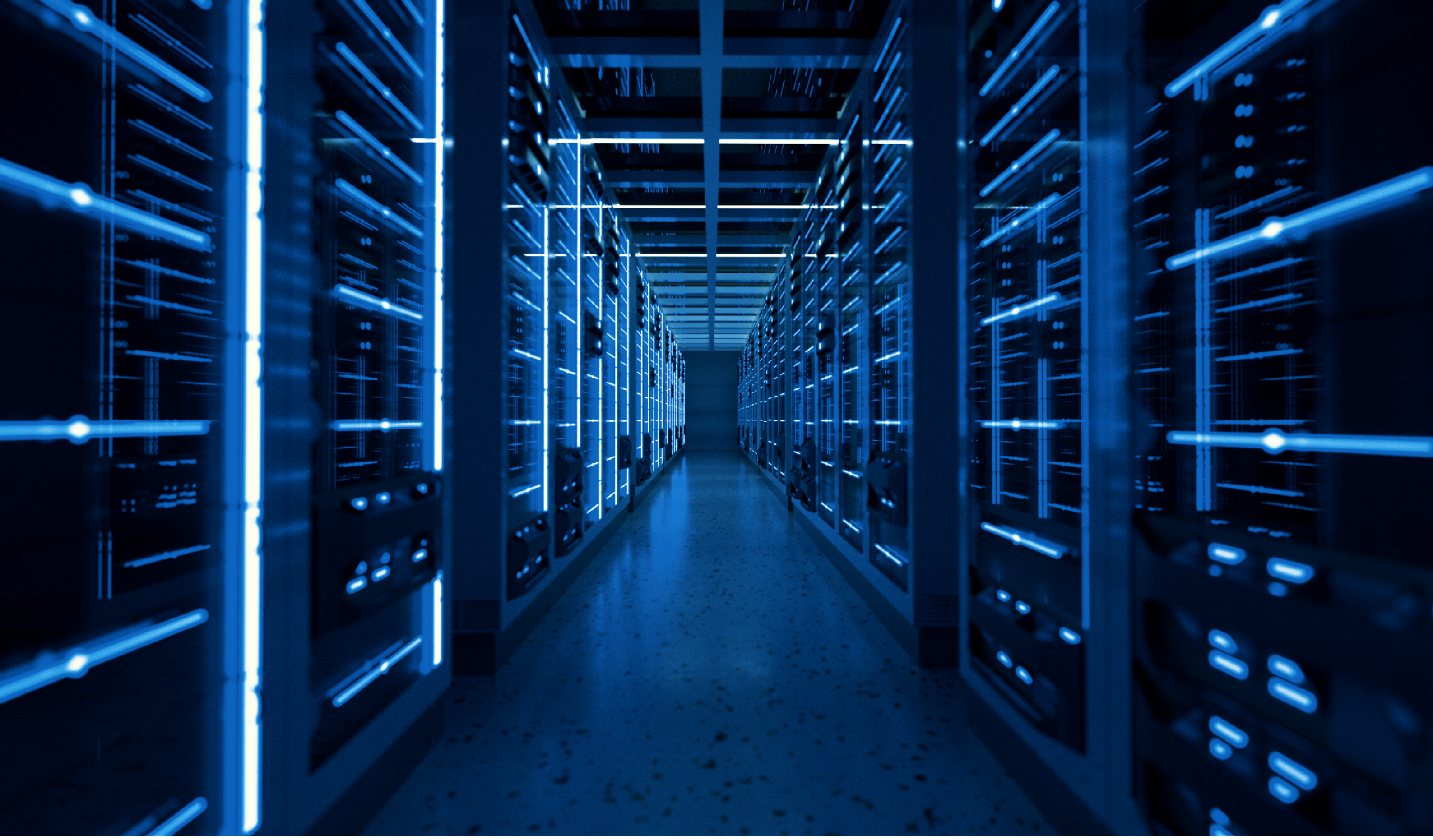


The Rise of the Data Economy and Policy Strategies for Digital Development

Digital Pathways Paper Series

Shamel Azmeh, Christopher Foster *and* Ahmad Abd Rabuh

BLAVATNIK SCHOOL OF GOVERNMENT
UNIVERSITY OF OXFORD

Digital Pathways at Oxford

Shamel Azmeh is a lecturer at the Global Development Institute (GDI) at the University of Manchester. *Christopher Foster* is a presidential research fellow at the Global Development Institute (GDI) at the University of Manchester. *Ahmad Abd Rabuh* is a data scientist and a lecturer at the University of Portsmouth.

Paper 10
February 2021

Digital Pathways at Oxford is a research programme based at the Blavatnik School of Government, University of Oxford. It produces cutting-edge research across the fields of public policy, law, economics, computer science, and political science to support informed decision-making on the governance of digital technologies, with a focus on low- and middle-income countries.

This paper is part of a series of papers on technology policy and regulation, bringing together evidence, ideas and novel research on the strengths and weaknesses of emerging practice in developing nations. The views and positions expressed in this paper are those of the author and do not represent the University of Oxford.

Citation:
Azmeh, S., Foster, C. & Abd Rabuh, A. (2021). *The Rise of the Data Economy and Policy Strategies for Digital Development.* Digital Pathways at Oxford Paper Series; no. 10. Oxford, United Kingdom

https://www.bsg.ox.ac.uk/research/research-programmes/digital-pathways

@DigiPathOxf

BLAVATNIK SCHOOL OF GOVERNMENT
UNIVERSITY OF OXFORD

Digital Pathways at Oxford

Abstract

The rapid advancement of digital technology is driving structural economic and technological changes in the world. The result of these changes can be seen in the rise of new modes of production, exchange, and consumption, the emergence of new economic actors, and the growing challenges to the policy landscape regulating the world economy. Data is central to these changes as the ability to collect, move, store, and analyse data is fundamental to new business models and actors. Debates around data and how to govern data have increased in recent years as states work to develop tools to develop their data economies and to limit the potential economic and social negative impact of new modes of production and trade. This paper uses the concept of data value chain