil, steel, and telecom industries, regulators are seeking to curb their influence. The EU in Brussels is often seen as leading this drive, but it is not alone. Indeed in 2018, many highlighted the potential role of the OECD to have a broad impact across the board. The October 2019 OECD proposal to shake up global taxation on the digital leaders is one of the first visible examples of this building momentum.[167]

In the EU, efforts to rein in firms that abuse their monopoly power, have resulted in, for example, a record $5 billion fine against Google - which is more than the tax that they currently pay.[168] In addition, GDPR is having a profound effect on the advertising and data gathering ecosystem.

“There needs to be a more clearly articulated government data strategy to enable community-driven initiatives that have wide public benefit.”

Singapore workshop

Elsewhere, California has already passed a sweeping data-privacy law, set to go into effect in 2020; the Indian government, as a reaction to what some saw as an attempt at colonialism, banned Facebook from allowing users to browse, without paying for mobile data[169]; even in China, the government is becoming more involved in controlling the dominance of Alibaba, Tencent, Baidu, and JD.com, rejecting, for example, a credit-scoring system by Alibaba's affiliated payment company, Ant Financial, in favour of one of its own. Some say that this is a cynical effort to benefit domestic actors - think of China's enormous tech industry, or India's burgeoning e-commerce giant, Flipkart. Others see it as evidence of the tide turning against the previous regulatory freedoms.

The problem of tax

Understanding how best to tax the different parts of the data value chain may become critical to enable a more equitable distribution of the profits that data-driven businesses can generate, while maximising the growth of the data-driven economy and ensuring good practice. The EU's proposed digital services tax, which seeks to tax revenues generated within national or bloc jurisdictions, and bypass the knotty issue of how to tax profits that are registered overseas, is one potential answer, but it also raises questions around fairness and application. In some of our workshops, it was pointed out that we are likely to see a period of piecemeal, attempts by different governments to tax digital and data-driven businesses, before we see more coherent strategies around digital/data wealth redistribution.

Surveillance and State Interference

The other major concern is the increasing control of data by government, and especially the focus on surveillance as a primary purpose. While the Russian, Chinese, and US instances are the most commonly shared globally, there were multiple additional examples. Control of data was brought up in Hong Kong and London, where the negative impact of government surveillance on democracy, particularly given the growing prevalence of facial recognition technology, is becoming a matter of public concern.

> “It is more likely that self-regulation will drive community standards. These in turn will drive localised regulation.”
>
> Manila workshop

What We Heard

There was also broad agreement that, given the extraordinarily rapid pace of technological change, it is unrealistic to expect governments to devise, update, and enforce effective data regulation without the cooperation of technology companies, particularly given the transnational nature of data. Some sort of collaboration between policy makers and technology companies is necessary. Although a number of business-driven consortia have cropped up to serve as independent standards-creation bodies, for example, not all have been effective, and the disconnect between regulation and industry remains.

The solution that was identified during our workshops was a global body to act as the focal point for governance activities. In Jakarta, the view was that there should be ***"an independent global data regulation framework (maybe like the G20)."*** In Bangkok, it was for *"a global data authority (like the WTO)."* In Singapore, there was the need for *"a global organisation (like the WEF, IMF, or WHO).* In Mexico, the proposal for 2030 was for *"an international body able to act at global level (like the UN),"* while in a London discussion, technology companies backed the role of the OECD in potentially coming up with an answer. All are looking for a higher authority to set the standards, define the common ground, and ensure balance and independence. All recognised that this may be a long way off.

Those in Jakarta, felt that regulatory change should be government-led, primarily because governments rather than corporates have a democratic mandate to represent the people. Others, such as those in Frankfurt and Bangalore, considered that co-regulation is more effective when the public and private sectors ideally *"co-design a regulatory framework for the digital age."*[170] In Hong Kong, a proposal was that this should be *"a framework of common principles allowing public and private use of data across multiple jurisdictions. To achieve this, first there has to be collaboration around a set of principles on standards."*

Rather than a global framework for data governance and a dedicated organisation to oversee this, many felt it would be more likely that a number of regulatory regions, within which common policies are adopted, will emerge. Europe, China, and the US are evidently three, and an ASEAN-focused approach building on the APEC data privacy framework is promised. In Africa and Latin America, some are considering their own regional regulatory methods. Europe's GDPR, which has harmonised data protection rules and given individuals greater rights over how their data is used, was often mentioned as a template for other nations to follow. *"GDPR will change the data landscape in Nigeria and bring in new standards."*[171] In Johannesburg, it was considered that *"as we move forward, we are likely to see more pockets of regulation that attempt to emulate or build on regulation elsewhere – such as GDPR."* That said, not everyone felt that regulation is necessary. In Manila, it was felt that it was *"more likely that self-regulation will drive community standards that in turn will drive localised regulation."*

> "There is a need to co-design a regulatory framework for the digital age."
>
> Frankfurt workshop

Either way, calls for a more joined up approach to regulation were common. So far, it was argued, the response to rapid technological change has been too piecemeal to be truly effective. From our first workshop in Bangalore; *"government policy is currently very scattered, with little uniformity of purpose,"* to our final meeting in Santiago; *"the challenge will be how different jurisdictions take control of the issues around data,"* there was recognition that the current plethora of different regulation does not solve the big issues.

