EVIDENCE REPORT
No 163

Policy Anticipation, Response and Evaluation

Big Data and International Development: Impacts, Scenarios and Policy Options

Stephen Spratt and Justin Baker

December 2015

The IDS programme on Strengthening Evidence-based Policy works across seven key themes. Each theme works with partner institutions to co-construct policy-relevant knowledge and engage in policy-influencing processes. This material has been developed under the Policy Anticipation, Response and Evaluation theme.

The material has been funded by UK aid from the UK Government, however the views expressed do not necessarily reflect the UK Government's official policies.

AG Level 2 Output ID: 253

BIG DATA AND INTERNATIONAL DEVELOPMENT: IMPACTS, SCENARIOS AND POLICY OPTIONS

Stephen Spratt[a] and Justin Baker[b]
[a] Institute of Development Studies
[b] University of Texas

December 2015

First published by the Institute of Development Studies in December 2015

IDS is a charitable company limited by guarantee and registered in England (No. 877338).

Contents

Acknowledgements

Thank you to everyone who contributed to this research, either as an interviewee, a workshop participant, or both. There are too many to list here, but their names are given in the annexes to this report. Thanks also to Alun Rhydderch for his wise guidance through the Foresight process. Any mistakes remain my own.

Abbreviations

DFID	Department for International Development
EU	European Union
GPS	global positioning system
ICT	information and communications technology
ICT4D	ICT for Development
IDS	Institute of Development Studies
IT	information technology
MGI	McKinsey Global Institute
MIT	Massachusetts Institute of Technology
NGO	non-governmental organisation
OECD	Organisation for Economic Co-operation and Development
R&D	research and development
RCT	randomised control trial
RFID	radio frequency identification
UK	United Kingdom
UN	United Nations
US	United States
WHO	World Health Organization

Introduction

Many people are excited about data, particularly when those data are big. Big data, we are told, will be the fuel that drives the next industrial revolution, radically reshaping economic structures, employment patterns and reaching into every aspect of economic and social life.

The numbers are certainly impressive. In 1946, one of the first computers weighed 30 tonnes and could do 500 calculations per second. Today, IBM's 'Watson' supercomputer can process 500 gigabytes of data per second. Every day, 39 per cent of the global population use the internet. Facebook has more than 1.3 billion active users, and after the United States (US) the countries with the most subscribers are India, Brazil and Indonesia. In 2007, Twitter had 400,000 tweets per quarter. By 2013, there were 500 million per day. Ninety per cent of data in existence were created in the past two years, and the quantity is doubling every two years. The size and cost of storage has fallen by a third every year since the 1970s, making it possible to store these vast new pools of data. New statistical techniques and tools such as machine-learning algorithms can process and analyse these data dynamically, at a scale and speed that would have been unimaginable just a few years ago.

These changes are already having major effects and will continue to do so. Beyond that little is clear, however. In the world of data, size obviously matters. But how much will it matter in the end, in what ways will these effects be felt and by whom. Perhaps most importantly, what can be done to influence this?

While considering the potential impacts of big data in a broad sense, this paper applies these questions specifically to developing countries. Section 1 introduces the research questions and methods used. Section 2 sketches the growing importance of data, considers some definitions of big data, and reviews the literature on the types of impact we could see. Section 3 develops a conceptual framework to assess how big data may create or destroy different forms of 'value', and identifies some of the most important determinants in this respect. Section 4 considers what could be done from a policy perspective over the next few years to increase the likelihood of positive outcomes for developing countries, and minimise the risks of negative effects.

1 Research questions and methods

The questions motivating this research are as follows:

1. Through what channels might big data impact on the options facing developing countries, and the development outcomes experienced in these countries, over the next 10–20 years?
2. What will be the most important determinants of the nature of these impacts?
3. What can be done from a policy perspective to encourage positive impacts, and vice versa?

Before focusing on developing countries, the first part of this study comprises a review of the general literature on the subject, much of which is also relevant for developing countries. Here we sketch the evolution of big data, variously defined, and map the most important current and future impacts that have been identified, both positive and negative.

The second part of the paper unpacks the 'impacts' with specific reference to developing countries, considering the different forms they could take. Not all of these will be positive, and neither will effects fall equally across individuals, groups and societies. Section 3 develops a conceptual framework to begin to capture these diverse impacts.

Nearly all future impacts are contingent upon wider factors, or determinants. A positive impact for a particular group may only happen if something else has already occurred. Similarly, negative results may require a set of circumstances based on preceding actions, or lack of actions. This is important. Those who argue for the benefits of big data often adopt an evangelical tone, while opponents tend to stress the dystopian nature of a 'big data future'. These visions of heaven or hell can sound preordained, but there is nothing inevitable about the impacts that big data will have, or how these will affect different groups. What we do, or do not do, matters. In the light of this, Section 4 considers the policy options that are available to encourage positive impacts and minimise the risks of negative effects in relation to different forms of value.

We are dealing with future events, however, and the future is inherently uncertain. While recognising the difficulties in doing so, we therefore need some way of projecting into the future and exploring ways that big data *could* evolve, and identifying the factors that will influence this. The methodology chosen for this purpose is the Foresight approach, which has been widely used by the private sector and governments[1] to develop and explore the implications of scenarios of the future.

The first step was to identify and interview key experts from a range of relevant fields. This was followed by two participant workshops to identify the key determinants of big data impacts, develop future scenarios based on these determinants, and examine policy options to increase the likelihood of positive impacts occurring. Details of the approach used, the interviews and workshop participants, and the four scenarios developed are contained in the annexes.

As well as the literature reviewed, the interviews and scenarios developed in the workshops are the primary source material used for the policy recommendations contained in Section 4. We would like to thank everyone who contributed their valuable time and expertise throughout this process.

[1] The United Kingdom (UK) Government has been one of the most enthusiastic users of the Foresight approach. For details of the subjects to which this has been applied, see Government Office for Science (2013).

2 Literature review

2.1 Key events and concepts in the 'data revolution'

Data can be defined as 'facts and statistics collected together for reference or analysis' (Oxford Dictionaries 2015). Though important in all human societies, data are being created from more sources and at a much faster rate than ever before. Around 2.5 quintillion bytes of data are created per day and this is doubling every two years (Aiden and Michel 2013). The literature suggests that the huge growth in data production and use has five elements, two social and three technological. The social components are: (i) the ubiquity of personal computing devices, and (ii) the growing importance of data across different markets and sectors. The three main technological shifts highlighted are: (i) more powerful and cheaper computing and networks, (ii) 'datafication', and (iii) a growing rate of data collection. Each of these elements is discussed below.

2.1.1 Ubiquity of personal computing devices

Computers are getting cheaper, more available and more mobile all the time. In the developing world there is much talk of 'leapfrogging' directly to internet-enabled mobile computing, skipping desktop computers altogether (Yonazi *et al.* 2012). While this allows more people access to the information ecosystem, 'leapfrogging' suggests moving to superior technology that will ultimately become the general standard. Although this may be the case with mobile versus fixed line telephones, desktop computing has advantages – for writing papers like this, for example. Mobile devices have advantages too, however, and by 2025, Schmidt and Cohen (2013: 4) estimate that 'the majority of the world's population will, in one generation, have gone from having virtually no access to unfiltered information to accessing all of the world's information through a device that fits in the palm of the hand'.

An important question is whether the information that people will ultimately have access to will be 'unfiltered'. An interesting recent initiative is the internet.org app, which is run by Facebook to expand internet access in developing and emerging countries. 'Free' access to 'basic internet' services is provided, with Facebook gaining access to the user data that are generated. This is not 'unfiltered' access, however, as those using the app only have access to a limited number of partners who are able to access a new group of potential customers. Neither is the access strictly 'free' in that Facebook acquires the users' data and their partners gain access to new potential customers. Rather than universal access to unlimited data, therefore, we may also see the expansion of limited, or 'tethered' approaches, where filtered access to selected sites is used to link firms to vast numbers of new customers in developing countries.[2]

2.1.2 More powerful computing and networks

In the past, processing power was the limiting factor in analysing large data sets. This has changed radically from the early days of computing. From the proliferation of transistors in the 1950s (Brock 2006: 11), to the creation of the first dynamic random-access memory (DRAM) chip in 1966, to the spread of 'cloud computing' today, advances in computing power have led to massive improvements in four key areas: (i) processing speed, (ii) data storage capacity (hardware), (iii) analysis of data (using software), and (iv) connections between data sources and processors via the internet.

[2] Thanks to Linnet Taylor for this point.

i. **Processing speed:** Although mechanical calculating machines existed as early as 1642 (Freeman and Louçã 2001: 309), it was the invention of electronic components that made it possible for modern computers to quickly and accurately complete complex tasks by breaking them into small ones. In 1946, the ENIAC (Electronic Numerical Integrator and Computer), one of the earliest programmable computers weighing over 30 tonnes, could process 500 additional problems per second. This was impressive at the time (CNN Money 2014). Today, IBM's 'Watson' supercomputer can process 500 gigabytes of data per second, can be accessed from anywhere, and is six inches high (BBC News 2014).

ii. **Storage:** Big data must be stored, whether on a computer hard drive, external 'thumb drive', or in the 'cloud' (a remote server). Between 1952 and 1956, IBM developed the first disk drive, which could store five megabytes of data (one three-minute audio file by today's standards), and was the size of a refrigerator. Advances in technology during the 1970s and 1980s rapidly brought down costs and the size of drives, while at the same time increasing storage capacity. In 1978, for example, a 20 megabyte drive took up 800 cubic inches, whereas in 1993 it might have only taken up 1.4 inches, more than a 30 per cent annual reduction in size (Christensen 1997: 7). Because of the far lower costs of storage, companies and governments today are able to store vast amounts of data without necessarily having a specific purpose for doing so. This was simply not possible in the relatively recent past.

iii. **Analytics:** New and flexible software has been developed specifically to handle large data sets and non-uniform data. Analysing data used to require spreadsheets and databases, it can now take any number of structured or unstructured forms. A growing number of companies are specialising in this area, offering customised data analysis for businesses, governments and non-governmental organisations (NGOs).[3]

iv. **Sharing:** The internet has revolutionised data sharing between computing devices, and increasingly connects sensors with computers in what has been called an 'information ecosystem' (Turner, Schroeck and Shockley 2013), or the 'internet of everything'. The growth in our ability to share data and the value we can derive from computer networks can be summarised by Gilder's Law and Metcalfe's Law, respectively. In 1997, Gilder proposed a tripling in internet bandwidth each year for the next 25 years (Gilder 1997). Metcalfe, responsible for inventing Ethernet communication, suggested that the value of any network is proportional to the square of the number of connections that make it up (Metcalfe 1995).

2.1.3 'Datafication'[4]

Waitman likens the recent explosion in the production of data – 'datafication' – to electrification, suggesting that many businesses today are coming to rely on 'data' as they do on electricity (Bertolucci 2013). Data come in different forms and may be created *passively* as a result of other activities (i.e. 'data exhaust'), for example, or generated *actively* (i.e. 'human-generated data'). The increasing abundance of computers and smart devices (those connected to the internet) and data storage, and new techniques, allows automated collection of 'data exhaust' (United Nations Global Pulse 2013). Data from sensors in homes, global positioning system (GPS) devices and records of economic transactions are all examples. Much of this can now be stored.

For actively generated data, social media and other applications (text messages, tweets, phone logs, [video] blogs, etc.) now produce vast amounts of data each year around the world. From 2009 to 2013, for example, Facebook experienced a near sixfold increase in

[3] For a listing of these companies, see The Big Data Landscape (2014).
[4] See Lycett (2013).

active users (Statista 2014). Twitter now records more than 500 million tweets per day (Twitter 2014).

2.1.4 Speed of data production

Many applications rely on the speed with which data can now be collected. Wal-Mart, for example, collects over 250 million transactions per day around the world in its 6,000 locations, making up around four petabytes (4,000 trillion bytes) of data stored daily (Bryant, Katz and Lazowska 2008). This was not possible only a few years ago, and is due to the low cost and ease of storage of large amounts of data.

2.2 Definitions and forms of impact

2.2.1 Definitions

Before considering definitions of big data, it is important to distinguish big data from two related concepts: information and communications technology (ICT) and 'open data'. For Osterwalder (2002), 'ICT encompasses all the technology that facilitates the processing, transfer and exchange of information and communication services'. ICT for Development (ICT4D), therefore, is about how ICTs 'can be used to help poor and marginalized people and communities make a difference to their lives' (Unwin 2009). While big data is built upon ICT infrastructure, it is less concerned with the data being exchanged and more about the value that the data bring. On the other hand, open data is accessible, digital, and free of restriction of use or redistribution (HM Government 2012). Big data is not always open, and at times will not be accessible without special skills or software.

The term 'big data' has its origins in the late 1990s, but only became widespread in 2009.[5] The year 2012 has been described as the 'breakout year' (Lohr 2012), when citations on Google Scholar began to grow rapidly. A number of big data journals have been established recently, including the American Journal of Big Data Research, Big Data & Society (SAGE), the International Journal of Big Data, Big Data (Mary Ann Liebert, Inc.), the Journal of Big Data (Springer) and Big Data Research (Elsevier).

Definitions of 'big data' vary by industries such as information technology (IT), computer science, marketing, social media, communication, data storage, analytics, and statistics. For Kolb and Kolb (2013: 55) this is 'because people use the term to mean a wide variety of things to suit their purposes. It has been used to describe data tools, data sets, questions, problems and answers'. Early definitions often stress the so-called three 'V's of big data, first presented in a paper in 2001 by Doug Laney:

- **Volume:** Big data involves large sets of usually complex data. As Kolb and Kolb (2013) note: 'you know you have it when the size of your data is part of your problem'.
- **Velocity:** A second issue is the increasingly rapid *speed* of data collection, which some practitioners even emphasise over size in their descriptions of big data.
- **Variety:** As described above, data are no longer restricted to spreadsheets or databases. Because it comes from a variety of sources, in a wide range of formats, big data can be structured or unstructured, fixed or constantly updated, or any combination of the above.
  (Laney 2001)

[5] In fact, a Wikipedia page was created for 'big data' in November 2009, but this was quickly deleted. It was recreated a few months later.

While helpful concepts, the 'three Vs' leave out what is perhaps the most important aspect of big data – what it actually does. In 2012, Beyer and Laney updated the three 'V's to address this:

> Big data is high-volume, -velocity and -variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision-making.
> (Beyer and Laney 2012)

The reference to 'insight and decision-making' is in line with other recent definitions, particularly those outside of the fields of data management, computer science, or information technology. Now 'big data' is increasingly defined in terms of its impact in the world. While any definition must involve some combination of quickly expanding, large, and possibly non-uniform data sets, big data is also 'greater than the sum of its parts' (Diebold 2012).

For Mayer-Schönberger and Cukier (2013):

> Big data refers to things one can do at a large scale that cannot be done at a smaller one, to extract new insights or create new forms of value, in ways that change markets, organizations, the relationship between citizens and governments, and more.

Table 2.1 gives some definitions of big data from across a variety of fields, where there is considerable diversity. While there is no need for consensus, defining big data in terms of computing power, data set size or speed seems restrictive, as these characteristics are all context-dependent. As Taylor (2013a) points out: 'in the poorest countries, almost anything becomes big data due to a lack of connectivity, storage and processing capacity'. Ultimately, computing power has always enabled bigger and faster data and this will no doubt continue. Therefore in this paper we emphasise the *impacts* that big data could have and define big data in broad terms:

> 'Big data' refers to our growing ability to generate, manage, analyse and synthesise data to create and destroy different forms of value.