to analyse the data economy and to examine the different policies states are following in different stages of the data value chain. We examine how these policies could translate into different pathways to achieve digital development by focusing on different stages within the data value chain. We identify four pathways to digital development: a) active data localisation, b) strategic data sharing, c) opportunities in low income data processes, and d) building sectoral specific application linked to data, and illustrate how different countries and economies could follow different policy pathways.

Table of Contents

1. Introduction

The rapid expansion of digital technologies in recent years is driving a major geographical and organisational restructuring of the global economy. As new technologies are developed, consumption habits are changing, production processes are being disrupted, and new types of products emerge. Reflecting those changes, new business actors are emerging with the potential to occupy central positions in global value chains and capture a substantial share of the value-added generated in those chains (Zuboff 2018, Foster et al. 2018). As economies and societies change as a result of those shifts, states are looking to new regulations and policies that aim to promote their economies and to tackle the economic, social, and political implications of these shifts (Azmeh et al, 2020).

Data and the data infrastructure have been central to these transformations. The growing infrastructure that collect, store, analyse, and use data has been a key driver of digital transformation (Mejias & Couldry 2019). As a result, debates around policy frameworks to deal with data have intensified (Singh & Vipra, 2019, Azmeh et al, 2020). At one end, global digital firms argue that policies that restrict the collection and exchange of data are harmful and lead to economic distortions and technological inefficiencies, and that liberal digital policies enable more economic benefits and drive innovation (Azmeh et. al. 2020). At the other end, more "data protectionist/interventionist" voices have emerged. Some of those voices focused on the implications of free flow of data on issues such as data protection and privacy (Mattoo & Metlzer 2018). Others have focused on the economic implications of a more liberal data regime, particularly the ability for dominant actors to control data and reap a major share of the economic benefits (Weber 2017).