In terms of regulatory levers, one suggestion was that if regulators can help put a value on data, or at least define the parameters by which data can be valued, then there could be a significant change in views around how it is managed. Putting a value on data, it was argued, would drive more informed debate on how that value should be better shared. As well as improving financial reporting, it could aid the formulation of tax policies, while also influencing organisations' own data strategies.

Another suggestion, which was also recently raised in the FT, is to shift to an earlier interpretation of antitrust regulation that focuses, not just on consumers, but rather on whether the larger economic ecosystem is being harmed.[172] Linked to this was the notion that better governance for data could unlock numerous positive opportunities for society. In India, for example, they looked forward to *"government guiding the private sector more on the development of 'social value of data' policies."* In Singapore, the call was for *"a more clearly articulated government data strategy to enable community-driven initiatives that have wide public benefit."* Participants in Nairobi wanted *"data to better drive development, become more accessible, and reduce poverty."*

"As we move forward, we are likely to see more pockets of regulation that attempt to emulate or build on regulation elsewhere – such as GDPR."

Johannesburg workshop

At the same time, there was widespread suspicion about governments and state actors, and the possibility that they could use new regulatory powers to assert their own control, especially over personal data for the purposes of surveillance. In Johannesburg, it was suggested that *"there is a risk that certain governments could increasingly use data regulation to drive top-down state control of very powerful data sets,"* while in Pretoria students debated how *"the centralisation of data creates a greater opportunity for government control."* Their fear was that, across Africa, *"some governments can limit access to data under the guise of national security."* In January 2019, the Zimbabwe government cut internet access for 3 days, to curb opposition protests. Further north in Abuja, the forecast was that *"government will want to control the data, while people do not realise the value."* In South America, anxiety about growing state surveillance collating more information about citizens, was also expressed in both Bogota and Santiago. The view in Hong Kong was that there are mounting instances of *"data creating power, shaping the wielding of power, the balance of power, and the accountability of power."*[173] Many there were concerned about the impact that this is having on society.

And there is the ever-present danger that regulation can create problems, as well as solve them.

In Hong Kong, concern was expressed that *"over-regulation could diminish the value of data and hinder innovation for social utility."*[174] They also observed the cultural effects of regulation. *"It is also important to consider the implication of the different ideologies within national boundaries, and their potential ambition,"* and *"what would be the implications of China winning the debate around data, and what would happen if it exported its values around the world?"* China's Great Firewall has already effectively caused two internets to develop. Looking ahead, if a US-China trade war deepens, and China's leaders feel they need to turn tech companies to their advantage, it is perfectly possible to see that those countries which are part of the Chinese Belt and Road Initiative, may well be encouraged to take on the Chinese tech infrastructure.[175]

> Governments will want to control the data, while people do not realise the value."
>
> Abuja workshop

Implications for Data Value

How then can regulators regulate effectively, given the challenges of technological change and the scale of the data revolution? Most in our workshops agreed that, to date, too little has been done to protect the public interest, and that governments need to step up to address this. There was also recognition that, despite the pressing nature of the challenges, global alignment may well be too hard to achieve in the short term, not just because of the scale of the ambition and agreements required, but also because of the distrust between some governments, existing international institutions, and large corporations. Indeed, in some countries, the disenchantment with globalisation and lack of enthusiasm for Western culture, alongside a growing recognition of China's increasing influence, there were strong indications that different regional policy models may well emerge, which could be to the detriment of the global data economy.

Many issues will require compromise – this will be hard for business leaders in particular, given they are not used to regulatory constraint, but it is something that they are beginning to acknowledge and accommodate. In the short term, this may also affect how data can be valued and may even limit the growth trajectories of some organisations. However, in the long term, many in our workshops agreed that multi stakeholder collaboration is a pragmatic stepping stone in the shift from reactive to proactive policymaking, which will ultimately better protect human rights and freedoms, and at the same time, ensure the long-term development potential of data-driven innovation which benefits us all.

"Overregulation could diminish the value of data and hinder innovation for social utility."