By disrupting the status quo, big data may both create and destroy value. As well as these impacts, we also assume that big data contains at least one of the following (Letouzé 2012):

- Real-time or near real-time data
- Data collected in novel ways (mobile phones, GPS, etc.)
- Structured or unstructured formats (i.e. social media vs mobile phone records).

Table 2.1 Some definitions of big data

General	'Big data refers to things one can do at a large scale that cannot be done at a smaller one, to extract new insights or create new forms of value, in ways that change markets, organizations, the relationship between citizens and governments, and more' (Mayer-Schönberger and Cukier 2013).
	' ... (The) ability to capture, store, and analyze vast amounts of human and machine data, and then to make predictions from it, is what's known as big data' (Feinleib 2013).
	'"Big data" is a term that describes the accumulation and analysis of information' (Smolan and Erwitt 2012).

(Cont'd.)

Table 2.1 (Cont'd.)

Computer Science	'A combination of data-management technologies that have evolved over time' and 'Big data is defined as any kind of data source that has at least three shared characteristics: (i) extremely large volumes of data, (ii) extremely high velocity of data, (iii) extremely wide variety of data' (Hurwitz *et al.* 2013).
	'Big data is a term describing the storage and analysis of large and or complex data sets using a series of techniques including, but not limited to: NoSQL, MapReduce and machine learning' (Ward and Barker 2013).
	'… big data shall mean the datasets that could not be perceived, acquired, managed, and processed by traditional IT and software/hardware tools within a tolerable time' (Chen, Mao and Liu 2014).
IT	'We define big data as (1) technology: maximizing computation power and algorithmic accuracy to gather, analyze, link, and compare large data sets as a cultural, technological, and scholarly phenomenon that rests on the interplay of: (2) analysis: drawing on large data sets to identify patterns in order to make economic, social, technical, and legal claims. (3) Mythology: the widespread belief that large data sets offer a higher form of intelligence and knowledge that can generate insights that were previously impossible, with the aura of truth, objectivity, and accuracy' (Boyd and Crawford 2012).
	'Big data technologies describe a new generation of technologies and architectures, designed to economically extract value from very large volumes of a wide variety of data, by enabling the high-velocity capture, discovery, and/or analysis' (Gantz and Reinsel 2011).
Business and marketing	'Big data refers to our ability to collect and analyze the vast amounts of data we are now generating in the world' (SmartData Collective 2013).
	'"Big data" refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyse' (Manyika *et al.* 2011).
	'... turning transactional business data into decisions and insight using cutting-edge analytics (regardless of where that data is stored)' (Foreman 2013).

2.2.2 Forms of impact

Drawing on the ICT literature (Hameed 2007), we classify the potential impacts of big data as *direct* or *indirect*.[6] Directly, big data may lead to the creation of new markets and new infrastructure, with producers and consumers of data feeding into a system connected by complex physical infrastructure (Bizer, Heath and Berners-Lee 2009).

Indirectly, big data will influence various sectors, institutions and networks through its effects on: (i) production and delivery of goods and services (including public services), (ii) marketing and communication, and (iii) organisational and network structures. In each case, big data has the potential to increase efficiency, reduce costs, and increase productivity, but may also have negative impacts, by increasing 'digital divides' across and within countries, creating privacy concerns and undermining civil liberties, or leading to job losses.

2.3 Direct impacts

2.3.1 New markets

The following sections examine the direct impacts big data may have in computer science, smart appliances and data analysis techniques.

[6] These two different forms of value creation by big data create a self-reinforcing feedback cycle: a growing amount of data can spur new technologies and uses for data, which in turn drives the creation of further data.

a. Data analytics and big data software

The three 'V's of big data have spurred innovations in software, networking, and data analysis and management. These include increasingly sophisticated algorithms, cloud computing and networking processes. New industries, firms and jobs are being created as a result. For example, open source big data software platform Hadoop (and its predecessor MapReduce) were created to store, process, and rapidly analyse massive amounts of data across a number of servers. This innovation has allowed the creation of several companies, which generated US$5bn revenue in 2013, and are growing at 60 per cent per year (Fast Company 2014; Smith 2013).

Data science, formerly a relatively obscure subject, has become a new profession, combining statistics and probability with advanced computing and programming (IBM 2013b). For Cukier (2010), a data scientist 'combines the skills of software programmer, statistician and storyteller/artist to extract the nuggets of gold hidden under mountains of data'. These employment opportunities are well paid and generally require specialised training and qualifications. This increased demand for specialised knowledge is also creating employment opportunities in the education sector.

These impacts may have particular importance in developing countries, where organisations outsource data analysis work to countries where labour costs are low but skills are high. Kenya and Rwanda, for example, are already known for their IT capabilities (Matinde 2013). Schmidt and Cohen (2013) note that 'as fewer jobs require a physical presence, talented individuals will have more options available to them' and will be able to compete across borders. Not all jobs will be high skill, however. Platforms such as Amazon's 'Mechanical Turk' allow the outsourcing of small, usually low-skilled jobs around the world. Many involve simple data manipulation tasks, and only require a basic internet connection.[7]

Table 2.2 Big data technologies

Techniques	**Infrastructure**
Supercomputing, Search and Decision Making, Textual Analysis, Distributed Software, Next Generation Internet, Data Mining, Data processing/analysis, Advanced Programming, Brain-computer Interface, Processing Real-Time Social Data, Crowdsourcing, Data Visualisation, Sentiment analysis, Network analysis.	Ubiquitous Computing, Next Generation Networks, Web hosting, Digitisation, Multisensory Input/Sensing, Sensor Networks, Data Processing and Storage, Cloud Computing, Ambient Intelligence Sensors/Network Sensors, Speckled Computing, Data storage/hosting.

Big data could also create negative effects in these areas. New and desirable jobs and skills could potentially keep less educated members of a workforce from upward mobility. In the case of developing countries, this could contribute to the reinforcement of a 'digital divide' in which the poor are disadvantaged because of their lack of access to technology (WhoIsHostingThis? 2014). For those who are able to gain employment in this space, big data could prove to be a source of 'brain drain' in low-income countries, as has already been the case in certain other fields (Beine, Docquier and Rapoport 2008).

b. New big data 'markets'

Besides software development and data analytics and processing, big data has also led to the creation of entirely new markets and business models, specifically revolving around data.

[7] For more information, see www.mturk.com/mturk/welcome.

Companies specialising in data collection, analysis, prediction and distribution are increasingly turning data into a commodity (Mayer-Schönberger and Cukier 2013). For example, both Monsanto and John Deere, two of the largest agriculture companies in the US, now offer services that will collect data from farmers with the goal of increasing productivity (John Deere 2014; Charles 2014). Monsanto also recently bought The Climate Corporation, a company which collects and distributes detailed, local, agriculturally relevant data (The Climate Corporation 2014). The banking industry has also seen rapid innovation in this space, with big data techniques enabling analysts to detect patterns of fraud, for example, even at the scale of billions of transactions (Turner *et al.* 2013).

As well as the benefits to these private companies, however, buying and selling data sets can create negative impacts, particularly with regards to privacy. Many consumers – given a choice – would object to data about them being bought and sold, but they are not necessarily able to control what corporations do with their data once they have been collected (Craig and Ludloff 2011). This is not just an issue in developed countries. Taylor (2013b) highlights several privacy concerns brought about by big data in developing countries, which often lack the legal and institutional frameworks to ensure data privacy.

2.4 Indirect impacts

Most discussion in the literature to date focuses on the indirect impacts that big data may have. Here we focus on three areas. Firstly, big data can affect the production and delivery of services by firms, governments and other actors through real-time data sharing, learning and automated decision-making, synthesising disparate data sets and improving predictive capabilities. Secondly, big data allows improved marketing through advances in data collection and market segmentation. Finally, big data methods can create organisational improvements in the structure and 'data architecture' of firms, public institutions, NGOs or social movements (OECD 2013; Manyika *et al.* 2011).

2.4.1 Production and delivery of goods and services

a. Real-time data collection and sharing

According to the McKinsey Global Institute (MGI), big data will create value across firms by allowing easier real-time and spatially disaggregated information sharing (Manyika *et al.* 2011). The increased speed with which data can be collected, organised and shared should improve the efficiency and productivity of firms. These effects will be most pronounced in disaggregated manufacturing processes. Radio frequency identification (RFID) tags give traceable signatures to individual products (rather than type of product, as with barcodes). Franks (2012) notes the use of RFIDs in asset tracking, stock management and transportation logistics.

In the power generation sector, real-time monitoring of energy demands and problems can 'transform the grid from a rigid analog system to a smart, dynamic, digital and automated energy delivery system' (IBM 2014). GTM Research predicts that globally utility (gas, water and electricity) companies will spend around US$3.8bn in 2020 on data analytics (up from US$700m in 2012) (Courtney 2014). According to Franks (2012), 'smart grids' would benefit both providers and consumers. First, they allow customers to understand their utility usage patterns, who in turn can lower consumption and save on bills. Second, utility providers are able to switch to 'time of day' or demand-driven pricing models, increasing efficiency and value for customers able to take advantage of differential pricing. Townsend (2013a) also notes that providers are able to better manage demand through load sharing or shedding, and on a 'real-time' response scale.

The real-time nature of big data also has significant potential in the transport sector (IBM 2013a). Much of this lies in the use of 'spatial big data', 'eco-routing', and 'temporally-detailed' roadmaps, made possible by sensors within the transportation network (*ibid.*). McKinsey estimates the potential value of 'smart routing' to be about US$500bn globally (by 2020) in fuel and time saved (Manyika *et al.* 2011). Big data harnessed in real-time could also create environmental benefits, as smart routing reduces transport emissions.

Real-time monitoring and feedback may also create positive impacts for human development. For example, in the health sector, remote patient monitoring (RPM) allows medical staff to monitor patients with smart devices and sensors. These techniques are likely to improve chronic care and home health, especially as life expectancies continue to rise (Blanchet 2008). In developing countries where doctors are often scarce, RPM could give more people access to healthcare, including from providers in different countries (Mechael 2009).

b. Learning, experimentation and automated decision-making

Big data enables firms to learn from the past using historical data to improve production and service delivery, reduce costs, and increase productivity. Cerra, Easterwood and Power (2013) note that companies are able to use experiments to 'help managers distinguish causation from mere correlation, thus reducing the variability of outcomes while improving financial and product performance' (see also Brown, Chui and Manyika 2011). For example, Wal-Mart used its own customer transaction data to find that sales of certain food items rose just before an incoming hurricane. It then began to stock these items near the entrances of its stores, and saw sales increase further (Mayer-Schönberger and Cukier 2013).

Big data also has significant potential for research and development (R&D) in health, with clinical trials potentially moving from randomised control trials (RCTs) to BCTs: 'big-data clinical trials' (Wang 2013). Big data may also facilitate the transfer of knowledge among physicians, by monitoring similar past and present cases (Murdoch and Detsky 2013). McKinsey Global Institute has estimated that the US could save US$200bn annually in reduced health care expenditure through the use of big data techniques.

Big data is increasingly replacing human decision-making with algorithms. In education, information collected about educational progress has advanced from just a few points per month to several thousand in some cases (Saxberg 2013). Big data can potentially help teachers and educational administrators through data-driven decision-making and mining of large amounts of data (Gmür and Schwab 2014: 4). As well as improving decisions about resource allocation, this could also help to target at-risk students. Gmür and Schwab note that big data techniques are able to help identify how likely students are to succeed in one area of learning, based on their past performance in other areas, and provide guidance on when and how to introduce a new concept to a given student (*ibid*.). This would also give individual students insight into their own learning habits (Siemens and Long 2011). Through the advance of ICTs across the world, some of these applications are becoming available in the developing world; already, students in developing countries have the ability to use massive open online courses (MOOCs) for free.

Automated decision-making also has the potential to create severe negative effects. As firms rely more on increasingly accurate decision modelling, traditional analyst jobs are likely to be lost. Mayer-Schönberger and Cukier (2013: 11) note that 'in the future – and sooner than we may think – many aspects of our world will be augmented or replaced by computer systems that today are the sole purview of human judgment'. Some companies have begun to use computer algorithms to make hiring decisions, reducing the need for human resources personnel (Walker 2012). Indeed, sophisticated computer algorithms could replace subject matter experts across sectors, leading to significant job losses.

More broadly, it has been estimated that 47 per cent of all jobs in the US are potentially at risk from automation, including automated decision-making. These tend to be relatively low-skill, low-wage jobs that are repetitive and/or rely on processing large amounts of information on an ongoing basis (Frey and Osbourne 2013).

c. Synthesising data sources

One of the defining features of big data is the ability to synthesise different data sets in ways that were not previously possible. Perhaps the most cited application in developing countries is the possibility of using remotely sensed and crowdsourced data to ‘map’ problems of many types, for example tracking and modelling the spread of malaria, influenza or dengue fever (Hay *et al.* 2013). Different sources of data can be combined with other conventional data (health, education, income, etc.) to build a more complete picture of how disease spreads in a particular context (Venture Beat 2013).

Another example in health is the use of the unstructured data in medical records to answer research questions, which otherwise might have been prohibitively costly. For example, researchers at Microsoft and Stanford University were able to show that mining search engine query libraries could identify drug interactions, providing a cheap and abundant source of medical data (White *et al.* 2013).

From an environmental perspective big data allows better imaging, and therefore greater accessibility, of remotely sensed data, which in turn enables a wide range of applications when joined with other sources of data. For example, in February 2014, Google and the World Resources Institute jointly launched the ‘Global Forest Watch’ (Saatchi *et al.* 2011), allowing internet users anywhere to explore trends in deforestation on a global scale. The Google Earth Engine processes data from the National Aeronautics and Space Administration (NASA) and United States Geological Survey (USGS) satellite services and enables non-technical users to draw meaning from them (Hampton *et al.* 2013).

More negatively, governments are increasingly able to synthesis data, including personal data. This potential has been called the ‘mosaic effect’ (Mazmanian 2014). While many stress the importance of such techniques in areas such as national security, there is also potential for governments to discriminate against opposition parties or even minority groups. Certain developing country governments already closely monitor citizen data – demographic and otherwise. Ethiopia, for example, is the only country in Africa with a nationwide internet filtering program in place, made possible by its monopoly over internet provision in the state (Freedom House 2013). While not all governments will look to use big data in a negative way, the potential to do so is expanding.

d. Prediction and insight

Traditional data modelling makes use of a small sample of observational values in order to make predictions. Big data allows the use of much larger sets of data, potentially increasing reliability significantly. A simple example is the predictive power of Google’s search engine. By aggregating past searches, Google is able to predict what a user would like to find from a few keywords. This has nearly unlimited applications in economic terms, from forecasting consumer prices to environmental and weather conditions, to economic projections for governments. Financial markets have also seen increasing use of big data techniques. Currently, algorithms can trace stock prices and refine predictions for computer-based trading, and financial firms commonly keep databases made up of trillions of data points (Easley, Lopez de Prado and O’Hara 2012).

For some this represents the ‘end of theory’ and the birth of a new way of doing science. Traditionally, the scientific method starts with the generation of a hypothesis based on theory. This is then tested in a sample of data and a conclusion is reached on the statistical significance of the result. By expanding the sample size, potentially to the whole population, some argue that this process is no longer necessary. Rather than using data to test theory, the data themselves become the source of theory, revealed using big data techniques. Others argue that this is misleading, and that it is rarely the case that all data are available. Analysing Twitter feeds, for example, biases results to Twitter users, who are not evenly distributed across socioeconomic and racial groups.[8]

More generally, prediction of future actions of people has obvious potential for negative impacts. The film *Minority Report*, where policemen arrest people for crimes they have not yet committed is an extreme, though thought-provoking extension. More immediately, there are increasing examples of firms using big data techniques to make decisions on credit and even hiring workers. Critics have identified bias, including racial bias, in such results. It is very difficult for these processes to be checked, as the underlying data used are often not transparent. Also, where decisions are based on algorithms, which may process hundreds of millions of data points, it may not be possible to identify the reasons for any particular ‘decision’ (Pasquale 2015).