In recent years, these debates around data have intensified in the context of developing and emerging economies (Weber 2017, Azmeh & Foster 2018b, Rossotto et al. 2018). The worry for many in developing countries is that the growing digitalisation of the economy will have serious implications, including the widening of the technological gap with the advanced economies and value-added being captured by global digital platforms. Many developing countries are already experiencing the impact of the digital transformation with regards to disruptions in their economies, with the growing role of digital platforms in areas such as e-commerce, media and advertising, amongst others. We have seen a diversity of reactions in different countries in responding to such challenges (Foster & Azmeh 2019, Azmeh & Foster 2018a, Azmeh & Foster 2018b). Some countries, especially smaller economies such as Estonia and Costa Rica, have followed a more liberal data approach, encouraging global digital firms to operate in their economies. Other larger economies, such as India, South Africa, and Indonesia, are looking towards "data protectionist" policies that aim to provide them with more control over data and to encourage the emergence of domestic digital firms.

Despite the growing debate around these issues, there is still a need for better understanding of data and the different types of roles, processes and actors involved in the data economy. Many data policies have been grouped under the label of "data localisation", but this underplays a diversity of policy approaches to data emerging in different contexts. This paper looks to unpack these issues of data and data policy with respect to developing countries, undertaking a more structured economic and technological analysis of those trends and these policy options. We aim to deliver such analysis by linking policy options to a stronger understanding of the data economy and by employing the notion of the "data value chain" (Li et al. 2019). By analysing the different stages in the data value chain, we aim to provide a better insight on processes, actors and relationships around data. Subsequently, we analyse the range of policies through this framework. An important outcome of this work is that given the range of different activities and specialisation in data, states are looking to different policy strategies through the value chain to develop their digital economies. Different developing economies are likely better suited to follow some pathways than others (Pathways for Prosperity Commission 2018).

The rest of this paper is structured as follows. Section two will dig deeper into the economics of data and the data value chain in order to provide a framework of the analysis. Section three examines the range of policies that affect the data value chain and each stage in this chain. Section four analyses how states are operationalising these policies shaping different digital pathways. Section five provides conclusions and policy implications.[1]

[1] In addition to the work of the authors in this area, this report is also based on conducting eight semi-structured interviews with firms and experts in different data-driven companies.

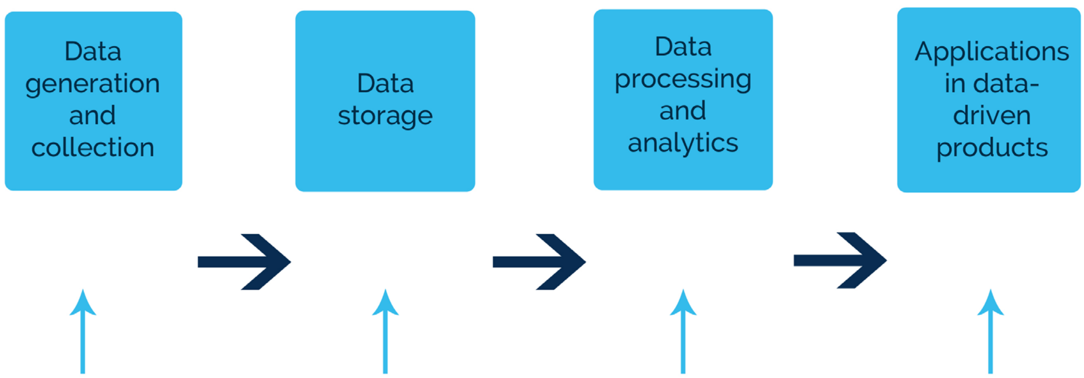
2. Data Value Chain and the Economic Value of Data in the Digital Economy

Different measures have been used to illustrate the growing importance of data. An underlying driver of data is the number of internet users, which has continued to grow rapidly in recent years (Cisco Annual Internet Report, 2018–2023). Internet users also generate growing amounts of data as they engage in more diverse and data-rich activities online. It is difficult to measure aggregate data flows, but estimates substantiate rapid expansions in data. According to market research firm, International Data Corporation (IDC), global volume of data increased from 20 zettabytes[2] in 2010 to 33 zettabytes in 2018 and is expected to grow to 175 zettabytes by 2025 (Reinsel et al. 2018).