Hong Kong workshop

5.0 Conclusion

This report identifies and explores the diversity and extent of the data revolution. Today's technological changes are so broad and deep, that they have fundamental implications for society. They offer a wealth of opportunity but also require careful consideration around how to deal with the pressing problems that have emerged. Our research took place during a critical time. The way in which large organisations share, manipulate, and profit from individuals' personal data was seldom out of the headlines, and the frustration and disappointment that this caused was often reflected in our discussions. The message from our workshops was clear. The organisations whose businesses are transforming the world, have a responsibility to help address the consequences of these changes. In a sector long focussed on growth and disruption, this is significant.

Irrespective of location, we heard acknowledgement that the tech sector cannot address these challenges on its own; there needs to be a collective effort and institutional change. Alongside self-regulation, government action is required to preserve the principles of social and economic consensus. Governments, separately, with each other and also in collaboration with business, must find ways to move faster. The EU is leading in this area, but other regions are keen to learn from their example and adapt legislation to their own local and regional requirements. In order to help regain public trust, technology companies need to be more transparent, more proactive, and more collaborative. Every discussion we heard acknowledged that the approach of driving change, watching what happens and then responding to it, has had grave consequences, and this can no longer be allowed to continue.

Ultimately, we need to reconcile a time of extraordinary technological transformation, with the preservation of human values – the things that, in the end, matter most to people. This means that individuals too need to step up. Every day, we are consenting to things we don't properly understand, failing to appreciate how important our privacy is in protecting our rights and freedoms, and making ourselves vulnerable to the influence and manipulations of bad actors. To preserve our values and the hard-fought rights that support them, we, as individuals, need to become more aware of the implications of this – we need to be more considered about how we manage our own data use, and indeed allow others to have access to it. Having a basic understanding of how to control and manage personal data is a social responsibility.

As we travelled, debated, and listened to different voices in very different cultures, and in countries at diverse stages of technological development, it was evident that there is widespread understanding of the challenges. Moreover, there is a global appetite to establish the processes and institutions that can tackle the particular issues, such as:

- How to develop a common language around data that will allow us to accurately describe and find solutions for some of the major challenges. These include:
    - the way organisations use personal data for profit;
    - the mechanisms that give individuals agency and control around their data;
    - a shared understanding of the implications of these, and what the potential solutions to these might be.
- Agreeing the legitimate use of power that is generated by a data-driven economy, who should wield this power, and how they should be governed and held accountable.
- Agreeing who has what rights to the different dimensions of value that data can generate. This is made doubly complex because data undermines absolutist notions of ‘private property’ (because it can be used without being used up, and because the same data can be used by many different parties for many different purposes).
- Agreeing which organisations should take responsibility for the governance and regulation of data and data-driven processes and activities, and defining how they should operate: national, regional, or global.
- Ensuring trusted interactions and relationships across the complete data ecosystem.
- Defining how organisations should account for their data, whether it can be recognised as a corporate asset and therefore become liable to taxation.

Throughout the programme, the importance of trust and the need for organisations, governments, and individuals to behave in a trustworthy manner, was a constant thread. The value of data can only be fully realised if we can be confident that it is shared responsibly, and is of good quality. Indeed, much of today’s business already relies on the ability to move accurate data, including personal data, across borders without restriction. However, few data organisations are considered to be trustworthy. Because of this, there are calls for them to act in a more transparent and accountable fashion. Despite the possibility that limiting global data flows may have the opposite effect to that which was intended, and reduce the value of data, this lack of trust is one of the core factors which has led to an increase in data nationalism, and the rise in data sovereignty and localisation.

Looking ahead, many in our workshops made it clear that organisations should also be held to account for the way they use and profit from data. They should be able to demonstrate, not only that they are capable of managing data, but also that they will do so ethically. Greater public understanding of what this should entail, and increased awareness of an individual’s rights and responsibilities for their own personal data, will be vital here, so again, a focus on digital literacy is key.

As we wrestle with these challenges, varied cultures and circumstances mean that people bring very different perspectives to the same issues. Given the diversity of opinion, the multiple levels of market maturity, and the manifold ways that data can be used, it is almost inevitable that there will be disagreement about how best to progress.

Our own conclusions from this are as follows:

- We need processes that bring all the key stakeholders together into a constructive, meaningful debate. Given the diversity and complexity of the issues involved, it is likely that this will need multiple forums operating at many different levels, but the shared ambition is to address and find solutions for the key issues around data.

- We need institutional reform, development, and innovation, in order to achieve an end state which is considered fair, workable, and beneficial to all stakeholders. Regulation alone will not be sufficient. Each of the issues we have discussed - data and digital literacy, privacy, consent, open data, machine data, issues relating to financial reporting and taxation, data localisation and sovereignty, data ethics - require their own tailored solutions. However, while these issues generate their own distinct requirements, they are not neatly self-contained. They are multi-faceted, multi levelled, highly contextual, entangled, and overlapping. Regulation alone cannot achieve consensus around these.