A positive case for using big data techniques to improve outcomes across a wide range of areas comes from a leading researcher at the Massachusetts Institute of Technology (MIT). In *Social Physics* Sandy Pentland argues that big data allows us to understand the complex ways that people interact, and the role of peer influence, uncovering the rules that govern the resulting outcomes – hence the ‘physics’.[9] Using an impressing array of examples from currency traders in the US to mobile phone users in Côte d’Ivoire, Pentland shows how big data analytics can improve outcomes through manipulating the social forces that influence behaviour. Much of this is convincing and it is clearly the case that very important improvements could be made in ways that enrich peoples’ lives significantly. The more extravagant claims for the benefits of an automated society run by technocrats and algorithms may be a little overdone, and somewhat Orwellian, but this should not detract from the potential benefits of these kinds of approaches.

A more fundamental critique of this approach to ‘optimising’ societies is that it ignores existing inequalities, and so risks entrenching them. A review of the book from the MIT Technology Review puts this critique very well and is worth reproducing here at some length:

> Social physics is a variation on the theory of behavioralism… and it suffers from the same limitations that doomed its predecessor. Defining social relations as a pattern of stimulus and response makes the math easier, but it ignores the deep, structural sources of social ills. Pentland may be right that our behavior is determined largely by social norms and the influences of our peers, but what he fails to see is that those norms and influences are themselves shaped by history, politics, and economics, not to mention power and prejudice. People don’t have complete freedom in choosing their peer groups. Their choices are constrained by where they live, where they come from, how much money they have, and what they look like. A statistical model of society that ignores issues of class, that takes patterns of influence as givens rather than as historical contingencies, will tend to perpetuate existing social structures and dynamics. It will encourage us to optimize the status quo rather than challenge it. (Carr 2014)

[8] See Harford (2014).
[9] See http://socialphysics.media.mit.edu/.

2.4.2 Marketing and outreach

Big data enables organisations to segment their markets, tailoring options to different groups. Cerra *et al.* (2013: xvi) note that this is 'a marketer's dream come true – the ability to use predictive analytics to intercept the consumer with the right message and right offer at precisely the right time to compel purchase or action'. Companies also collect personal data to improve marketing strategies and predict future consumer behaviour based on past personal actions. Online retailers such as Amazon have built business on suggesting future sales based on past purchases by customers. Banks and financial institutions, often as dealers in information, also stand to gain through these techniques (Turner *et al.* 2013). While this looks appealing from the perspective of the firm, others may take a different view. Being at the receiving end of targeted advertising is not necessarily a pleasant experience. There are also serious privacy concerns, particularly as firms increasingly sell personal data to other firms. Cerra *et al.* (2013: xvi) note that 'these relatively uncharted waters are fraught with challenges as marketers struggle to walk the line between consumer exploitation and empowerment'.

As these techniques require the tracking of citizens' and consumers' 'online lives' for 'behavioural targeting' (Craig and Ludloff 2011), they therefore require a minimum level of connectivity to be effective. As more people in developing countries go online these issues will become more salient. Companies looking to expand into fast growing markets in the developing world will continue to focus on those inadvertently sharing their personal details and preferences through online platforms. Ohm (2013) makes a good case against the 'benefits' of big data when weighed against these privacy costs.

2.4.3 Organisational structures

Big data should allow better management and coordination within organisations of all kinds, including loose networks. One of the most important potential applications is in the public sector, which McKinsey calls 'the most fertile terrain for change' with big data (Manyika *et al.* 2011). Governments hold massive amounts of data and are some of the largest purchasers of database software in the world – the US spent more than US$700m on Oracle's data services in 2012 alone (Townsend 2013b). While the idea of 'e-Government' is not new (Howard 2001: 80), big data can help governments to organise and share these large and growing amounts of data.

Many governments have launched open data initiatives, with an emphasis on research and increasing efficiency (Commonwealth of Australia 2013; HM Government 2012; New Executive Office of the President 2012). In 2012, for example, the Obama administration committed more than US$200m to 'new commitments that, together, promise to greatly improve the tools and techniques needed to access, organize, and glean discoveries from huge volumes of digital data' (New Executive Office of the President 2012). Better organised service providers should decrease waiting times and improve outcomes, benefiting the public.

This reorganisation of data within governments (national or local) could also enable greater transparency, creating opportunities for citizen feedback and input into future planning. Assaf Biderman of the MIT likens this to ancient Greek democracy, suggesting that we 'now have a capacity to influence, design, make decisions and suggestions about life in the city by learning about how our city and our environment is managed' (quoted in Kellmereit and Obodovski 2013). Big data thus creates new opportunities for citizens to be informed about what their governments are doing, and new opportunities for governments to be informed about the needs and desires of their citizens.

The use of social media in social movements of various forms is an increasingly studied phenomenon. From the 'colour' revolutions in Eastern Europe to the uprisings of the

Arab Spring, mobiles and social media have been identified as important tools enabling collective action of various kinds. These technologies make it possible for people to coalesce and organise rapidly and over large distances at very low cost. Theoretically, some interesting work argues that, by lowering transaction costs radically, new forms of organisation are now possible that simply could not have happened with traditional (and expensive) management and institutional structures (see, for example, Shirky 2008).

It should be noted that it is not just peaceful social movements who have embraced social media. Terrorist groups such as Islamic State in Iraq and the Levant have become increasingly sophisticated in their use of such techniques, for example. Just as authoritarian governments increase the surveillance of new media to monitor opposition groupings, democratic governments are doing the same to monitor the activities of potential terrorist groups. As with much in the world of big data, the same techniques can be used both positively and negatively.

3 Conceptual framework

The purpose of this report is to explore how big data could have the greatest positive impact in developing countries. There are a number of different types of impact, which can be thought of in terms of the different forms of value that could be created, or destroyed by the evolution of big data. For this report, we suggest the following broad categories:

1. **Economic:** The monetary value that is created for companies through profits, for people (income) or governments (tax revenue). Economic value is destroyed when companies go bust, livelihoods are lost or tax revenues are reduced.
2. **Human Development:** The value created through advancing health, education, or other measures of development, and cultural and social value that may be created or lost (UNDP 2014).
3. **Rights:** The individual or societal utility allowed through freedoms of speech, expression, movement, etc., and through privacy and the right to protection by law (United Nations 2014). Rights value can be destroyed by invasion of privacy or impeding other freedoms.
4. **Environment:** Both the intrinsic value of the environment and the utility gained through environmental benefits (forest and ocean commons, for example). Reducing strain on natural resources can create environmental value.

In some parts of the world, big data will have profound effects in many of these areas. In others, life may go on much as before for the foreseeable future. Where effects are significant, they will also be felt unevenly. Some countries – and stakeholder groups within countries – may see significant value generated, while others could be affected negatively (i.e. lose value) through the same processes.

A country as a whole, for example, could 'gain' through increased national income, while certain groups 'lose' as old industries and jobs are negatively affected. More broadly, some economic sectors are likely to see significant benefits, while others shrink sharply or disappear altogether. In some countries, some groups in society may be empowered by big data techniques and other technologies. In other countries, the same innovations may be a source of oppression and social control. While it is beyond the scope of this report to assess these differential effects, this is an important avenue for future research.

As well as the scale and distribution of impacts, what we might call 'contingent factors' will be crucial in determining outcomes. By this we mean that the nature and distribution of impacts will be affected by other factors. The introduction and effective enforcement of universal rights to privacy at the global level, for example, would have a massive impact on how 'value' is created in this area.

Table 3.1 sketches some of these potential direct and indirect impacts in terms of different types of value creation. In each case, impacts that are potentially relevant for stakeholders in developing countries – particularly low-income countries – are highlighted in bold. We have not differentiated by stakeholder, though this is an obvious area for future research. For direct effects, big data analytics and software are creating new industries, creating a new source of profitability for companies. From a development perspective, however, it is unlikely that these firms will be located in low-income countries, although there is the potential for new sources of employment through outsourcing.

On the other hand, human development may be adversely affected in these countries through 'brain drain'. For contingent factors, the size of the 'digital divide' between higher- and lower-income countries is likely to determine how strong these impacts are, both positively and negatively.

We have already seen the emergence of a large marketplace for data. Many companies buy and sell data routinely, and specialised 'data-broking' companies have emerged. While this generates new sources of profits and therefore employment, these are largely based in high-income countries and are likely to remain so. On the negative side, people in countries at all levels of income may see their privacy invaded and data security breached.

The most important contingent factor here is the extent to which companies are restricted in their ability to trade personal data. As well as straightforward prohibition or laissez-faire policies, individuals could be given rights over their own data to control, including selling elements if they chose to.

Table 3.1 Big data value creation and destruction

Mechanism	Economic	Human development	Rights	Environment	Contingent factors
1. Direct					
i. New markets					
a. Data analytics and software	(+) Profits (+) **Jobs**	**(-) 'Brain drain'**			Technological inequality; 'digital divide'
b. Data marketplace	(+) Profits (+) Jobs		**(-) Privacy**		Rights over use of personal data
ii. New infrastructure					
a. Storage	(+) Profits (+) Jobs				–
b. Hardware	(+) Profits (+) Jobs	**(+) Social (ICT connections)**	**(-) Surveillance**	**(-) e-waste**	Availability of hardware and broadband; privacy rights; recycling
c. Electrical power	(+) Profits			**(-) Pollution** **(-) Degradation**	Means of energy generation
2. Indirect					
i. Production and services					
a. Real-time	(+) Profits	(+) Health (personal)	(+) Security (-) Surveillance	**(+) Environmental monitoring;** smart grids; transport	Availability of health monitoring; privacy rights
b. Learning/ prediction	(+) Profits	**(+) Health** **(+) Education**	**(+) Security** **(-) Surveillance** **(-) 'Fate'**	(+) Lower waste/pollution	Control over personal data
c. Automated decision-making	(+) Profits **(-) Loss of jobs**		**(-) Algorithmic biases**	**(+) Conservation and monitoring**	Cost of automation vs workers; design of algorithms; transparency
d. Synthesising sources, research and insight	(+) Profits	**(+) Early warning** **(+/-) 'Social physics'**	**(+) Security** **(-) Surveillance (mosaic effect)** **(-) Privacy**	**(+) Improved scientific understanding**	Monitoring potential in developing countries; control over personal data

(Cont'd.)

Table 3.1 (Cont'd.)

Mechanism	Economic	Human development	Rights	Environment	Contingent factors
ii. Marketing	(+) Profits		**(-) Privacy**		Control over personal data
iii. Organisational structure	(+) Profits (-) Loss of jobs	**(+) Improved bureaucracy** **(+) More resources**	**(+) Government transparency** **(+) New social movements and collective action** **(-) Terrorist or criminal groups**		Savings are used in a productive way; mechanisms for transparency are designed in; positive forms of collective action are not suppressed while dangerous groups are prevented from flourishing

For infrastructure, new forms of data storage generate profits and some employment, but again this is likely to be focused in the most technologically advanced countries. The same is true for new hardware, but here there will also be human development benefits as connectivity is improved – between members of the diaspora and their families, for example, and between old and new friends as new forms of community emerge. The same hardware may also create risks with respect to rights, as surveillance opportunities are increased, and the growth in e-waste creates environmental risks. Whether and how much increased demand for electricity creates negative environmental impacts depends entirely on the means by which the power is generated.

Connectivity benefits will only materialise if hardware is available and affordable, and internet connections are of high enough quality. The extent to which privacy and liberties are protected is the most important contingent factor with respect to risks around surveillance. While this is a risk in all countries, current trends suggest that the risks are greater in some developing countries. The scale of the negative environmental impacts of e-waste will be determined by the extent of recycling that is undertaken in different jurisdictions.

Turning to indirect impacts, the use of real-time data to increase the efficiency of commercial processes will boost profitability. The development of increasingly small, cheap and sophisticated sensors makes it possible to monitor people's health in real time, potentially creating health benefits. The same types of devices may also be used for national security purposes, with potentially positive impacts (more security) to be weighed against negative effects (more surveillance).

The same is true for the positive environmental benefits that may result from the use of 'smart grids' and 'smart transport', both of which are significant but are likely to be restricted to technically advanced countries. In contrast, using real-time data to monitor and respond to environmental issues such as deforestation has significant potential in developing countries.

The availability and cost of sensors for health monitoring, and the supporting infrastructure, will determine the extent to which this potential can realised. Achieving security benefits, whilst avoiding excessive infringement of personal liberties, requires a delicate balancing of rights. Reasonable people are likely to disagree on where this should be struck.

For learning and prediction, large companies – particularly transnational corporations – are likely to benefit from more efficient processes and decision-making based on mining their own data.

Health benefits, including in low-income countries, may come from big data 'clinical trials', as well as from the tailoring of 'stratified' medicines to individuals, though these benefits are more likely to be felt in higher-income countries. Learning from data will also allow educational inputs to be designed to suit individuals, improving educational outcomes. More negatively, the ability to predict health outcomes may adversely affect the availability or cost of insurance, while the same abilities in education may see those predicted to have less potential being written off, or consigned to this 'fate'.

Enhanced predictive ability also has the potential to improve security, by anticipating and preventing threats. This also raises severe privacy concerns, however, where people may be condemned for acts they have not actually committed, and perhaps never would have. The most important contingent factors relate to control over personal data, particularly whether individuals have the right and ability to retain control over how their data are used and by whom, including government. As above, there is a potential trade-off here with the ability of the state to enhance security and there is no consensus on this question, including between countries.[10]

Automation in the workplace has already boosted profitability in many manufacturing sectors, often at the expense of 'blue-collar' jobs. Automated decision-making based on algorithms has the potential to extend this to 'white-collar' sectors, as it becomes increasingly possible to replace clerical and analyst jobs. While this creates significant economic value for the owners of firms and investors, it destroys economic value for workers. As described above, it has been estimated that 47 per cent of jobs in the US are at risk from automation, and these are disproportionately low and middle-income jobs. While firms in low-income countries are unlikely to adopt these techniques in the foreseeable future, workers in these countries are at risk due to the potential impact on the outsourcing of manufacturing and back office functions.

Algorithmic decision-making is being increasingly used in areas like personal credit and even over decisions on whether to hire particular people. While not consciously designed to do so, such techniques can lead to biases against certain groups, including ethnic minorities (Pasquale 2014). As algorithms are used to make decisions in an increasingly wide set of circumstances, avoiding these risks requires careful design, which is unlikely to happen unless there is transparency and expert scrutiny. Given the proprietary, commercial nature of most decision-making algorithms, this is not currently the case.

The ability of big data techniques to generate insights through synthesising data from a range of sources may hold the greatest potential and carry the greatest risks of all. As with the other impacts, this is likely to be beneficial to firms, boosting profitability. Again, however, it is large firms from technologically advanced countries that will benefit most.

An exception is the possibility of using big data to fine-tune societies to improve outcomes using 'social physics'. As described above, these techniques clearly have significant potential, which should be explored in full. There are limits, however, to what is achievable and also some risks with respect to democratic control and the legitimacy of decision-making.

[10] Interestingly, citizens of the UK appear relatively relaxed over the state infringing personal privacy to protect security. In other countries, such as Germany, the attitude is very different. Historical experience appears very important in shaping these broad differences.