These numbers, while impressive, are not very useful. In order to understand the economic value of data, we need a deeper analysis of when and how data becomes economically useful and how is value-added generated and captured. Data-intensive technologies are at the heart of these processes. Data-intensive technologies can be defined as technologies that are capable of processing high volumes of data at very high speeds. These processes are supported by data models, infrastructure services and tools that allow obtaining (and subsequently processing of) data from a variety of sources and delivering data-driven products and services to users (Demchenko et all 2014).

Building on Li et al. 2019, we can adapt such outlines of data transformation and processes into a simplified "data value chain" (figure 1). The activities in the data value chain enable the transformation of data from trillions of contextual transactions and interactions into economically useful tools and applications. It is important to highlight that this is a simplified version of the data economy. Often, the separation of the different stages and the sequence of activities is not as straightforward as suggested. Similarly, while we can think of application in data-driven products as the final stage, it is also important in shaping the entire process including all previous stages -product design affect what data and what data processing are needed (as indicated by the dotted line in the Figure). Notwithstanding these caveats, this framework is useful analytically to think of the distinct stages through which data is generated/collected, stored, analysed, and used to feed into decision making and a range of products and services.

Figure 1: Data Value Chain



[2] A zettabyte is a trillion Gigabytes

Using the data value chain concept suggests a variety of potential sources of economic value: controlling the raw data; having the infrastructure to store this data; having the human and computing capabilities to process and analyse this data; and having the ability to develop data-intensive goods and services. Moreover, given the use of data across a broad array of sectors, understanding how the data value chain translates into capturing value in specific economic sectors is crucial.

To understand the challenges that nations (and firms) might face in entering and evolving within the data value chain, it is useful to analyse the barriers to entry to stages of the data value chain. Data generation involves data collected from human activity or the environment, with the collection of this data largely done by a range of automated tools and devices. At this stage, factors such as the network effect and the high switching costs between different digital platforms often create high barriers to entry (Zuboff 2018, De Hert et al. 2018). Barriers to entry can also be significant in the data storage stage of the chain, especially due to the huge capital investments required in data centres and infrastructure especially with the growing scale and complexity of data centers (IDA 2018). The data processing and analytics stage requires a high degree of human capabilities and highly skilled labour in order to develop extractive and predictive models from the data. Similarly, the application stage in producing data-driven goods and services require product developers with knowledge in business-to-business (B2B) and business-to-consumers (B2C) trends and usage.

The data value chain also highlights the complex geographies of data. Certain activities in the data value chain can theoretically be performed anywhere in the world, provided workers are connected to the Internet. Nevertheless, the value chain remains grounded geographically in locations of data collection, sites for data centers, cables, key locations with skilled data scientists and engineers, etc.

In sum, the data value chain provides a basis for this study by allowing a more systematic analysis of the specific issues around skills, capacity, capital and geography of each stage. This can then allow a clearer understanding of the potential, and policy directions in developing countries. To begin this analysis, in the next sections we move to examine each of the stages of the data value chain in more detail.

a) Data Generation and Collection

Data generation and collection is the first step in the data value chain. Data is generated in many aspects of daily life such as movement and transport, health, consumption, production, amongst others. Data is collected from multiple sources including individuals and organisations, public and private. Digitalisation is enabling a rapid expansion of data capture in real time, and with increasing granularity and accuracy. With the expansion of data storage, this data is typically archived and can be integrated into datasets.