- When workable ways forward are found, we will almost certainly need to adapt the operation of existing institutions, such as how regulation works and/or invent completely new ones. One of the challenges of building such institutions is that appropriate functions, powers, boundaries, checks, and balances, all need to be negotiated in a world where jurisdictions will overlap. Across the world, we heard multiple independent calls for the establishment of a World Data Council capable of getting the many different (and often hostile) national governments and multinational corporations to address global data governance challenges. Many organisations, from the IMF and the EU, to individual governments and corporates, are already trying to build a structure around this, in order to control and manage the flow of data in more meaningful ways. These are welcome steps. Different, additional institutions and processes may also be needed.

- This is a complex and competitive environment, and the very fact that these challenges have been recognised and are being addressed is a positive step. We recognise that the ideal of a global framework, able to quantify the value of one data set against another is a long-term goal. But with good will, it would be possible to achieve. However, as this process is iterative, and as many of our workshops pointed out, it may be that different regional or industry solutions will emerge to set standards, following on from the example set by Europe's GDPR.

Finally, we recognise that we can only reflect what we heard at any given time, but given the context in which our research took place, we believe the views we have reflected are important. No one wants to create a world which is worse for the next generation than the one we enjoy today. Without greater consideration of the consequences of our actions, however, it is entirely possible that we could. Data is making the world a smaller and more intense place to live in. In order for us to operate in this sort of environment, there must be clearly defined and widely recognised rules. We all need to hold ourselves more accountable for consequences of the decisions we are making.

5.1 Questions

From the work we have undertaken around the world and the follow-on synthesis in this document, we can see that this project has perhaps raised as many new questions as it has provided conclusive answers. By the very nature of the interlinked topics of the value and role of data that are both currently undergoing substantial change across multiple sectors, this should not be a great surprise. While the previous chapter provided some conclusions from the research, we can also see further issues to be discussed.

To help provoke additional dialogue, especially related to the specific implications for various organisations, below we have suggested some questions that may be useful to initiate conversations. We have proposed ten questions each for individuals, for companies, and for governments.

Questions for Individuals:

1. How can education help us to navigate the internet and digital platforms, and engage with social media? Who is best placed to teach us?

2. How can we ensure that we have the skills needed to work in a digital age? Do we need to train or retain so we can actively participate in the digital economy?

3. How would you assess if your data is being valued fairly, when it is used in exchange for something else?

4. Is 'ownership' a useful/practical concept when it comes to certain types of data, such as personal data? If not, what alternative concepts can we use to replace it?

5. How can we become more aware of our individual rights and responsibilities online? Should citizens be more proactive in making decisions around how to gain value from their data?

6. Who can we trust most to manage our data? Why?

7. What do you think the most significant digital rights should be and should they vary dependent on culture and region?

8. Given that we live in an era of increased surveillance, does privacy matter? Is it possible to achieve?

9. Would you be prepared to pay for services in exchange for greater privacy?

10. Would you be happy for data about you to be shared for social causes?

Questions for Companies:

1. Organisations collecting and using large quantities of data can generate significant value for individuals, society, the economy, and for themselves. At the same time, however, they may create excessive concentrations of power and/or use the power they do have unfairly or inappropriately. How should these dangers best be addressed? By who?

2. Aside from ownership, what ways can we use to allocate rights, benefits, and responsibilities relating to data across stakeholders including governments, technology companies, multinational corporations, and individuals?

3. Is it possible to create a 'common language' where, across the world, key stakeholders all use the same terms and definitions to describe what is happening with data?

4. Is there sufficient understanding amongst policy makers to manage the transition to and the impact of digital technologies successfully? Can regulators better support digital literacy?

5. If it is impossible to deliver "informed consent" in any practical form, what should replace it?

6. How should these decisions be implemented and enforced?

7. If the momentum towards data sovereignty continues, will it be possible to ensure an international market for data?

8. What would encourage you to make your data sets available for public good? What constitutes 'good quality' open data?

9. Given the race to collect, store, and use data, and the commercial opportunities that this creates, how can businesses ensure that ethics are not sidelined? How can this incorporate the enormous variety in moral and ethical beliefs between different cultures?"

10. What does it take to be trustworthy?

Questions for Governments:

1. If a corporate entity is deemed to have too much power or to be exercising its power irresponsibly, what are the appropriate mechanisms for effective action?

2. How can government enable citizens to have a more active role in decisions around how to gain value from their data, either for themselves or for others?

3. When is it necessary/desirable for data to flow across national borders? What different rules should be applied to different types of data (e.g. personal, non-personal), and different circumstances and use cases?

4. How can/should disputes between different entities and jurisdictions (local, regional, global) relating to the collection and use of data be handled?