From a development perspective, the ability to synthesise data from different sources has the potential to support powerful early warning systems in areas such as natural disasters, disease, famine and human security. Understanding better and sooner where risks are building up allows more effective targeting of resources and earlier interventions.

We are also only just beginning to grasp the implications of these advances in the physical and social sciences. For some, the ability to synthesis vast pools of diverse data, combined with ever-more sophisticated machine-learning algorithms, means the 'end of theory' and a revolution in the scientific method. While others warn of hubris and point to the limits of such techniques, their impact will be undoubtedly significant.

Set against this potential, the ability to put together information on individuals from a range of structured and unstructured sources – the so called 'mosaic effect' – would give governments unprecedented ability to 'see' their citizens, potentially in real-time. While this could create security benefits, the risks to privacy and civil liberties, particularly with unscrupulous governments are obvious.

These potential benefits and risks apply in countries of all kinds, though the ability to strike the right balance in terms of encouraging innovation and protecting citizens' rights will vary significantly. Developing effective early warning systems requires an extensive and sophisticated real-time data infrastructure. Transparency in how algorithms are used and designed is again important, as is the ability of people to know what data are held on them and by whom, and to be in control over this process. As well as effective legislation, there are technical contingent factors, such as individuals to see all the information that is held, but others – i.e. governments and firms – to do likewise.

Although the greatest potential and risks for big data may lie in these areas, the most noticeable application to date may actually be in marketing. It is now the norm to receive personalised advertisements based on browsing history, or perhaps activities on social media. These innovations have created a new source of profits. Indeed, it is largely this kind of advertising that profits the revenues for companies such as Google and Facebook. Some people may appreciate being presented with tailored advertising. Others find it annoying and possibly slightly unnerving.[11]

Firms such as Jana are combining advertising and marketing directly with expanding internet access in developing countries. The business model links advertisers with customers by providing free internet access to people in exchange for being able to directly advertise to them. This is not 'free' of course, in that the companies must expect to obtain significant business through gaining direct access to millions of new customers.[12]

Where marketing is based upon personal data collected and sold on, there are obviously privacy concerns. For contingent factors, once again the most relevant issue is whether individuals have control over how their personal data are collected, stored and traded.

The final form of big data impact considered in Table 3.1 is on organisational structure. In the private sector, this has the potential to increase profitability through more efficient operations, but may again have negative implications for jobs. For public sector institutions, improvements in organisational efficiency could also generate significant cost savings,

[11] A famous anecdote on this subject is of an irate father complaining when the US retailer Target started sending his teenage daughter pregnancy-related materials, only to discover that the company had correctly identified his daughter's pregnancy from her browsing behaviour.
[12] See https://jana.com/.

creating the potential for value creation via increased frontline spending (human development value) or lower taxes (economic value for citizens).

The potential for big data to increase citizens' ability to better scrutinise their governments and hold them to account is an important final source of potential value. The availability of relevant data and the organisational means to make it available creates this potential. For this to be realised, however, governments need to create the organisational systems and infrastructure to make this technically feasible and democratically legitimate.

Finally, innovations in data software and hardware have enabled new forms of social movement to come together around issues, from boycotting particular brands to overthrowing governments. More darkly, the same techniques have been used by terrorist and criminal groups.

In the final section of the report, we consider policy options in each of these areas.

4 Policy issues and options

In the table below, only potential impacts for developing countries are included, which are coloured, red for negative and green for positive. As we can see, these are quite evenly balanced. In the following subsections, each of the potential forms of value is considered in turn, with policy options to realise positive effects and minimise risks considered.

Table 4.1 Potential big data value impacts for developing countries

Mechanism	Economic	Human development	Rights	Environment	Contingent factors
1. Direct					
i. New markets					
a. Data analytics and software	(+) Jobs	'Brain drain'			Technological inequality; 'digital divide'
b. Data marketplace			Privacy		Rights over personal data
ii. New infrastructure					
b. Hardware		Social (ICT connections)	Surveillance	e-waste	Availability of hardware and broadband; privacy rights; recycling
c. Electrical power				Pollution Degradation	Means of energy generation
2. Indirect					
i. Production and services					
a. Real-time				Environmental monitoring	Availability of health monitoring; privacy rights
b. Learning/ prediction		Health Education	Security Surveillance 'Fate'		Control over personal data
c. Automated decision-making	(-) Loss of jobs		Algorithmic bias	Conservation and monitoring	Design of algorithms; transparency
d. Synthesising sources, research and insight		Early warning systems Social physics	Security Surveillance (mosaic effect) Privacy	Improved scientific understanding	Monitoring potential in developing countries; control over personal data
ii. Marketing			Privacy		Control over personal data
iii. Organisational structure		Improved bureaucracy More resources	Government transparency Social movements Terrorist/criminal groups		Savings are used in a productive way; mechanisms for transparency are designed into the system

These options are based on two sources: the interviews with experts in the field, and two workshops undertaken in London in late 2014. See Annex 1 for a list of interviewees; Annex 2 for details of drivers of change identified during the first workshop; Annex 3 for the four narrative scenarios developed over the two workshops; and Annex 4 for a full list of participants.

4.1 Policy options for economic value creation

The first column contains potential economic impacts. As we can see, this is the least important for developing countries of any of the columns. This highlights an extremely important point: many of the largest potential benefits that could arise from big data are economic in nature, particularly profits for firms. But, to a very large extent, these benefits are likely to be captured by firms in the most technologically advanced countries.

The economic impacts that are most likely to be felt in developing countries concern employment patterns, where both positive and negative impacts are likely to be seen. Workers in developing countries may benefit from increased opportunities to work remotely on data analytics, for example, while outsourced jobs are likely to be lost through the increasing use of automated decision-making. It is not clear where the balance will fall in terms of net impacts, but the evidence suggests that low and middle-income jobs will be most threatened by automation in a general sense. To the extent that high-income, high-skill jobs remain disproportionately concentrated in developed countries, however, the potential benefits may be less in lower-income countries.

The primary reasons why economic benefits are likely to be concentrated in developed countries is that inequality in ICT infrastructure is extreme, and the 'digital divide' in terms of individual skills and access to technology is similarly acute. Addressing this is a huge task that extends well beyond big data. For skills, this has to start with ensuring universal access to basic literacy and numeracy. Beyond that, statistical knowledge and computing skills need to be built into curricula from an early stage. It is impossible to teach computing skills without computers and good quality access to the internet, however. Given the scale of this task, a few projects here and there in low-income countries are not going to make much difference. Technology infrastructure needs to be prioritised and funded, which will require external support in many countries. Investment in the teachers and trainers needed to design and deliver high-quality education is also essential. Once the hardware and connections are in place, this need not all be delivered in person, however, as there is considerable scope to expand cross-border linkages between educational institutions and deliver education digitally.

Even with increased investment, it will take time to develop the skills and infrastructure needed. Realistically, the possibility of most developing countries being able to build successful companies in the big data world appears remote, as they would simply be unable to compete with incumbent firms. This is not a new issue of course, but the same 'infant industries' problem that has been understood for a century or more.

Policy options range from doing nothing and allowing free access to markets for all firms, or restricting foreign firms' access to enable domestic firms to reach a scale and sophistication where they could compete. There are trade-offs, as restricting access to foreign firms is likely to increase costs and reduce quality in most cases, at least over the medium term. There are also limits to what different countries might realistically hope to achieve, even in the longer term. China places severe restrictions on what firms such as Google can do. Part of the reason comes from the desire to censor the material citizens can access, but China is also providing space for its own competitor firms to grow.[13] For most countries, developing a viable competitor to Google will not be an option, but this does not mean that there are no other economic 'niches' where competitive firms could be incubated. This will not just happen, but would need a coherent and effective industrial/ICT policy to be feasible.

[13] Examples are Baidu.com and Soso.com.

More broadly, as we have seen with firms like Google and Facebook, there are significant returns to scale in this sector, suggesting that new sources of economic value may be captured by fewer and fewer firms. There are potential advantages, as these institutions will have access to more data, increasing the big data applications that are possible. There are also obvious risks, however, as this concentrates power in huge, global corporations.

There are three main policy options to address this. First, do nothing and rely on the good will of the owners of big data firms. Second, forcibly break-up these firms to reduce their market power. While the US has a long history of such actions against monopolistic firms, the issue is complicated, as it is the fact that such firms have near monopolies that creates the potential to maximise what can be done with the data. Third, allow the tendency to scale to continue, but regulate these firms strongly.

The second form of economic value considered is employment, where the issues are somewhat different. As more analytics and processing work can be done remotely, internet connections improve and people acquire the necessary skills, geography becomes less important. There is thus scope for workers in developing countries to have access to new forms of employment, though this needs to be set against the risk to such jobs by the increasing automation of decision-making.

What may decide this balance is relative cost, suggesting that as the cost of automation falls then workers in all countries will need to lower wages to compete. As has been the case in some aspects of manufacturing, the risk is of a ‘race to the bottom’ with countries competing to offer low-cost data services. While a number of low-income countries, such as Kenya and Rwanda, for example, have prioritised the development of these kinds of jobs, and invested heavily in the necessary infrastructure, the results remain modest.

As has been shown to be the case in manufacturing, this type of strategy is likely to be more sustainable if combined with the development of viable domestic firms. Again, therefore, the need to build data ambitions into a broader industrial strategy may be important.

Key policy options:

Global

- Explore whether breaking up incumbent monopolies to create smaller firms would lead to better outcomes than regulating these firms to increase the public benefit of big data
- Consider these options in the light of what is likely to be most feasible from a political economy perspective.

Developing countries

- Close basic skills gap in developing countries
- Build ICT infrastructure in combination with statistical and computing skills base
- If desirable, develop industrial policy to create competitive domestic firms in big data ‘niches’ that is potentially feasible given country circumstances
- If not, do not waste resources, but focus policy towards maximising the potential benefits and minimising the risks of other forms of big data value.

4.2 Policy options for human development

From a developing country perspective, this form of value has the greatest potential for developing countries, as shown by the balance of positive and negatives in Table 4.1.

For the latter, the primary concern is that the growth of the big data industry exacerbates the ‘brain drain’. This relates to the issue of the ‘digital divide’ in the sense that efforts to narrow the skills gap and build capacity could result in (expensively) trained workers emigrating to a developed economy.

From a policy perspective, this underscores the importance of developing viable domestic firms. Other than direct restrictions on emigration, the only way to prevent a brain drain is to provide attractive jobs in the domestic economy, which is only likely to be possible if potentially viable economic niches are identified and developed as part of a long-term, strategic industrial policy.

An area with the potential to create significant value is the enabling of communication and interactions via online platforms such as Skype, Facebook, Twitter and Instagram. The huge growth in the popularity of these applications, particularly in developing countries in recent years, shows how much value they create, and how they are reshaping the ways in which people interact and even what we mean by terms such as community. For this value to be realised, however, people need access to internet-enabled devices, a good quality internet connection and a reliable source of power. Very large numbers of people in developing countries have none of these things, and could not afford them even if they were available. As one interviewee put it nicely: ‘[big data proponents] need to realise that not everyone in the world is middle class, lives in New York and owns an iPhone.’

One of the areas that has created the most excitement is using big data for learning and prediction, particularly in health and education. In health, individualised ‘stratified medicine’ has the potential to improve health outcomes by tailoring interventions to people’s biological profiles. As well as physical health, this can be applied to wellbeing more generally, where big data could increase understanding of the determinants of people’s wellbeing and help design interventions accordingly. In education, using big data to understand individual patterns of learning and designing education programmes accordingly offers similar potential to improve the effectiveness of interventions. The risk, as described previously, is that predictions of individuals’ health or education prospects become ‘self-fulfilling destinies’ that they are unable to escape as it is assumed that interventions that could change this will fail. Moreover, the identification of particular biological markers is likely to radically change the terms on which health and life insurance is available.

While these are profound issues, it is not clear how important they will be in developing countries in the foreseeable future. As well as requiring a significant pool of data from which to learn, which is unlikely to exist in many contexts, designing and delivering stratified healthcare or bespoke education to individuals is likely to only be affordable for the relatively wealthy. As illustrated in one of our scenarios, the risk is that the health and educational inequalities are increased by the application of big data, with those that can afford the improved interventions benefiting accordingly, particularly in health.

The use of technology does have the potential to make education available far more widely than in the past, but this is not necessarily the big data-powered, bespoke form that proponents describe. Bringing education to those that previously did not have access is a good thing of course, but if the quality of the education received by the relatively wealthy also increases, educational inequalities will remain pronounced.

Perhaps the most important potential application of big data from a development perspective is through early warning systems based on the synthesis and analysis of diverse sources of data. As some of this is real-time or close to it, and sources are diverse and independent, such systems can be far quicker and more accurate than traditional approaches.

Applications range from natural disasters to disease outbreaks to food insecurity, but could also be used to identify stresses in livelihoods.

Much excitement was generated when Harvard's HealthMap big data programme broke the news of the Ebola outbreak in West Africa before the World Health Organization (WHO). For enthusiasts, this showed that big data could 'beat the news', enabling a quicker and more effective response. A closer look at the story, however, shows that this was not strictly true. In fact, the government of Guinea had already announced the outbreak at a press conference broadcast on national television, and informed the WHO. It was this news item that HealthMap picked up (Foreign Policy 2014).

To be really ahead of the game, big data needs to be able to monitor grassroots-level activity, particularly on social media, before they are filtered through news outlets. A problem is that a lot of these data are not in English, and big data remains heavily skewed towards English language material, despite improvements in translation software and analytics (*ibid*.).

The most ambitious applications of this kind are of the 'social physics' variety discussed previously. There is clearly potential to design social systems and infrastructure more effectively in developing countries. The fact that these are currently lacking in many instances, creates the potential for intelligent design from the outset, rather than restructuring existing systems, which is inherently more difficult. For this potential to be realised, however, the resources needed to invest in these systems still need to be found. The more general critique of these approaches also applies in developing countries, perhaps more so given very large inequalities and power differentials that apply in many countries. Designing policy to 'optimise the status quo' in such conditions, risks entrenching inequalities and also distracting attention from the need to address them.

Applying these techniques to organisations, particularly government institutions, has similar potential to improve efficiency, saving resources and improving outcomes. The often cited McKinsey study on potential savings in the healthcare sector applies to the US, however, which is one of the richest countries in the world and also spends more on healthcare than any other country. The situation in developing countries is completely different of course, where health expenditure is a much smaller slice of a far, far smaller cake. The real benefits of these approaches may also lie in the intelligent design of new systems, therefore, rather than the reform of existing ones. Again, this raises the issue of where the resources to invest in these systems can be found.

Key policy options:

Global

- Develop mechanisms to transfer the health and education benefits of big data that emerge to developing countries.

Developing countries

- Develop successful domestic tech businesses to prevent 'brain drain'
- Ensure citizens have access to internet-enabled devices, the internet and affordable power
- Address the English language bias of big data early warning programmes, and increase their ability to tap into grass-roots real-time data sources
- Generate revenues to invest in 'smart' urban design and transport infrastructure
- Generate revenues to invest in 'smart' health and education systems

- Do not allow the ability to improve the ‘intelligence’ by which social systems, infrastructure and government services can be delivered detract from problems of inequality.

4.3 Policy issues for rights

Questions regarding the right to privacy and control over personal data are not straightforward. Some of the most promising applications of big data, such as its ability to increase the speed, power and accuracy of early warning systems, could be compromised under certain scenarios. As one interviewee put it: ‘people have the right to privacy, but they also have the right to food and to health.’ The questions, therefore, are where the balance should lie, how this should vary according to the type of data, and how can this be enforced in very different country contexts?