Data collection is also growing as a result of a growing list of "traditional" goods being equipped with data collection capabilities, such as appliances and cars. Data devices are increasingly interconnected with data collected and transmitted through M2M (machine to machine) communication or Internet of Things (IoT) networks (Tan & Wang 2010). Overall, data collection is expected to continue to grow rapidly, especially for the latter M2M and IoT areas (Cisco Annual Internet Report, 2018–2023). Throughout the economy, firms operating in manufacturing, services, and agriculture, are working to capture growing amounts of data from their operations (Basso & Antle 2020, Pilloni 2018, Oliff & Liu 2017).

While data collection is undertaken by multiple economic actors, large digital firms such as Amazon, Google, Microsoft, Apple, and Facebook are key actors in this process due to their control of leading online platforms and dominance of consumer devices (Zuboff 2015). These companies compete in developing services and devices that enable them to collect data from individuals and organisations. For individuals, these companies compete to develop new devices to collect data such as mobile phones, tablets, smart assistants, smart watches, smart TVs, amongst others. From their origins with a consumer focus, large digital firms have also expanded into provision of tools for firms in multiple sectors that allow them to capture data (Wolfert et al. 2017).

Notwithstanding the central role of these digital giants, it is important to recognise that data is collected by a large number of actors beyond the leading digital firms. Firms such as internet service providers, telecommunication companies, app developers, financial institutions, media companies, amongst others, are very active in this arena. As other sectors are digitising, a range of more traditional economic actors are becoming significant data collectors. As self-driving technologies expand, for example, automotive producers need to become involved in the business of data collection in order to compete (Bloom et al. 2017).

b) Data Storage

As data is collected in a range of forms and locations, the storage and associated infrastructure that supports movement of data globally is an important part of the data value chain. A key trend in recent years has been the shift from storage-on-device into the use of cloud computing for data storage.

Historically, enterprise data has been stored in more ad-hoc ways, with each firm building their own data and storage infrastructures. However, as data volumes have grown, there has been a trend towards the use of large and centralised data storage and the use of cloud infrastructure (Low et al. 2011, Lian et al. 2014, Alshamaila et al. 2013). For consumer data, there is growing use of the public cloud within consumer application and services for the storage of data (IDC 2018).[3] The rise of the public cloud is driven by multiple factors. Users, individuals and enterprises, benefit from the higher flexibility, scalability and security offered by cloud storage (Talukder & Zimmerman

[3] The public cloud refers to infrastructure services offered by third parties to individuals and companies in contrast to the private crowd which is infrastructure owned directly by those individuals and companies. Cloud services offered by companies such as Amazon, Google, and Microsoft are examples of the public cloud.

2010). Users can access the cloud from any location which provides a high degree of flexibility. Expanding (and reducing) storage capacity through the cloud is also simple in comparison to investments in in-house storage capacities.

The use of centralised data stores, cloud resources and associated infrastructure is expected to accelerate over the coming years as additional data is generated and as both business and consumer applications and devices become ‘cloud first’, relying on online processing at the core of their operation. Forecasts from the International Data Corporation (IDC) suggest that the share of data stored in consumer devices has declined from 60% of data in 2010 to around 40% of data by 2018 and is expected to reach around 20% by 2025. Data stored in the cloud has, in the meantime, increased from less than 10% in 2010 to 20% in 2018 and is expected to reach almost 50% by 2025. Data stored in enterprise data centres remains an important part of the location map with about 30-40% of all data (IDC 2018).