5. Which bodies, at what level (local, regional, global), are best placed to take a lead on the on this, and how can we ensure a) their legitimacy in the eyes of key stakeholders and b) their effectiveness?

6. What is the best way to address key stakeholders' concerns (e.g. the dangers of a new 'data imperialism', the risks that constrained data flows could undermine innovation and economic prosperity)?

7. Will IoT data have greater value if it is proprietary or open to all? How do you ensure clarity about what data should be opened up, for what uses, and by who?

8. How can we create a regulatory environment which encourages competition, while making information-intensive organisations more accountable for the data in their care?

9. New commercial sources of value are being created from public, academic, and government information, which are being used for private enterprise. Is it possible to limit the 'privatisation' of open data?

10. Do new innovations around AI and Machine Learning need a different form of governance and regulatory approach?

Workshop Locations

Australia

Sydney

Canada

Toronto

China

Shanghai

Colombia

Bogota

Denmark

Copenhagen

Germany

Frankfurt

Hong Kong

Hong Kong

India

Bengaluru

Indonesia

Jakarta

Ivory Coast

Abidjan (x2)

Japan

Tokyo

Kenya

Nairobi

Mexico

Nairobi

Nigeria

Abuja

Lagos

Philippines

Manila

Senegal

Dakar

Singapore

Singapore

South Africa

Pretoria

Johannesburg

Spain

Madrid

Sweden

Stockholm

Thailand

Bangkok

United Kingdom

London (x2)

United Arab Emirates

Dubai

USA

San Francisco

Washington DC

References

[1] https://www.chathamhouse.org/chatham-house-rule

[2] https://www.slideshare.net/futureagenda2/future-of-value-of-data-interim-summary-aug-2018compressed

[3] https://integratedreporting.org/

[4] https://www.weforum.org/agenda/2017/09/the-value-of-data/

[5] London Workshop

[6] Dakar

[7] Jakarta

[8] http://reports.weforum.org/rethinking-personal-data/executive-summary/?doing_wp_cron=1547984248.0330200195312500000000

[9] https://www.michalsons.com/wp-content/uploads/2017/11/PoPI-Regulations-Final-14-Dec-2018-3-languages-42110_14-12.pdf

[10] Nairobi workshop

[11] Johannesburg workshop

[12] Nairobi workshop

[13] Pretoria workshop

[14] https://www.bigbangerp.com/blog/data-localization-laws/

[15] https://www.econstor.eu/bitstream/10419/174802/1/ecipe-pb-2016-03-Unleashing-Internal-Data-Flows-in-the-EU.pdf

[16] London Workshop

[17] Madrid workshop

[18] London workshop

[19] Bogota workshop

[20] Lagos workshop

[21] Bogota workshop

[22] London Dinner 2018

[23] https://hbswk.hbs.edu/item/do-experts-or-collective-intelligence-write-with-more-bias-evidence-from-encyclopdia-britannica-and-wikipedia

[24] Bangkok workshop

[25] Madrid workshop

[24] Tokyo Workshop 31

[25] Digital Skills Crisis report from UK Government

[26] https://publications.parliament.uk/pa/cm201617/cmselect/cmsctech/270/270.pdf

[27] e.g. https://www.weforum.org/reports/the-future-of-jobs-report-2018, https://ainowinstitute.org/AI_Now_2017_Report.pdf and

[28] https://www.pwc.com/gx/en/services/people-organisation/workforce-of-the-future/workforce-of-the-future-the-competing-forces-shaping-2030-pwc.pdf 29

[29] http://www3.weforum.org/docs/WEF_Future_of_Jobs_2018.pdf 30

[30] Tokyo Workshop 31

[31] Washington DC Workshop

[32] Nairobi workshop

[33] Tokyo Workshop

[34] Madrid Workshop

[35] https://www.theguardian.com/global-development/2019/feb/27/millions-of-ugandans-quit-internet-after-introduction-of-social-media-tax-free-speech

[36] https://www.cbronline.com/opinion/data-privacy-day-2018

[37] https://www.wsj.com/articles[40]/why-i-put-my-dogs-photo-on-social-media-but-not-my-sons-11552404655?mod=searchresults&page=1&pos=12

[38] http://reports.weforum.org/rethinking-personal-data/executive-summary/?doing_wp_cron=1547984248.0330200195312500000000

[39] https://www.pewresearch.org/fact-tank/2016/09/21/the-state-of-privacy-in-america/

[40] https://dipp.gov.in/sites/default/files/DraftNational_e-commerce_Policy_23February2019.pdf

[41] http://cyberlaw.stanford.edu/publications/thanks-amazon-government-will-soon-be-able-track-your-face

[42] https://searchsecurity.techtarget.com/feature/Ciscos-chief-privacy-officer-on-the-future-of-data-after-GDPR?