The issue of security is a good example. A primary responsibility of any government is to protect its citizens. Governments of all kinds have used big data techniques to identify and monitor potential threats. These same techniques, however, are also used to identify and monitor opponents of the government, or just groups with views the government does not like. In practice, it is not obvious how big data techniques could be restricted to the first type of example, as the way they are employed is dependent on what the government considers to be a ‘threat’, and this varies hugely.

The same techniques, as we have seen, have been used to mobilise social movements. How policy can facilitate – or at least not oppose – the emergence of a plurality of ‘positive’ social movements, while doing the opposite to ‘negative’ groups is not obvious. This is particularly the case as different governments will have a very different view of which groups are ‘positive’ or ‘negative’. It is very difficult for any government to argue that certain groups should not be suppressed in other countries, even where their governments disagree with them, if other groups – no matter how distasteful – are suppressed in their own country. For this reason, there seems no option but to err on the side of freedom of speech and association, within the law, for countries of all kinds.

The citizens of different countries take very divergent views on issues of security versus liberties. As well as historical experiences, there is also the issue of awareness. Understanding of what is now possible with big data is limited to a very small group of people in most countries. The first step in striking the right balance is therefore to raise awareness. In the words of one interviewee: ‘to equip people with a big data mindset.’ As well as individuals, encouraging the establishment of expert NGOs to act as watchdogs of government and corporate behaviour with respect to big data would also be a positive move.

An important issue is whether these questions should be settled nationally or at the global level. A concern with the former is that countries with the most lax data standards will attract unscrupulous operators – the classic ‘race to the bottom’. A concern with the latter is that global agreements are extremely difficult and time-consuming to negotiate, and very hard to implement.

Assuming for a moment that an agreement was possible – the United Nations (UN) Declaration of Data Rights, for example – who should be involved in its design? It is important to note that this is not primarily technical but a question of ethics and law. As a result, it should not be left to computing and data experts, but should include social scientists and legal experts. Given the very limited understanding of where the correct balance should be struck between data being open and closed in different circumstances, there is a good *prima facie* case for establishing such an expert panel under the auspices of the UN.

An obvious distinction to make is on types of data. Most people would be in favour of open data for governments, for example, as this has huge potential to improve the accountability and effectiveness of governments across the world. The question is how to do this, as the countries where this would be the most beneficial are those where governments are likely to be most resistant. There is no simple solution, though enshrining citizens' rights to data about their own government's activities would be the ideal way of addressing the issue. In the absence of such a measure, it is incumbent on those governments which take a progressive approach to develop mechanisms and demonstrate their advantages in terms of more effective and legitimate governance.

The issue of personal data is different and raises important questions. What counts as 'personal' in this sense, and under what circumstances would restricting its use prevent important big data benefits being created? One way of looking at this is to consider who the beneficiaries are. Early warning systems may rely on what could be considered personal data under some definitions. The intended beneficiaries, however, are the same people whose data are being used. This is different from commercial firms using personal data to increase their profitability, either by targeting individuals directly, or selling data to other firms. An important issue is the opacity of the algorithms that companies use for 'behavioural targeting', which have been shown to be biased towards particular groups in some instances, though not necessarily deliberately. Distinct from issues of whether this should be restricted, it is clearly important that such approaches are more transparent and subject to expert scrutiny.

For governments, it may be less clear who the beneficiary is, with most governments of all kinds generally claiming – and no doubt believing in many cases – that their surveillance activities are for the public good. As with the commercial sector, it is important that the algorithms that security agencies use are open to scrutiny by independent experts to avoid bias and to be able to reassure the public that the limits placed on government action – whatever these may be – are being respected.

Particularly important appears to be preventing the 'mosaic effect', where governments are able to use big data to piece together disparate information about citizens and 'see them' in real time. On this technical solutions may exist, with Estonia widely cited as an example of best practice. Estonian citizens are able to see all the information that the state holds on them, but state agencies can only see the data that are relevant for them. If there are errors, citizens are able to correct them. Being able to see what information is held about you is unarguable with respect to government data, and there seems to be no reason why this should not be extended to firms. On this point, people have very different levels of tolerance about the personal information that they are comfortable being used, with younger people generally being more relaxed. While it is important to raise awareness on these issues, it is also the case that genuine differences exist, and should be taken account of.

One proposed solution is to ensure that all individuals have the right to control their own personal data, and can choose to sell as much or as little of this as they like. For some, the result would be a global 'marketplace' of data which people could choose to engage with on their own terms, trading data as they saw fit.

The principle that people should have control over their own data is surely right. For many concerned with issues of privacy, the Organisation for Economic Co-operation and Development (OECD) Privacy Principles, which underpin the approach adopted in the European Union (EU), offer a good template. There are important differences between the approach adopted in the EU and the US in this regard, however, with EU regulation being considerably tighter. This is seen most topically in disputes between the European

Commission and US companies such as Google, which are characterised in Europe as defending citizen's rights and in the US as protectionism.

These disputes make it clear that reaching global agreement on these issues would be far from straightforward. An alternative with respect to the private sector would be for individual countries to extend the privacy rights their citizens enjoy to other countries where their companies operate. In developing countries where issues of personal privacy and data rights are not prioritised, firms from more heavily regulated markets should not be able to exploit the personal data of people in ways they could not do at home.

The vision of a 'marketplace of data' in the context of extreme inequality should be treated with care. Market systems are based on free exchange. The circumstances of the parties, however, has a strong influence over what they 'choose' to exchange. Desperate people will sell anything they have, including their personal data, of course. As one workshop participant put it: 'the risk is that the rich pay for their privacy, and the poor pay with their privacy.' This suggests that, at the very least, the implications of such an approach need careful research.

Thus far, issues of rights have been discussed from a legal perspective. There is also a technical approach, however. To work, the Estonian model described above needs it to be impossible for governments to circumvent controls on the mosaic effect, even if they want to. Some institutions are working to make unbreakable encryption software open source and available to all. The question, which others dispute, is whether it is actually possible to make unbreakable encryption devices. If it is, then technical solutions to privacy concerns are possible. If not, they are not.

Key policy options:

Global and developing countries

- Raise awareness of citizens: create a 'big data mindset'
- Encourage the establishment of expert NGOs to act as watchdogs of government and corporate behaviour
- Establish a UN panel to explore the design of a Declaration of Data Rights to balance privacy with the benefits that access to data could bring and ensure a level playing field. The panel should be comprised of social scientists, ethicists and legal experts, as well as technical experts
- Enshrine citizens' rights to access open data on their government's activities in the process and develop mechanisms to encourage countries to adopt this approach in the meanwhile
- Establish the principle that all individuals should have the right to see and control the information that is held about them by governments
- Establish a similar principle with respect to the commercial sector. Explore the implications of allowing people to sell their personal data if they choose to, with a focus on the impacts of this in poorer countries and potential for exploitation
- Practice and encourage a presumption in favour of freedom of speech and association for all groups, regardless of how disagreeable their views are to the majority, within clear legal boundaries
- Require companies from developed countries to employ the same approach to data privacy in all countries that they operate in as they do at home
- Where technical solutions are possible, employ them to enable the enforcement of global and national rights principles reached through democratic and legitimate means (rather than allow technical innovations – or their lack – to determine the approach to privacy and data rights).

4.4 Policy issues for the environment

From an environmental perspective, big data has the potential to both create and destroy value. On the negative side, big data applications require significant energy inputs. This has the potential to generate carbon emissions and other pollutants, raising climate change and localised air quality concerns. On the other hand, if the energy is generated from renewable sources, there may be no negative environmental impacts. The policy implications are clear.

From desktops, to laptops, tablets and mobile phones, the massive increase in data-related hardware has generated e-waste on a huge scale, with significant environmental impacts. The extent to which these can be managed depends on whether recycling techniques are improved and made the norm. In developing countries, a large proportion of this work takes place in the informal sector and can be surprisingly efficient and cost-effective. From a policy perspective, there is a good case for learning from and building on these successes, rather than trying to eliminate and replace them (Widmer *et al.* 2005).

The potential to create positive environmental impacts is significant. First, big data applications have become increasingly important ways of monitoring ecosystems such as forests with far greater accuracy and timeliness than was the case in the past. Accurately tracking deforestation is fundamental to preventing illegal logging, and is also likely to be central to the agreements reached under processes such as REDD+. Big data techniques are also being used to support biodiversity by monitoring species and to identify the most suitable locations for endangered species to flourish (Worth 2013; Sanders 2014). From a policy perspective, increasing support for research in this area, and facilitating collaboration between researchers in different parts of the world, would be likely to yield benefits.

Perhaps the largest long-term positive environmental impact may come from the scientific advances that big data enables. By definition it is not possible to be specific about breakthroughs that have not yet happened, but if only a small number of the major environmental challenges we face are mitigated through big data-enabled science, the results could be very significant.

5 Concluding remarks

There is no doubt that the impacts of big data – broadly defined – will be very large. These will be varied in nature, relating to economic, social, environmental and rights-based outcomes. As we have seen, there is much potential for value to be created in all of these areas, but also for existing value to be destroyed. It is very clear that these costs and benefits will not be evenly distributed between and within countries, but also that these outcomes can be influenced by policy if we choose to act.

Doing nothing is also a choice of course. The pace of innovation and change in the world of big data continues to accelerate. This will no doubt continue. Companies will continue to find new ways to improve their production and marketing by obtaining and analysing ever-larger, more accurate and timely data sets. Some firms will be scrupulous about the use of these data, others less so. Governments will continue to gather and analyse data for their own purposes. Some of these governments will be concerned with the wellbeing of their citizens, and respectful of people's rights to privacy and other civil liberties. Others will be less so.

In the absence of a strong, effective and enforceable framework to regulate this process, then some of the more dystopian visions of surveillance states are likely to become a reality in some countries. Similarly, the power of large, data-hungry corporations to manipulate personal data – and the individuals who generate these data – for own commercial ends can only continue to grow.

These risks apply in all countries, but may be more acute in some developing countries, given less well-developed protections for civil liberties, and less ability to curtail the activities of global corporations. At the same time, however, many of the largest potential benefits of big data in areas such as health and education are to be found in developing countries also.

Realising these benefits requires a robust framework that protects people's rights but also gives them the confidence to share the data that are needed for these benefits to be realised. Giving individuals, communities and societies access to and control over their own data is at the heart of this.

In this report we have suggested a number of policy interventions with the potential to further this goal. If the impacts of big data are to go beyond profits for global corporations and greater surveillance capacity for governments, the process of agreeing and implementing policies to achieve this needs to start urgently.

Annex 1 Interviews

Fifteen interviews were undertaken with experts from government, the private sector, academia and NGOs. The aim was to elicit views on: (a) the types of impact big data could have (both generally and specific to developing countries), (b) which factors will be most important in shaping these outcomes, and (c) what policy interventions could be made today to have the greatest positive impacts in the future?

Interviewees

- Michael Talbot and Chris Fleming (Respectively: Head of Government Horizon Scanning Unit and Head of Data & Analytics, Government Office for Science)
- Viktor Mayer-Schönberger (Director, Oxford Internet Institute)
- Linnet Taylor (Academic, University of Amsterdam)
- Jim Fruchterman (President, Benetech)
- Carly Nyst (Director, Privacy International)
- Emma Uprichard (Academic, University of Warwick)
- Owen Barder (Director, Centre for Global Development, Europe)
- Chaitan Baru (Distinguished Scientist, San Diego Supercomputer)
- Robert Kirkpatrick (Director, UN Global Pulse)
- Nathan Eagle (Co-Founder, Chief Executive Officer, Jana)
- Kerry Albright (Senior Advisor, Data and Innovation, UK Department for International Development)
- Duncan Green (Strategic Advisor, Oxfam)
- Tim Unwin (United Nations Educational, Scientific and Cultural Organization (UNESCO) Chair in ICT4D)
- Mark Lycett (Professor of Intelligent Information Management, Brunel University)

Annex 2 Identifying drivers of change

In the September workshop participants identified key drivers of change that might shape future states of the world and affect big data outcomes. They were categorised using the 'STEEP' framework:

- S: Social
- T: Technology
- E: Economy
- E: Environment
- P: Politics and government

Participants were asked to think as broadly as possible and to identify what they considered the most important drivers in each category. Each driver was translated into a spectrum of possible outcomes – e.g. privacy versus surveillance, or better health versus more disease. The results of these discussions are summarised below.

Social

(S1) Privacy and digital identities: The prevalence of social media and their impact on societies was a common theme identified by the participants. The main concern was the intrusion on privacy, and the construction of (public) digital identities, as well as the reaction to these phenomena which may discourage the use of social media. Spectrum: privacy and ethics → surveillance and identity theft.

(S2) Population dynamics: Participants shared ideas around population growth, migration patterns, and demographics, especially the ageing population. Drivers of change identified ranged from the current population dynamics (with young poor population in developing countries and old rich population in developed countries) to the construction of a less stratified global population.

(S3) Education: Another important driver considered was education, particularly the extent to which people (especially women) will become more educated in the next 20 years, especially through new forms of education (digital/online). Spectrum: current education rates → universal education.

(S4) Knowledge creation: How will the changes in education and access to information influence knowledge generation? On the one hand, an excess of information may reduce the perceived value of research and evidence. Alternatively, new technological advances can promote a stronger involvement of citizens in data generation and knowledge creation. Spectrum: open, accessible and participatory knowledge → closed and commercialised knowledge.

(S5) Health: Lifestyle diseases overtaking infectious diseases, diseases induced by technology as well as antibiotic resistance were some of the health concerns shared by the participants. Spectrum: increase of health solutions → increase of disease/pandemics.

(S6) Socioeconomic dynamics: Potential changes in socioeconomic dynamics were discussed in a variety of ways. For example, new forms of employment (no workplace/working remotely) may drive changes and shifts in peer networks and lead to more individualistic societies. The further growth of the 'top 1 per cent', and the consequent increase of inequality was also considered. Other participants proposed dynamics leading to the opposite outcomes. Spectrum: unequal and individualistic societies → more equal societies (in terms of gender and income).

(S7) Religion: Several participants shared thoughts regarding the future of religion, faith and spirituality in its diverse forms. The reversal of enlightenment and the rise of religious fundamentalism and sectarianism were contrasted with other hypotheses around the rise of less divisive religious approaches, or new forms of organising (or not organising) religion. Spectrum: fundamentalism/sectarianism → common spirituality/tolerance.

(S8) Culture: The impact of new economic and technological forces on culture and behaviour was considered, with the central question being whether these would encourage plurality and diversity, or the opposite. Spectrum: cultural diversity → homogeneity.

Technology

(T1) Data: The data deluge and 'datafication' of everything has implications for data infrastructure, storage and transport. Ideas around data being processed by technology were also shared, drawing on the concept of the 'internet of things'. Spectrum: data collected and stored through technology → data analysed and processed by technology.

(T2) Information and communications technology: For ICT, participants envisioned two main possibilities in terms of the direction of change: a world of increasing connectivity through mobile and internet technology, or the 'Balkanisation' and privatisation of the World Wide Web (www). Spectrum: an open interconnected world → closed and privatised access to ICT.

(T3) Cybercrime: The increase in data collection and the value associated has led to the rise of cybercrime in its different forms (cyberterrorism, cybercrime tailored to individuals, etc.). Spectrum: secure data storage and use → continuing growth of cybercrime.