The main providers of cloud services are capturing a growing share of global data storage and are rapidly expanding their infrastructure. While different segments include specialist actors, the global public cloud market is dominated by three leading firms, Amazon Web Services (AWS), Microsoft Azure, and Google Cloud, in addition to Alibaba which leads in the Chinese market. Cloud computing is a capital-intensive business with the need to invest billions in storage facilities, buildings, security, skilled staff, software, power, cooling systems, amongst other areas. As a result, cloud computing facilities tend to be concentrated in countries with advanced infrastructure, trained labour, and a high degree of political and regulatory stability. There are also technical requirements to store data close to the market in order to reduce costs of transit, risks of saturated networks and limit latency, the time it takes the data to be transferred through the infrastructure. This preference means that cloud computing companies make decisions based on a combination of variables, including proximity to market and backbone fibre connections. Such decisions are also linked to policy. As regulations around data are adopted in different countries, compliance with those regulations often require certain decisions in terms of storage locations (Doelitzscher et al. 2010). Decisions around data might also link to the notion of ‘redundancy’ to ensure that even if systems and locations have issues, there are backup options. Cloud firm may account for redundancy through selecting a variety of locations. For organisations, data may be stored in more than one location in order to maintain data resiliency, the ability to recover the data.

An illustration of these points is Microsoft Azure. Microsoft divides the world into sixty-one regions. The company defines a region as “a set of data centers deployed within a latency-defined perimeter and connected through a dedicated regional low-latency network.”[4] As shown below, these regions are grouped into geographies that include two regions or more and that have similar regulatory and technical requirements.

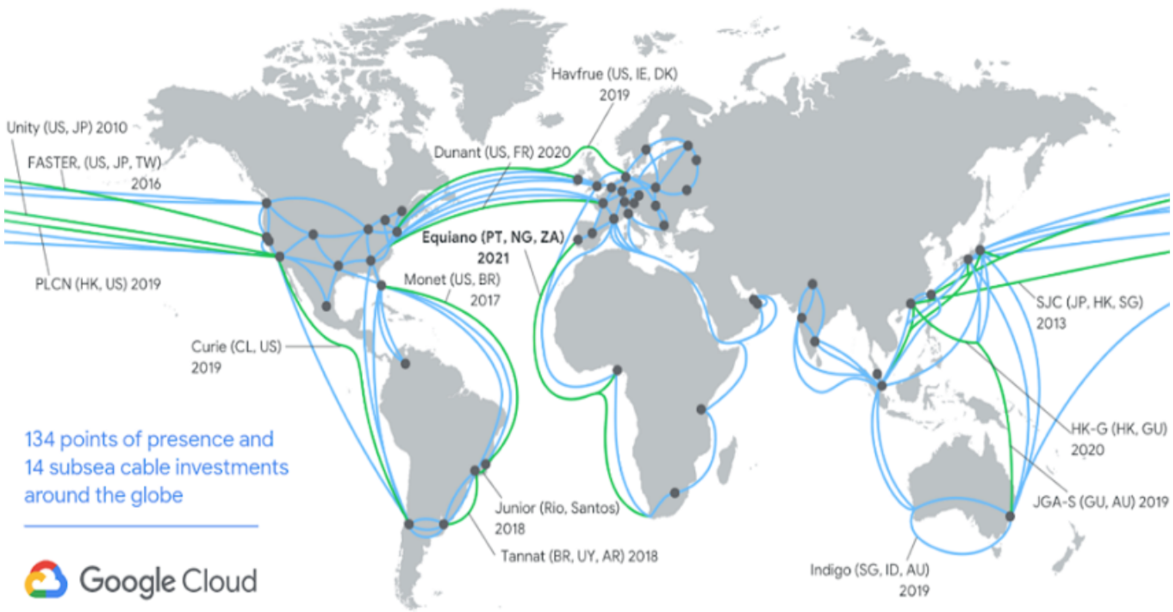
[4] Latency measures the time between the request of the user and the processing and response to that request by the server. In other words, it measures the round trip from a browser to a server.

Figure 2: The Location map of Microsoft Azure infrastructure

Source: Microsoft Azure

Cloud providers used to rely on leasing fibre optic links provided by dedicated infrastructure and telecommunication firms. But more recently as the dominance of the largest firms have grown, they are investing in cable networks either as part of a consortium with other companies including telecommunications companies or fully owned by them. Google has been the leading company in this arena investing US$ 47 billion between 2016 and 2018 in a network of underwater sea cables including Dunant which is a 6,600 kilometre cable that connects France and the United States and Curie, a 10,500 kilometre cable that connects California with Chile, and Equiano which runs along the West Coast of Africa from Portugal to South Africa with branching units along the way (figure 3). The aim of these investments is to improve the speed of data, increase data bandwidth, and reduce latency.