[43] https://www.ft.com/content/ff2d9600-259a-11e9-8ce6-5db4543da632?

[44] Marreiros, H., Tonin, M., Vlassopoulos, M., & Schraefel, M. C. (2017). “Now that you mention it”: A survey experiment on information, inattention and online privacy. Journal of Economic Behavior & Organization

[45] http://www.meaningfulconsent.org/reports/2017/consentful_economy/

[46] This is an area which a number of regulatory authorities are already examining (e.g. the UK’s CMA and the EC).

[47] Hong Kong Workshop

[48] Bangkok workshop

[49] https://webfoundation.org/2017/11/the-future-of-open-data-and-the-openness-agenda/

[50] https://data.worldbank.org

[51] https://data.oecd.org

[52] https://data.europa.eu/euodp/en/home

[53] http://data.un.org

[54] https://opendatabarometer.org/

[55] http://odin.opendatawatch.com

[56] https://index.okfn.org/place/

[57] https://www.microsoft.com/en-us/open-data-initiative

[58] Argast and Zvyagintseva 2016; Weerakkody et al. 2017; Robinson and Ward-Mather 2017.

[59] Janssen, Marijn, Yannis Charalabidis, and Anneke Zuiderwijk. “Benefits, adoption barriers and myths of open data and open government.” Information systems management 29.4 (2012): 258-268

[60] https://webfoundation.org/2017/11/the-future-of-open-data-and-the-openness-agenda/

[61] https://www.ft.com/content/f8e9c2ea-b29b-11e8-87e0-d84e0d934341

[62] https://blog.okfn.org/2017/05/31/open-data-quality-the-next-shift-in-open-data/

[63] https://okfn.org/

[64] https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information

[65] https://www.govloop.com/community/blog/future-open-data-relationship-private-sector/

[66] https://www.waze.com/en-GB/about

[67] https://www.theverge.com/transportation/2015/5/19/8622831/uber-self-driving-cars-carnegie-mellon-poached

[68] https://academic.oup.com/jlb/advance-article/doi/10.1093/jlb/lsy021/5114251

[69] https://www.ft.com/content/94e86cd0-44b6-11e9-a965-23d669740bfb

[70] https://www.theguardian.com/science/2005/oct/14/genetics.research

[71] https://eu.usatoday.com/story/news/2018/02/12/trump-budget-continues-support-air-traffic-control-privatization/324331002/

[72] https://dg.dk/wp-content/uploads/2017/11/Open-Access-to-Data-–-It’s-not-that-Simple.compressed-1.pdf

[73] https://www.nature.com/articles/s41467-019-10933-3

[74] https://www.cigionline.org/publications/open-data-endgame-countering-digital-consensus

[75] https://www.infineon.com/cms/en/discoveries/internet-of-things-2030/

[76] https://www.intel.com/content/www/us/en/internet-of-things/infographics/guide-to-iot.html

[77] https://www.cisco.com/c/dam/en/us/products/collateral/se/internet-of-things/at-a-glance-c45-731471.pdf

[78] https://www.researchgate.net/publication/300366604_5G_Vision_and_Requirements_for_Mobile_Communication_System_towards_Year_2020

[79] https://www.seagate.com/gb/en/our-story/data-age-2025/

[80] https://www.bain.com/insights/choosing-the-right-platform-for-the-industrial-iot

[81] https://www.frontier-economics.com/media/1167/201803_the-economic-impact-of-iot_frontier.pdf

[82] https://www.pwc.fr/fr/assets/files/pdf/2017/03/2017_ai_and_iot_v13b.pdf

[83] Frankfurt workshop

[84] https://www.futurefarming.com/Tools-data/Articles/2018/10/Data-ownership-questions--and-why-theyre-important-340743E/

[85] https://www.flagshipfarmers.com/en/about-the-program/

[86] https://newsroom.toyota.co.jp/en/detail/18135029/

[87] https://www.futureagenda.org/news/the-future-of-automotive-data-and-its-potential-value

[88] https://www.futureagenda.org/news/future-of-autonomous-vehicles-building-the-informed-view

[89] https://www.futureagenda.org/news/the-future-of-automotive-data-and-its-potential-value

[90] https://www.sciencedirect.com/science/article/pii/S0963868717302615

[91] https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf

[92] https://www.thetimes.co.uk/article/nhs-data-is-a-such-precious-asset-it-must-be-given-a-proper-valuation-7967nbmvd

[93] https://www.imf.org/~/media/Files/Conferences/2018/6th-stats-forum/presentations/session-3-wendy-li-value-of-data-presentation.ashx?la=en