(T4) Innovation: The path of technological innovation was another driver identified. From new applications of already developed technologies (such as 3D printing, bio-tech and nanotech, and changes in ownership of satellites) to nascent technologies with strong transformative potential: cryogenics, genetic engineering, robotics and artificial intelligence, etc. Spectrum: stagnation of technological innovation → explosion of technological innovation.

Economy

(E1) New economic models: The possible decline of existing economic models and rise of new models was discussed. Possibilities such as a radical fall in employment due to technological change, the death of the physical economy and the collapse of social security systems were imagined. For new economic approaches, the rise of digital money and the shared economy, personalised products and pricing, or the flourishing of environmentally sustainable business models were examined. Spectrum: crisis of traditional economic systems → innovative economic models.

(E2) (De-)Globalisation: Participants discussed the possibility of 'de-globalisation' and the growth of regional or local markets, or the reshaping of the global order by new powerful actors such as the BRICS (Brazil, Russia, India, China and South Africa). Spectrum: continuing globalisation (including the possibility of a new global landscape) → de-globalisation.

(E3) Inequality: Concerns about the rising income inequality (global and national) were also included in this section. Spectrum: increasing economic inequality → reducing economic inequality.

(E4) Finance: Several participants envisioned the possibility of further financial crises at the global level. Spectrum: financial system stability → instability.

Environment

(V1) Climate change: The drivers of change in this subcategory range from climate change not being true/being controllable, to a catastrophic event with regards to climate, triggering a radical shift in public behaviour towards the environment. Spectrum: climate chaos/crisis → increase of mitigation/adaptation techniques.

(V2) Resources: Resource scarcity was a recurrent theme. As well as food, water and mineral scarcity, special emphasis was placed on energy. Participants considered possibilities such as new forms of energy (wireless, fusion, etc.), free energy or increasing energy scarcity. Spectrum: scarcity/constraints → availability/new solutions.

(V3) Health: Public health concerns such as new diseases, or the health consequences of intensifying urbanisation and pollution were also considered in relation to the environment. Spectrum: solutions to environmental health problems → rise of environmental diseases/threats.

Politics and government

(P1) Global order/geopolitics: Within this subcategory drivers of change ranged from the end of the distinction between developed and developing countries to the irrelevance of multilateral institutions, to the formation of a new global order where the rising powers (especially China) have a growing influence. Spectrum: old global order → new global order.

(P2) Political regimes: The expansion or decline of democratic models is another important area of uncertainty over the next 20 years. Spectrum: democratic regimes (including new models of democracy) → authoritarian regimes.

(P3) Public sector/private sector: Participants considered whether power would shift from governments to businesses (i.e. controlling data, information, content, etc.). Spectrum: power and relevance resides in public sector → power and relevance resides in private sector.

(P4) Data/ICT: Options ranged from new forms of control (colonialism 3.0, manipulation and cyber-tactics for mass social control) to a more open approach to data and ICT based on freedom of information and intellectual property reform to promote openness. Spectrum: closed access/perverse use of data and information → open access and constructive implementation of the potential of ICT technologies.

(P5) Conflict: Participants envisioned several possible forms of conflict and war: Third World War (which could take the form of a virtual/data war), global ethnic and ideological cleansing and terrorism. Spectrum: war → peace.

In summary, the main drivers of change identified were:

1. Social: population dynamics; privacy; religion; education and knowledge creation; health; socioeconomic dynamics and culture
2. Technology: ICT; data; cyberterrorism; technological innovation
3. Economy: new economic models; globalisation; inequality; financial systems
4. Environment: climate change; resources; health and energy
5. Politics and government: geopolitics; democracy; public vs private sector; data/ICT; conflict.

Annex 3 The Foresight scenarios

As described in Section 1, the Foresight methodology was used during two workshops to develop future scenarios and examine policy options to encourage better outcomes. The starting point for participants was the identification of drivers of change identified in the STEEP[14] framework. See Annex 2 for a detailed breakdown. Through discussion and voting, participants reduced the drivers identified to the ten they thought most important in terms of their potential impact on future 'states of the world'. This shortlist was then examined to identify common features, and assess whether drivers could be combined into more generally applicable concepts. We also tried to conceptualise these drivers in terms of a spectrum.

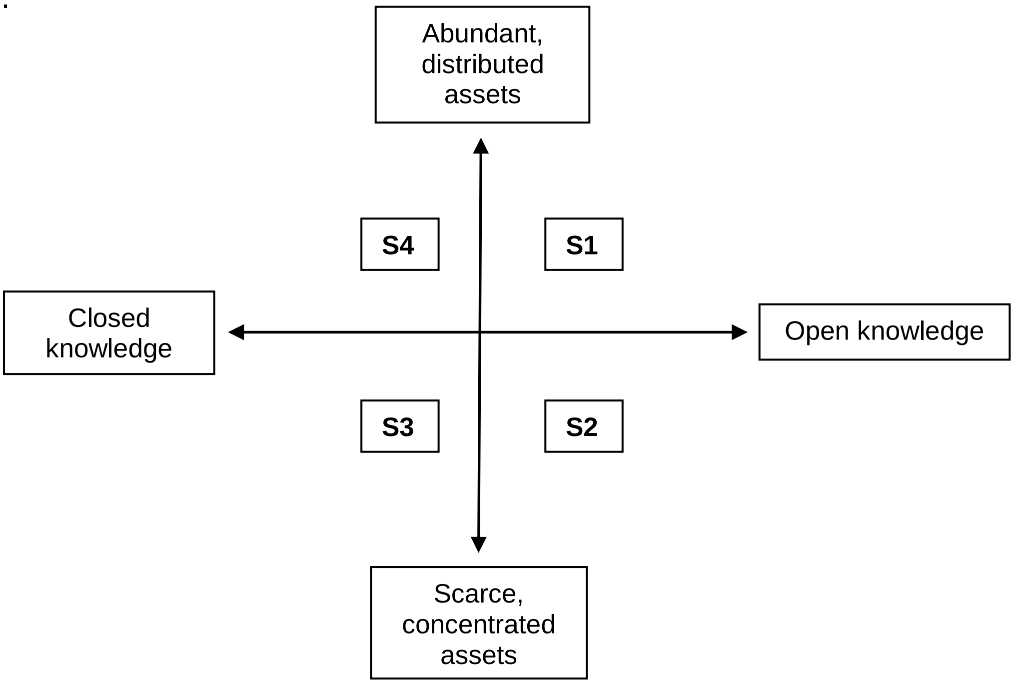
The results of this exercise saw two main drivers emerge as the most important in terms of their likely impact upon future outcomes, and broad enough to encompass many of the more specific drivers discussed in Annex 2. These were:

i. Access to knowledge and data

This driver captures whether knowledge of all kinds is openly accessible or closed and controlled.

ii. Asset scarcity and distribution

This driver captures whether resources are abundant and widely distributed, or scarce and concentrated in terms of ownership. 'Assets' are defined broadly, including wealth, energy and environmental resources, as well as human assets such as health and education. The two drivers were then used as axes to form the basis of four distinct scenarios, as shown below.



[14] Social, Technology, Economy, Environment, Politics and government.

S1: Scenario 1 is a world of abundant assets and relatively dispersed ownership, where knowledge is openly available, but regulation and surveillance are pervasive.

S2: Scenario 2 is also a world of freely available knowledge, but in a context where many important assets are scarce, and ownership of these assets is relatively concentrated.

S3: Scenario 3 is also a world of resource scarcity and concentrated ownership, but where access to knowledge is tightly managed and controlled.

S4: Scenario 4 combines tight management and control of knowledge with abundant assets, the ownership of which is relatively widely distributed.

In order to bring these scenarios to life, participants were asked to develop narratives to describe these 'worlds' in more detail, and give some sense of what life in these circumstances could be like. The following four narratives draw heavily on the original versions, but have also been informed by interviews with experts in the field.

Following the Foresight approach these are presented as fictionalised accounts of the world in 15 years based on the logic of the drivers behind each scenario. They are deliberately extreme versions. In reality, the future will no doubt contain elements of each of these scenarios and there is plenty here that will not happen. Considering development implications in these stylised scenarios, however, will help us to identify the range of outcomes that *could* happen, to examine the most important development implications in each case, and to explore policy options.

Each scenario is followed by a summary of the key drivers that underpin them.

Scenario 1: a friendly Big Brother?

A world of abundant assets and relatively dispersed ownership, where knowledge is openly available, but regulation and surveillance are pervasive.

It was 2019 that everything changed. That was the year the economies of the world cracked after the Great Financial Crisis. The scale of the crash was way beyond anything that had been seen before, and since then we've been slowly putting things back together, but better and fairer.

Today, most people have what they need to live a decent life, and the huge inequalities that used to exist within and between nations are largely a thing of the past. The shock of the crisis was big enough to make politicians really cooperate, and to wipe out the biggest obstacles to change. Without this we would have no doubt continued to muddle through, but faced with the collapse of the financial systems that trade and the real economy relied upon, there was no option but to change. The global economic system was irreparably broken – we needed something new.

Following a series of emergency summits at the UN, nations pooled and invested what resources they had to build infrastructure and fund R&D into new technologies. The results were rapid and impressive. Many infrastructure needs were met and much needed jobs created in all countries. Demand began to be rebuilt. Today we all enjoy abundant clean energy as a result of the research started at that time.

A new monetary system was introduced after the collapse of national currencies, with a global reserve currency managed for the interests of the world economy rather than just one country. Just as quantitative easing had supported national economies after the financial crisis of 2007–08, we learned that the same applied at the global level, and that – if well managed – financial resources could be created and used to meet people's needs and support economic development.

The successes of cooperation led to more cooperation, until something resembling a global government, or global technocracy, emerged. The elites that managed this system had clear goals: to prevent a return to the economic and environmental instability of the past, and to manage the global economy such that people's basic needs were met everywhere and the natural environment was safeguarded for future generations.

Information is now freely available. Everyone in the world who wants it has access to all the information that has ever existed. As with much else, it was the 2019 crash that wiped out the corporations that had horded vast troves of data. Combined with freely available education and the investments made in infrastructure, this broke down many of the economic advantages of what we used to call developed countries. The 'internet of things' provided young, data-savvy entrepreneurs with rapid feedback, enabling products to become increasingly tailored to individuals. Automated production and the 'industrial internet' turned these dreams into physical reality. Entrepreneurship flourished across the world, as people used their ingenuity to identify and meet new sources of demand. Geography ceased to matter very much, and the world became far more meritocratic. Because people's starting points in terms of education and access to resources were increasingly equal, economic and social outcomes also became increasingly equal.

Some say that all this has come at a cost. Given the availability and free access to data, and the ubiquity of the 'internet of things', there is little in the way of privacy any more. And yes, we have central planning to maintain stability in the global economy and provide the infrastructure and basic services that we all rely upon. In some areas of life there has never been more choice, while in others choices are now made for us. Politics is the big one, but most people are happy to leave it to the technocrats nowadays, who need to regulate all our lives quite closely to hold all this together.

We had come to think that central planning couldn't work – it could never solve the 'information problem'. The problem was data. In the past, the market was often smarter than government because information was too dispersed, the data were too 'big', for any central planner to grasp, much less harness. But now we have data in unimaginably larger quantities, and computing power that seemed impossible in the 2010s. Our algorithms make their own new algorithms without us and constantly improve. Maybe you could say we are not far away from being governed by a benevolent machine but we have only to look around at the abundance of knowledge and resources at our disposal to remind us of the huge benefits of this.

Despite these, there are still people who don't think everyone should know everything about everything. Who feel that having your health, habits, and skills quantified and made public is an affront to dignity. For them, the costs to liberty and privacy are too great and they dream of getting out somehow. I have even heard tales of people 'going off the grid' and living completely different forms of lives, though I can't imagine that there is any truth in these.

Key drivers for this scenario

- **Privacy and digital identities:** Nearly all information ever produced is available and open to all. While this is often used to serve the public good, privacy is largely a thing of the past. There is opposition, but many are now used to it and cannot really remember a world where this was not the case.

- **Socioeconomic dynamics:** Economic inequality is greatly reduced. Most people are not rich, but hunger, malnutrition, illiteracy, infant mortality, and other indicators of poverty are low.
- **Data:** 'Datafication of everything' is a reality and machines are now learning to analyse it better.
- **Innovation:** The data and tools available to everyone have allowed for an explosion of innovation.
- **New economic models:** The neoliberal model has been discredited. Centralised control over key parts of the economy works in a way previously unimaginable due to the availability of data and tools with which to analyse it.
- **Resources:** Energy is abundant and distributed.
- **Political regimes:** Along with the economic model, liberal democracy and its institutions were also discredited in the crash, leaving people more comfortable with benevolent authoritarianism that could guarantee economic and environmental stability, ensure fairness and material wellbeing.

Scenario 2: knowledge isn't power

A world of freely available knowledge, but in a context where many important assets are scarce, and ownership of these assets is relatively concentrated.

For many people living in developing countries in 2030, some of the worst trends from the early 2000s continued. The promise of renewable energy never materialised, as technological advance was unable to improve the economics sufficiently to compete with fossil fuels. The shale gas boom and low oil price also proved short-lived, so that those that do have access to energy pay a high price for it. Global economic growth has been severely constrained ever since.

Following the continuing failure to achieve global agreement on greenhouse gas emissions, talk of climate change mitigation died away years ago. Changes to the climate have been more incremental than some thought, but significant warming is now a fact of life, and weather systems have become increasingly volatile and unpredictable. As was long predicted, these impacts are felt hardest in countries near the equator and tropics, where droughts have intensified, and tropical storms become more extreme.

Because of weak growth and high energy costs, living standards for many failed to improve and even deteriorated in some countries. As resources became increasingly constrained, countries and companies competed fiercely for them. More effort was expended in attempts to obtain access to energy supplies, for example, than in providing the goods and services that would improve people's lives. What cooperation for the global good had existed has largely disappeared.

But there have also been positive changes in the world. Most importantly, knowledge is now open and available to all who wish to access it. Not everyone is able to use these data to the same extent, however, as global corporations have the technology, algorithms and resources generally to make the best use of them. Also, having abundant information does not necessarily mean empowerment in a resource-scarce world. In many countries it has just fed feelings of resentment, as the poor and marginalised are all too aware of the inequalities that contribute to their situation.

While people are able to understand their health problems with an ease and accuracy that was unimaginable in the 2010s, this does not mean that they can afford to buy the drugs they need. An alternative to paying the prices of the global pharmaceutical companies is generic medicines which have become more widely available as access to scientific knowledge improved, though these are less reliable. Even these are unaffordable for many people though. In the absence of government support, people in poorer countries have turned to online crowdfunding to raise finance for many of the health services they need, and online petitions to lobby government to address social problems are now everywhere.

While the ability of people to identify needs and use new techniques to try and raise the finance to address them is a positive side of open knowledge, few of these efforts are successful. With so many people attempting to crowdfund at the same time, the resources available are a fraction of what is needed. Political tensions, therefore, are often high. As a result, the wealthy have shut themselves off behind security cordons. To maintain law and order the state is now engaged in extensive digital surveillance, gathering information on all aspects of people's lives. Drones are everywhere.

Privacy is affected in other ways. The sharing of medical records for crowdfunding means anyone can access them, reducing the ability of sick people to get health insurance or access to finance, and adversely affecting their job opportunities. Health data are traded between global pharmaceutical companies. More generally, the vast amount of personal data now available – often in real-time – is increasingly in the hands of a small number of corporations, who analyse these to provide goods and services that are perfectly tailored to individuals' needs, if they can afford them. People have no control over the data that exist about them.

Some cannot afford to participate in the system, while others remove themselves by choice. Some of these people have started to band together in larger groups. While a few use violence, most try to persuade the public and governments that extreme inequalities are damaging to all, not just to the marginalised. It remains to be seen if these people will be able to affect change, but the power of the vested interests they confront has never been greater.

Key drivers for this scenario

- **Privacy and digital identities:** Massive data collection and data matching coupled with open knowledge means that privacy and data security is easily compromised.
- **Knowledge creation:** Through the use of the social web (crowdsourcing) and other technologies that collect, store and analyse data, there are new forms of knowledge creation.
- **Culture:** Having open knowledge but a scarcity of resources has an impact on how people relate to each other and how they behave.
- **Information technology:** ICTs are accessible to most people and at affordable cost. The internet has essentially become the computer in this new era of cloud computing. Mobile devices that connect to the internet are affordable for most people in the developing world.
- **Innovation:** Advances in biotechnological pharmaceuticals coupled with systems in e-health make 'personalised' medicine the norm. Sensors are more powerful, cheaper and consume less power; they can also be embedded in anything due to miniaturisation. Research into computer algorithms, machine learning, and data processing produce smart applications. Robots become more intelligent and autonomous and play a critical role in helping humans. Drones are increasingly used in civilian settings.
- **Globalisation:** Globalisation works in favour of stakeholders that are already powerful, ranging from individuals, nation states and corporations.
- **Inequality:** Without active intervention inequality increases in a world of scarce resources.
- **Data/ICT:** The ownership of data and ICT infrastructure will play a key role in this scenario.

Scenario 3: ignorance is bliss?

A world of resource scarcity and concentrated ownership, but where access to knowledge is tightly managed and controlled by corporations and government.

In 2030, resources have become scarce. The price of everything, including information and knowledge, has increased. A small political and economic elite control these resources, including people's personal data, the media, and the means of producing and disseminating knowledge. Competition, between both countries and corporations, is largely a thing of the past. Those that have managed to retain or obtain control over resources have such an advantage that their positions are unassailable. Inequality between the general populations in different countries has continued to fall, while the gap between this majority and the global elite widens inexorably.

Energy is scarce, expensive and unreliable in developing countries. Public infrastructure is either non-existent or of poor quality, with those that can afford to using their own systems. Most people with jobs work for a branch of one of the global mega-corporations, which have monopolies in different sectors. In broadcasting, for example, the same firms control programmes, channels and the screens on which they are shown. Although viewers can no longer choose what to watch, the production values and scripts of the available offerings are generally good due to the mega-corporations employing almost all the creative talent.

Similar quality exists in the health sector, for those able to afford it. A patient's response to medication is monitored in real-time, with dosages and formulations varying to optimise results. Data are encrypted and sent to different parts of the world to feed into proprietary research trials aiming to produce lucrative 'life-extending' medicines for the elite. Public health systems continue to wither away, as the availability of data on health prospects reduces the incentive for people to pool risks and collectively insure themselves. As uncertainty over people's future health reduced, those with the resources tailor their expenditure to meet their own need in the private sector. Those without the resources have little option but to rely on an increasingly cash-strapped public system.

Millions of young workers across the globe transcribe discrete chunks of data for the global corporations, without ever knowing what it is they are working on. Geography is increasingly irrelevant, with most jobs going to the brightest and cheapest, wherever they live in the world. While intelligent and curious, few of these young people can afford to go to university. Online courses are an alternative, but they are also expensive, with content controlled by the monopoly education provider. Employers train their workers well to do particular tasks, but not more broadly.

Some disillusioned employees dream of releasing data and making things like medical advances available, but security is generally too tight to get away with this, and the resources needed to use the data are concentrated in the mega-corporations anyway. Not all hackers are altruistic though – some with the skills to break into closed systems do so for personal gain. Many don't blame them.

By staying away from the large urban centres, independent spirits can avoid seeing prescribed broadcast content, although they still have compulsory online profiles, created from population data and updated using their personal data. Some prefer to get information from underground networks, which are informal and uncontrolled. A few of these local networks have even developed into small self-governing enclaves, usually where somebody in the community has the – increasingly rare – knowledge to bypass the various forms of government and corporate surveillance. Many in these communities wonder why most people seem so accepting of the status quo, even happy with it.

Key drivers for this scenario

- **Natural and energy resources:** Elite capture of scarce resources.
- **Medical technology:** Has advanced to significantly prolong life for the richest. The rest of the population is kept moderately healthy and used to test new treatments.
- **Information technology:** Widely available and advanced but monitored and constrained by the elite.
- **Data management capacity:** Has expanded as the volume of data has increased, driven by the elite's need for social control through access to personal information.
- **Privacy and digital identities:** Social networking technologies are advanced with near universal participation, but used for surveillance by the corporate and government elite.
- **Social stratification:** Big gap between elites and everyone else. Enclosure and manipulation of knowledge, information and media keeps the 99 per cent ignorant and relatively content.

Scenario 4: privatised paradise

A world of abundant and widely distributed resources, where knowledge is largely privately controlled but by a range of different actors.

To a visitor arriving from 2014, the world would seem overflowing with wealth, riches and gadgets. The global system looks fairly stable, and everyone has a job and access to leisure and entertainment. You can't argue that much has changed for the better in a material sense. Following global agreements in 2015 between governments, big companies, and international NGOs, the Sustainable Development Goals (SDGs) have been met in many countries. 2015 also saw the historic agreement at the Paris Conference of the Parties (COP), where a global deal to control carbon emissions was struck, and a stable climate financing framework was agreed. Yes, emissions continued to rise, and the two-degree warming threshold was quietly abandoned, but emissions have been declining for a few years now. The introduction of the global carbon market was a key driver of change, with the major financial institutions intimately involved in its development and day-to-day management.

An era of 'green growth' saw new industries emerge, with many jobs created in the energy sector, or to improve the efficiency of older buildings and to make new buildings and transport systems sustainable. As always, some countries are further ahead than others, but most are making progress, helped by the flow of commercial climate finance to poorer countries where reducing carbon emissions is cheapest.

Many diseases have been controlled, and we have such good early warning systems with sensors in everything (including us) that any new risks can be quickly assessed and monitored. Nutrition has improved, as has life expectancy. Basic online education is now widely available, but is strongly geared to meeting the needs of the economy. Broader, high-quality education is also available, but it is very expensive and is increasingly tailored to individual abilities and needs. The private sector has proved extremely innovative at developing and maintaining many of these systems.

The UN Declaration on Data Privacy was agreed in 2016, giving individuals control over their personal data. Unequal access to technology and information meant that this was of most benefit to people in developed countries, with people in poorer countries sometimes selling their personal data to get access to the internet, or just for the income it could generate. This was not ideal of course, but a vibrant market for personal data soon sprang up in developed countries too, as younger people in particular saw it as just an extension of personal choice. The situation is becoming more equal now.

When corporations began exercising their rights to data privacy, efforts to obtain their data for public benefit were thwarted. This did prevent some of the more utopian benefits of ‘big data’ being realised, but they were probably never realistic. In contrast, the commercial potential of big data proved to be greater than even the optimists had hoped.

Not everything is better, of course. Most things that used to be free 20 years ago now have to be paid for. Societies seem more individualised, with less talk of things like ‘public goods’. Yes, poverty in all its forms is less noticeable and more people have some assets to build upon, but the countries that were rich in 2015 have got much richer. Life within poorer countries has got better for most people, but there too it is the relatively wealthy that have benefited most.

Democracy is now the norm pretty much everywhere, though what that means in practice varies a lot. Private corporations and wealthy individuals exert a big influence, of course, but there is nothing new there. And we have our share of dissidents, of course. The anti-capitalist movement never went away, but found it difficult to compete for attention with commercial interests.

Some former ‘techno-libertarians’ have become activists, teaming up with the hackers of old to find gaps in the security systems of corporations and government. These were supposed to be unbreakable, but this has been shown not to be the case. High profile data leaks have caused disquiet at the top of global corporations and also in government. There are rumours that some firms and governments are now copying the hacker’s technology, and using it to access the personal data banks of individuals. But everyone knows that the encryption of this is completely unbreakable, so that obviously can’t be true…

Key drivers for this scenario

- **Privacy and digital identity:** Access to data is privately controlled, by individuals, corporations and government. Data may be sold or traded, but whether this is a free ‘choice’ may depend on the circumstances of the seller.
- **Education:** This is privatised and expensive for those who want high-quality education. For others it is free, open access, but highly tailored and closely monitored and controlled, to ensure the workforce matches the needs of the economy.
- **Socioeconomic:** The society is still divided into the ‘haves’ and ‘have-nots’. The gap between them is wider, despite improvements of the living standards of all groups.
- **Data:** There is an abundance of data, but limited knowledge of how to analyse and use them. Those who have this are either among the elite, or part of opposition movements.
- **Technology:** Society is highly connected, but net neutrality has long since gone, despite the US ruling of 2015. There is widespread use of sensors and surveillance, the output of which is controlled and traded on a commercial basis.
- **Social media:** People are highly networked, with networks designed to respond to need based on market analysis. Alternative networks are also evident. Crowdsourcing has followed a similar path, and is automated in many respects.
- **Innovation:** This is often crowd-managed but driven by commercial interests. The most interesting social innovations are coming from alternative groups who are trying to bring about structural change.
- **Globalisation:** The economy is now more environmentally sustainable following the introduction of strong economic incentives. Large economies, financial institutions and international corporations have disproportionately benefited from these developments.

Points emerging from scenarios

Each of the four scenarios paints a vivid image of the future. These are deliberately extreme caricatures in some respects, but they also describe contexts which are internally coherent. An important point is that none are utopian or dystopian, though some may resemble each of these somewhat depending on your perspective. Each scenario combines features that many people would consider positive with others that they would consider negative. Whilst not as extreme as the stylised world portrayed here, it is highly likely that the future will combine 'good' and 'bad' in this way, that people will disagree on what these terms mean in practice, and that effects will be differentially felt.

These outcomes are not random, however. Another purpose of developing a narrative account of a future state of the world is that it helps to see how things that happen now – or do not happen, of course – can result in significant changes in the future. If we want to increase the likelihood that some things happen and others do not, it is necessary to understand these relationships.

Section 4 of this report explores policy options in the light of these considerations. As described at the outset, the purpose of this work has been to identify the most important ways that big data could impact upon developing countries, and to understand how positive impacts could be maximised and negative effects minimised. The question, therefore, is what policy can do to influence this?

Annex 4 Big data workshops, 2014: participants

- Kerry Albright (Senior Advisor, Data and Innovation, Department for International Development (DFID))
- Justin Baker (University of Texas)
- Craig Bardsley (Head, International Development, Economic and Social Research Council)
- Evangelina Berdou (Research Fellow, Institute of Development Studies (IDS))
- Nason Bimbe (Information Systems Manager, IDS)
- John Brownlee (Partnership and Capacity Development Senior Convenor, IDS)
- Graham Brown-Martin (EDlabs)
- Julia Chandler (Digital Strategy Implementation, DFID)
- Nicolas de Cordes (Vice President, Marketing Vision, Orange)
- Graham Eele (Royal Statistical Society)
- Emily Gillingham (Chair of the Executive Council, R4L)
- Paul Gooding (Research Fellow in Digital Humanities, University of East Anglia)
- Jon Gregson (Head of Knowledge Services, IDS)
- Jon Harle (Senior Programme Manager for the Research Access and Availability team, International Network for the Availability of Scientific Publications (INASP))
- Gus Hosein (Executive Director, Privacy International)
- Caroline Hotham (Fund Manager, Start Network)
- Neil Jackson (Chief Statistician, DFID)
- Theo Keane (Senior Programme Manager, Innovation Skills, Nesta)
- Chris Keene (Library Technical Development Manager, University of Sussex)
- Richard King (Policy Research Adviser, Oxfam)
- Robert Kirkpatrick (Director, UN Global Pulse)
- Jeff Knezovich (Policy Influence and Research Uptake Manager, IDS)
- Melissa Leach (Director, IDS)
- Laura Mann (Postdoctoral Researcher, Leiden African Studies Centre)
- Jenny Molloy (Coordinator, Open Knowledge Foundation)
- Isabel de la Penna (IDS)
- Rachel Playforth (Repository Coordinator, British Library for Development Studies, IDS)
- Emma Prest (General Manager, DataKind UK)
- Alun Rhydderch (Horizon Scanner)
- Emma Samman (Research Fellow, Overseas Development Institute)
- Stephen Spratt (Research Fellow, IDS)
- Naaman Tammuz (Data Scientist, Duedil)
- Steve Tovell (Programme Coordinator, IDS)
- Katherine Van Wyk (Business Development, Every1Mobile)

References

Aiden, E. and Michel, J-B. (2013) *Uncharted: Big Data as a Lens on Human Culture*, New York: Riverhead Books

BBC News (2014) *IBM Seeks App Developers to Harness Watson*, www.bbc.com/news/technology-26366888 (accessed 28 February 2014)

Beine, M.; Docquier, F. and Rapoport, H. (2008) 'Brain Drain and Human Capital Formation in Developing Countries: Winners and Losers', *The Economic Journal* 118.528: 631–52

Bertolucci, J. (2013) 'Big Data's New Buzzword: Datafication', *Information Week*, www.informationweek.com/big-data/big-data-analytics/big-datas-new-buzzword-datafication/d/d-id/1108797 (accessed 3 March 2014)

Beyer, M.A. and Laney, D. (2012) *The Importance of 'Big Data': A Definition*, www.gartner.com/doc/2057415/importance-big-data-definition (accessed 16 June 2015)

Bizer, C.; Heath, T. and Berners-Lee, T. (2009) 'Linked Data – the Story so Far', *International Journal on Semantic Web and Information Systems* 5.3: 1–22

Blanchet, K.D. (2008) 'Remote Patient Monitoring', *Telemedicine and e-Health* 14.2: 127–30, http://online.liebertpub.com/doi/abs/10.1089/tmj.2008.9989 (accessed 16 June 2015)

Boyd, D. and Crawford, K. (2012) 'Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon', *Information, Communication & Society* 15.5: 662–79

Brock, D.C. (ed.) (2006) *Understanding Moore's Law: Four Decades of Innovation*, Philadelphia PA: Chemical Heritage Foundation

Brown, B.; Chui, M. and Manyika, J. (2011) 'Are You Ready for the Era of "Big Data"?', *McKinsey Quarterly* 4: 24–35

Bryant, R.; Katz, R. and Lazowska, E. (2008) *Big-Data Computing: Creating Revolutionary Breakthroughs in Commerce, Science and Society,* www.cra.org/ccc/files/docs/init/Big_Data.pdf (accessed 9 July 2015)

Carr, N. (2014) 'The Limits of Social Engineering', *MIT Technology Review*, 16 April, www.technologyreview.com/review/526561/the-limits-of-social-engineering (accessed 30 April 2014)

Cerra, A.; Easterwood, K. and Power, J. (2013) *Transforming Business: Big Data, Mobility, and Globalization*, Indianapolis IN: Wiley

Charles, D. (2014) *Should Farmers Give John Deere And Monsanto Their Data?*, www.npr.org/blogs/thesalt/2014/01/21/264577744/should-farmers-give-john-deere-and-monsanto-their-data (accessed 23 March 2014)

Chen, M.; Mao, S. and Liu, Y. (2014) 'Big Data: A Survey', *Mobile Networks and Applications* 19: 171–209, http://mmlab.snu.ac.kr/~mchen/min_paper/BigDataSurvey2014.pdf (accessed 16 June 2015)

Christensen, C.M. (1997) *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail*, Boston MA: Harvard Business Review Press

CNN Money (2014) *A Timeline of Computing Power*, money.cnn.com/interactive/technology/computing-power-timeline (accessed 28 July 2014)

Commonwealth of Australia (2013) *The Australian Public Service Big Data Strategy*, www.finance.gov.au/sites/default/files/Big-Data-Strategy_0.pdf (accessed 16 June 2015)

Courtney, M. (2014) 'How Utilities Are Profiting from Big Data Analytics', *E&T Magazine*, 20 January, www.eandt.theiet.org/magazine/2014/01/data-on-demand.cfm (accessed 17 March 2014)

Craig, T. and Ludloff, M.E. (2011) *Privacy and Big Data*, Sebastopol CA: O'Reilly Media

Cukier, K. (2010) 'Data, Data Everywhere', *The Economist*, 25 February, www.economist.com/node/15557443 (accessed 16 June 2015)

Diebold, F.X. (2012) *A Personal Perspective on the Origin(s) and Development of 'Big Data': The Phenomenon, the Term, and the Discipline*, http://economics.sas.upenn.edu/~fdiebold/papers/paper112/Diebold_Big_Data.pdf (accessed 24 June 2015)

Easley, D.; Lopez de Prado, M. and O'Hara, M. (2012) 'The Volume Clock: Insights into the High Frequency Paradigm', *The Journal of Portfolio Management* 39.1: 19–29, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2034858&download=yes (accessed 16 June 2015)

Fast Company (2014) *The World's Top 10 Most Innovative Companies in Big Data,* www.fastcompany.com/most-innovative-companies/2014/industry/big-data (accessed 19 March 2014)

Feinleib, D. (2013) *Big Data Demystified: How Big Data Is Changing the Way We Live, Love and Learn*, San Francisco: The Big Data Group

Foreign Policy (2014) *Why Big Data Missed the Early Warning Signs of Ebola,* www.foreignpolicy.com/2014/09/26/why-big-data-missed-the-early-warning-signs-of-ebola/ (accessed 26 September 2014)

Foreman, J.W. (2013) *Data Smart: Using Data Science to Transform Information into Insight*, Indianapolis IN: John Wiley & Sons, Inc.