[94] https://www.ft.com/content/93ffec82-ed2a-11e8-8180-9cf212677a57

[95] https://www.reuters.com/article/us-cybersecurity-hospitals/your-medical-record-is-worth-more-to-hackers-than-your-credit-card-idUSKCN0HJ21I20140924

[96] https://www.experian.com/blogs/ask-experian/heres-how-much-your-personal-information-is-selling-for-on-the-dark-web/

[97] https://ig.ft.com/how-much-is-your-personal-data-worth/

[98] http://reconanalytics.com/2017/05/digital-advertising-takes-the-lead-dominated-by-google-and-facebook/

[99] https://www.ft.com/content/e38b60ce-27d7-11e8-b27e-cc62a39d57a0

[100] https://www.ft.com/content/a8c36a90-dba5-11e8-9f04-38d397e6661c

[101] https://www.ft.com/content/f00d2f70-8a6f-11e9-a1c1-51bf8f989972

[102] https://www.theguardian.com/commentisfree/2018/apr/27/chris-hughes-facebook-google-data-tax-regulation

[103] Manila Workshop

[104] https://ec.europa.eu/digital-single-market/en/news/study-emerging-issues-data-ownership-interoperability-re-usability-and-access-data-and

[105] http://integratedreporting.org

[106] https://www.bbc.co.uk/news/business-46050724

[107] https://data.worldbank.org/indicator/NY.GDP.PCAP.CD?locations=UG

[108] http://ecipe.org/publications/restrictions-to-cross-border-data-flows-a-taxonomy/

[109] https://www.ft.com/content/6f0f41e4-47de-11e8-8ee8-cae73aab7ccb

[110] https://www.bigbangerp.com/blog/data-localization-laws/

[111] https://www.forbes.com/sites/ronakdesai/2019/04/30/indias-data-localization-remains-a-key-challenge-for-foreign-companies/#45fe7713e0a3

[112] https://itif.org/publications/2017/05/01/cross-border-data-flows-where-are-barriers-and-what-do-they-cost

[113] Nairobi workshop

[114] https://economictimes.indiatimes.com/tech/ites/all-about-indias-data-localisation-policy/articleshow/66297596.cms

[115] https://www.business-standard.com/article/companies/google-facebook-protesting-data-localisation-to-evade-taxes-phonepe-118091300023_1.html

[116] https://economictimes.indiatimes.com/news/economy/policy/localisation-heat-data-bank-here-may-help-government-debit-taxes/articleshow/67120911.cms

[117] https://www.ft.com/content/92bb34a8-b4e5-11e8-bbc3-ccd7de085ffe

[118] http://knowledge.freshfields.com/m/Global/r/3824/where_are_we_now_with_data_protection_law_in_china_

[119] https://jsis.washington.edu/news/chinese-data-localization-law-comprehensive-ambiguous/

[120] https://www.theguardian.com/technology/2016/feb/22/tim-cook-apple-refusal-unlock-iphone-fbi-civil-liberties

[121] https://www.theguardian.com/commentisfree/2018/mar/04/apple-users-icloud-services-personal-data-china-cybersecurity-law-privacy

[122] https://www.endgame.com/blog/technical-blog/march-toward-data-localization

[123] https://www.gsma.com/newsroom/press-release/gsma-free-flow-of-data-across-borders-essential-for-asias-digital-economies/

[124] https://theaseanpost.com/article/data-localisation-southeast-asia

[125] https://www2.deloitte.com/insights/us/en/focus/tech-trends/2018/data-sovereignty-management.html

[126] https://restoreprivacy.com/5-eyes-9-eyes-14-eyes/

[127] https://itif.org/publications/2017/05/01/cross-border-data-flows-where-are-barriers-and-what-do-they-cost

[128] Nairobi workshop

[129] https://www.nytimes.com/2018/05/04/books/review/automating-inequality-virginia-eubanks.html

[130] https://www.forbes.com/sites/davidshaywitz/2018/02/18/the-deeply-human-core-of-roches-2-1b-tech-acquisitionand-why-they-did-it/#e2d838e29c21

[131] https://hbr.org/2018/08/what-data-scientists-really-do-according-to-35-data-scientists

[132] https://qz.com/1427621/companies-are-on-the-hook-if-their-hiring-algorithms-are-biased/

[133] https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

[134] 2017 survey by talent software firm CareerBuilder

[135] https://www.technologyreview.com/s/612775/algorithms-criminal-justice-ai/

[136] https://www.futureagenda.org/news/future-of-digital-identity-report-launch

[137] Madrid workshop

[138] Washington DC workshop

[139] https://www.accenture.com/t00010101T000000Z__w__/gb-en/_acnmedia/PDF-22/Accenture-Data-Ethics-POV-WEB.pdf#zoom=50

[140] https://www.ft.com/content/24aaaa9a-674a-11e8-8cf3-0c230fa67aec

[141] https://www.blog.google/technology/ai/ai-principles/

[141] https://www.salesforce.com/company/ethical-and-humane-use/

[143] https://www.bbc.co.uk/news/technology-47825833

[144] https://www.techatbloomberg.com/blog/time-data-ethics-conversations-dinner-table/?utm_source=dsorg

[145] https://www.theguardian.com/commentisfree/2018/mar/04/apple-users-icloud-services-personal-data-china-cybersecurity-law-privacy

[146] https://www.bloomberg.com/news/articles/2017-09-28/microsoft-ceo-urges-tech-to-focus-on-self-policing-not-regulation-fears

[147] https://www.partnershiponai.org

[148] https://www.thebritishacademy.ac.uk/sites/default/files/JBA-6s1-Hamdani-Hashai-Kandel-Yafeh.pdf

[149] https://www.thebritishacademy.ac.uk/future-corporation-research-programme

[150] https://www.thebritishacademy.ac.uk/sites/default/files/JBA-6s1-Hamdani-Hashai-Kandel-Yafeh.pdf

[151] https://marianamazzucato.com/publications/books/value-of-everything/

[152] https://www.shoshanazuboff.com/new/

[153] https://www.thebritishacademy.ac.uk/future-corporation-research-programme

[154] https://www.theguardian.com/uk-news/2018/feb/04/cambridge-most-unequal-city-population-divide-income-disparity

[155] https://www.cityoflondon.gov.uk/business/economic-research-and-information/research-publications/Documents/research-2017/total-tax-report-2017.pdf

[156] https://csimarket.com/stocks/singleProfitabilityRatiosy.php?code=AAPL&itx

[157] https://www.ft.com/content/ec84ac4e-4653-11e9-a965-23d669740bfb

[158] https://medium.com/@teamwarren/heres-how-we-can-break-up-big-tech-9ad9e0da324c

[159] https://www.gov.uk/government/publications/unlocking-digital-competition-report-of-the-digital-competition-expert-panel

[160] http://umich.edu/~thecore/doc/Friedman.pdf

[161] https://www.ft.com/content/a84647f8-0d0b-11e9-a3aa-118c761d2745

[162] https://www.ft.com/content/3732eb04-c28a-11e9-a8e9-296ca66511c9

[163] https://www.thebritishacademy.ac.uk/programmes/future-of-the-corporation

[164] https://www.wired.com/story/paris-call-cybersecurity-united-states-microsoft/

[165] B Smith and C Browne, “Tools and Weapons, The promise and the peril of the digital age,” Penguin, 2019

[166] https://www.theguardian.com/commentisfree/2018/sep/16/wanted-in-digital-monopoly-age-powers-to-curb-online-giants

[167] https://www.ft.com/content/b16fd228-ea72-11e9-a240-3b065ef5fc55

[168] https://www.wsj.com/articles/google-to-be-fined-5-billion-by-eu-in-android-case-1531903470?mod=article_inline

[169] https://www.theguardian.com/technology/2016/may/12/facebook-free-basics-india-zuckerberg

[170] Frankfurt Workshop

[171] Nigeria Workshop

[172] https://www.ft.com/content/d72fb54c-ffc1-11e8-b03f-bc62050f3c4e

[173] Hong Kong workshop

[174] Hong Kong workshop

[175] https://www.wsj.com/articles/the-global-tech-backlash-is-just-beginning-1540476151?ns=prod/accounts-wsj

FUTURE
AGENDA
Open Foresight

Delivering Value Through Data

This report shares the findings from the **Delivering Value Through Data** research programme which is based on the output from 30 expert discussions held across 24 countries on every continent. It is part of a series that explores some of the key issues facing society over the next decade.

Delivering Value Through Data highlights several important, emerging issues that are the source of major differences of opinion around the world. These range from the issue of data sovereignty and localization; how best to manage consent and control; concerns around trust and trustworthiness, and, given these, how to address the need for more effective regulation and greater public engagement.

Some of the challenges and opportunities are technical in nature, but many are concerned with different ethical, philosophical and cultural approaches to data and how its value can be delivered.

About Future Agenda

Future Agenda is an open source think tank and advisory firm. We help organisations, large and small, to explore emerging opportunities, identify new growth platforms and develop game-changing innovations. Founded in 2010, Future Agenda has pioneered an open foresight approach that brings together senior leaders across business, academia, NFP and government. The aim is to connect the informed and influential, to challenge assumptions and build a more comprehensive view about the future that will help deliver positive, lasting impact.

For more information and to have access to all our insights please visit **www.futureagenda.org**

Contact:
Dr. Tim Jones:
tim.jones@futureagenda.org
Caroline Dewing:
caroline.dewing@futureagenda.org