Franks, B. (2012) *Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics*, Hoboken NJ: John Wiley & Sons, Inc.

Freedom House (2013) *Ethiopia*, www.freedomhouse.org/report/freedom-net/2013/ethiopia#.UzeJ5K1dVic (accessed 30 March 2014)

Freeman, C. and Louçã, F. (2001) *As Time Goes By: From the Industrial Revolutions to the Information Revolution*, Oxford: Oxford University Press

Frey, C.B. and Osbourne, M.A. (2013) *The Future of Employment: How Susceptible are Jobs to Computerisation?*, OMS Working Paper, Oxford: Oxford Martin School

Gantz, J. and Reinsel, D. (2011) ‘Extracting Value from Chaos’, *IDC iView* 1–12

Gilder, G. (1997) ‘Fiber Keeps its Promise’, *Forbes ASAP*, 8 April, 91–94

Gmür, M. and Schwab, K. (2014) *Education and Skills 2.0: New Targets and Innovative Approaches*, Geneva: World Economic Forum

Government Office for Science (2013) *Foresight Projects*, www.gov.uk/government/collections/foresight-projects (accessed 16 June 2015)

Hameed, T. (2007) *ICT as an Enabler of Socio-Economic Development*, www.itu.int/osg/spu/digitalbridges/materials/hameed-paper.pdf (accessed 9 July 2015)

Hampton, S.E.; Strasser, C.A.; Tewksbury, J.J.; Gram, W.K.; Budden, A.E.; Batcheller, A.L.; Duke, C.S. and Porter, J.H. (2013) ‘Big Data and the Future of Ecology’, *Frontiers in Ecology and the Environment* 11.3: 156–62

Harford, T. (2014) ‘Big Data: Are We Making a Big Mistake?’, *Financial Times Magazine*, 28 March

Hay, S.I.; George, D.B.; Moyes, C.L. and Brownstein, J.S. (2013) ‘Big Data Opportunities for Global Infectious Disease Surveillance’, *PLoS Med* 10.4, http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001413 (accessed 16 June 2015)

HM Government (2012) *Open Data White Paper: Unleashing the Potential*, The Stationary Office, www.data.gov.uk/sites/default/files/Open_data_White_Paper.pdf (accessed 16 June 2015)

Howard, M. (2001) *e-Government Across the Globe: How Will ‘e’ Change Government?,* www.academia.edu/2259802/e-Government_Across_the_Globe_How_Will_e_Change_Government (accessed 16 June 2015)

Hurwitz, J.; Nugent, A.; Halper, F. and Kaufman, M. (2013) *Big Data For Dummies*, Hoboken NJ: John Wiley & Sons, Inc.

IBM (2014) *Big Data Industry Energy & Utilities*, www-01.ibm.com/software/data/bigdata/industry-energy.html (accessed 23 March 2014)

IBM (2013a) *Big Data and Analytics in Travel and Transportation*, www-01.ibm.com/software/data/bigdata/industry-travel.html (accessed 16 June 2015)

IBM (2013b) *What Is a Data Scientist? – Bringing Big Data to the Enterprise*, www-01.ibm.com/software/data/infosphere/data-scientist (accessed 16 June 2015)

John Deere (2014) *John Deere FarmSight™*, www.deere.com/wps/dcom/en_US/campaigns/ag_turf/farmsight/farmsight.page (accessed 23 March 2014)

Kellmereit, D. and Obodovski, D. (2013) *The Silent Intelligence – The Internet of Things*, San Francisco: DND Ventures LLC

Kolb, J. and Kolb, J. (2013) *The Big Data Revolution*, Applied Data Labs Inc.

Laney, D. (2001) *3D Data Management: Controlling Data Volume, Velocity, and Variety*, Stamford CT: META Group

Letouzé, E. (2012) *Big Data for Development: Challenges & Opportunities*, New York: United Nations Global Pulse, www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulseJune2012.pdf (accessed 16 June 2015)

Lohr, S. (2012) 'How Big Data Became So Big', *The New York Times*, Unboxed, 11 August, www.nytimes.com/2012/08/12/business/how-big-data-became-so-big-unboxed.html?_r=0 (accessed 2 March 2014)

Lycett, M. (2013) '"Datafication": Making Sense of (Big) Data in a Complex World', *European Journal of Information Systems* 22.4: 381–86

Manyika, J.; Chui, M.; Brown, B.; Bughin, J.; Dobbs, R.; Roxburgh, C. and Byers, A.H. (2011) *Big Data: The Next Frontier for Innovation, Competition, and Productivity*, McKinsey Global Institute, www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_innovation (accessed 16 June 2015)

Matinde, V. (2013) 'Will Rwanda Overtake Kenya as East Africa's ICT Hub?', *IT Web Africa*, www.itwebafrica.com/ict-and-governance/389-rwanda/231618-will-rwanda-overtake-kenya-as-east-africas-ict-hub (accessed 29 July 2014)

Mayer-Schönberger, V. and Cukier, K. (2013) *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, New York: Eamon Dolan/Houghton Mifflin Harcourt

Mazmanian, A. (2014) *The Mosaic Effect and Big Data*, www.fcw.com/articles/2014/05/13/fose-mosaic.aspx (accessed 16 June 2015)

Mechael, P. (2009) 'The Case for mHealth in Developing Countries', *Innovations: Technology, Governance, Globalization* 4.1: 103–18, www.researchgate.net/publication/227627505_The_Case_for_mHealth_in_Developing_Countries (accessed 16 June 2015)

Metcalfe, R. (1995) 'From the Ether; Metcalfe's Law: A Network Becomes More Valuable as It Reaches More Users'*, Info World Magazine*, www.Infoworld.com (accessed 23 April 2014)

Murdoch, T.B. and Detsky, A.S. (2013) 'The Inevitable Application of Big Data to Health Care', *The Journal of the American Medical Association* 309.13: 1351–52

New Executive Office of the President (2012) *Obama Administration Unveils 'Big Data' Initiative: Announces $200 Million in New R&D Investments*, Office of Science and Technology Policy

Ohm, P. (2013) 'The Underwhelming Benefits of Big Data', *University of Pennsylvania Law Review Online* 161:339–46, www.pennlawreview.com/online/161-U-Pa-L-Rev-Online-339.pdf (accessed 16 June 2015)

Organisation for Economic Co-operation and Development (OECD) (2013) *Exploring Data-Driven Innovation as a New Source of Growth*, OECD Digital Economy Paper 222, www.oecd-ilibrary.org/science-and-technology/exploring-data-driven-innovation-as-a-new-source-of-growth_5k47zw3fcp43-en (accessed 16 June 2015)

Osterwalder, A. (2002) *ICT in Developing Countries*, Lausanne, Switzerland: University of Lausanne, pp.1–13

Oxford Dictionaries (2015) www.oxforddictionaries.com/us/definition/english/data (accessed 16 June 2015)

Pasquale, F. (2015) *The Black Box Society: The Secret Algorithms That Control Money and Information*, Cambridge MA: Harvard University Press

Pasquale, F. (2014) 'The Dark Market for Personal Data', *The New York Times,* 16 October, www.nytimes.com/2014/10/17/opinion/the-dark-market-for-personal-data.html (accessed 23 June 2015)

Saatchi, S.S.; Harris, N.L.; Brown, S.; Lefsky, M.; Mitchard, E.T.A.; Salas, W.; Zutta, B.R.; Buermann, W.; Lewis, S.L.; Hagen, S.; Petrova, S.; White, L.; Silman, M. and Morel, A. (2011) 'Benchmark Map of Forest Carbon Stocks in Tropical Regions Across Three Continents', *Proceedings of the National Academy of Sciences of the United States of America* 108.24: 9899–9904, www.research.ed.ac.uk/portal/en/publications/benchmark-map-of-forest-carbon-stocks-in-tropical-regions-across-three-continents(cc9e0a0e-dde5-414d-89cb-b525260f087a)/export.html (accessed 16 June 2015)

Sanders, R. (2014) 'Scientists Enlist Big Data to Guide Conservations Efforts', *Berkley University of California*, http://news.berkeley.edu/2014/07/18/big-data-guides-conservation-efforts (accessed 16 June 2015)

Saxberg, B. (2013) 'A River of Data', *Education Next* 13.1, www.educationnext.org/a-river-of-data (accessed 22 March 2014)

Schmidt, E. and Cohen, J. (2013) *The New Digital Age: Reshaping the Future of People, Nations and Business*, New York: Knopf

Shirky, C. (2008) *Here Comes Everybody: The Power of Organizing Without Organizations*, London: Penguin Books Ltd.

Siemens, G. and Long, P. (2011) 'Penetrating the Fog: Analytics in Learning and Education', *Educause Review* 46.5: 30–32

SmartData Collective (2013) *Is This the Ultimate Definition of 'Big Data'?*, www.smartdatacollective.com/bernardmarr/128486/big-data-ultimate-definition (accessed 19 February 2014)

Smith, S.V. (2013) 'Data is the Economy's New Oil', *Marketplace Tech*, www.marketplace.org/topics/tech/data-economys-new-oil (accessed 25 March 2014)

Smolan, R. and Erwitt, J. (2012) *The Human Face of Big Data*, Sausalito, California: Against All Odds Productions

Statista (2014) *Facebook: Figures of Monthly Active Users 2008–2013*, www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide (accessed 31 March 2014)

Taylor, L. (2013a) *Big Data in the Developing World*, blog, 4 March, linnettaylor.wordpress.com/2013/03/04/big-data-in-the-developing-world (accessed 21 March 2014)

Taylor, L. (2013b) *The Scramble for Africa's Data: Resource Grab or Development Opportunity?*, blog, 27 June, linnettaylor.wordpress.com/2013/06/27/the-scramble-for-africas-data-resource-grab-or-development-opportunity (accessed 23 March 2014)

The Big Data Landscape (2014) *Big Data and Relevant Companies*, www.bigdatalandscape.com (accessed 25 February 2014)

The Climate Corporation (2014) *Climate Basic*, www.climate.com/products/climate-basic/ (accessed 23 March 2014)

Townsend, A.M. (2013a) *Smart Cities: Big Data, Civic Hackers, and the Quest for a New Utopia*, New York: W.W. Norton & Company

Townsend, A. (2013b) 'Smart Government and Big, Open Data: The Trickle-Up Effect', *Institute for the Future*, www.iftf.org/future-now/article-detail/smart-government-and-big-open-data-the-trickle-up-effect (accessed 22 March 2014)

Turner, D.; Schroeck, M. and Shockley, R. (2013) *Analytics: The Real-World Use of Big Data in Financial Services*, Somers NY: IBM Corporation, www-935.ibm.com/services/multimedia/Analytics_The_real_world_use_of_big_data_in_Financial_services_Mai_2013.pdf (accessed 16 June 2015)

Twitter (2014) about.twitter.com/company (accessed 31 March 2014)

United Nations (2014) *The Universal Declaration of Human Rights*, www.un.org/en/documents/udhr/#atop (accessed 28 July 2014)

United Nations Development Programme (UNDP) (2014) *Human Development Index*, www.hdr.undp.org/en/content/human-development-index-hdi (accessed 28 July 2014)

United Nations Global Pulse (2013) *Big Data for Development: A Primer*, UN Global Pulse

Unwin, T. (ed.) (2009) *ICT4D: Information and Communication Technology for Development*, Cambridge: Cambridge University Press

Venture Beat (2013) *IBM Uses Big Data to Predict Outbreaks of Dengue Fever and Malaria*, www.venturebeat.com/2013/09/29/ibm-uses-big-data-to-predict-outbreaks-of-dengue-fever-and-malaria (accessed 2 April 2014)

Walker, J. (2012) 'Meet the New Boss: Big Data', *The Wall Street Journal*, 20 September, http://online.wsj.com/news/articles/SB10000872396390443890304578006252019616768 (accessed 16 June 2015)

Wang, S. (2013) 'Opportunities and Challenges of Clinical Research in the Big-Data Era: From RCT to BCT', *Journal of Thoracic Disease* 5.6: 721–23

Ward, J.S. and Barker, A. (2013) *Undefined By Data: A Survey of Big Data Definitions*, www.arxiv.org/abs/1309.5821 (accessed 16 June 2015)

White, R.W.; Tatonetti, N.P.; Shah, N.H.; Altman, R.B. and Horvitz, E. (2013) 'Web-scale Pharmacovigilance: Listening to Signals from the Crowd', *Journal of the American Medical Informatics Association* 20.3: 404–8, www.ncbi.nlm.nih.gov/pubmed/23467469 (accessed 28 July 2014)

WhoIsHostingThis? (2014) *Internet Censorship Around the Globe*, www.whoishostingthis.com/blog/2014/02/20/internet-censorship (accessed 28 July 2014)

Widmer, R.; Oswald-Krapf, H.; Sinha-Khetriwal, D.; Schnellmann, M. and Böni, H. (2005) 'Global Perspectives on e-Waste', *Environmental Impact Assessment Review* 25.5: 436–58

Worth, D. (2013) 'HP Big Data Tools Help Wildlife Charity Save the Planet', *V3 News*, www.v3.co.uk/v3-uk/news/2318103/hp-big-data-tools-help-wildlife-charity-save-the-planet (accessed 16 June 2015)

Yonazi, E.; Kelly, T.; Halewood, N. and Blackman, C. (eds) (2012) *The Transformational Use of Information and Communication Technologies in Africa*, Washington DC: World Bank

Institute of
Development Studies

Brighton BN1 9RE

T +44 (0)1273 606261
F +44 (0)1273 621202
E ids@ids.ac.uk
www.ids.ac.uk