Figure 3: Google Cloud Undersea Cable Network



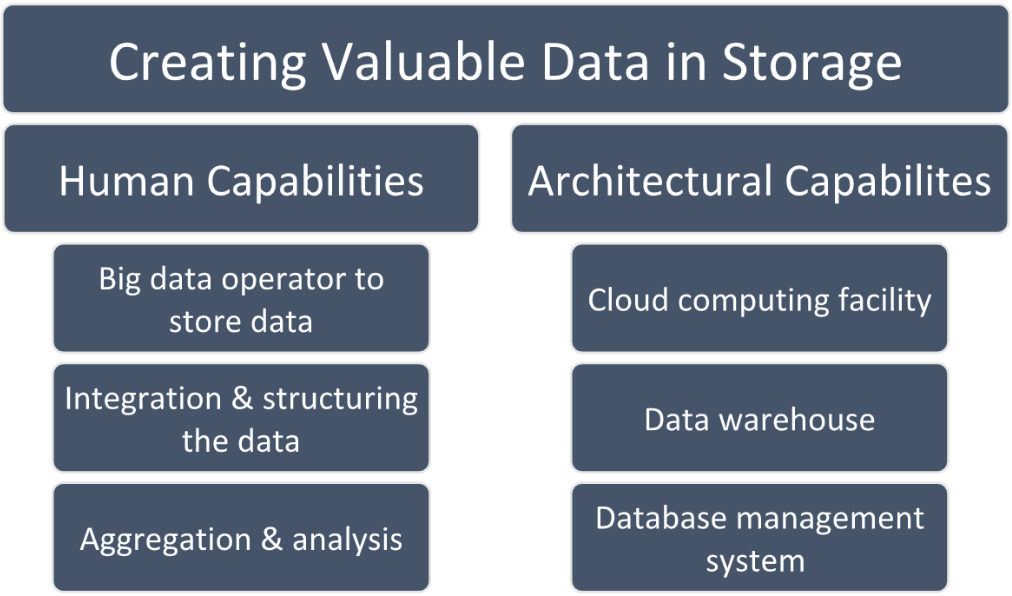
Source: Google Cloud

Data storage also involved a need for human capabilities (figure 4). A data warehouse is one of the most commonly-used repository architectures in today's big-data-enabled firms. Data analysts and professionals are those who are likely to have more direct interaction with raw data and are often involved in data cleaning, integration, refactoring data and online-analytical processing (OLAP) as part of managing data. Such activities are undertaken by database management professionals who are versed in understanding, and indeed designing, the structure of databases and can join different datasets using query languages such as structured query language (SQL). Low level data processes at scale also require significant network administration and interconnection management to ensure that low level operating systems and network services are running smoothly.

Many of these data tasks can be performed remotely. As a result, the physical storage of data in data centres does not necessarily create jobs within the location in which these data centers are established. It may result in a relatively small number of jobs for network engineers, technicians, administrative, and security jobs, but most data work will be done remotely.

However, even if most work is done remotely, the large capital investments needed for data centers and accompanying infrastructure can often create a "signalling effect" that drives investments in other areas of the digital economy and create economic spillovers (Andreosso-O'Callaghan et al. 2015, WRC 2013, IDA 2018). Such investments signal a number of positive factors with regards to issues such as business environment and incentives, political stability, land cost, and energy reliability. Moreover, large investment in data centres and associated infrastructure provides upgraded facilities that can be used by others. In Ireland, for example, it has been argued that initial investments in data centres have driven additional infrastructure investments leading to the expansion of nearby metropolitan areas as "digital clusters" (IDA 2018).

Figure 4: Data Warehousing Capabilities



Source: prepared by the authors

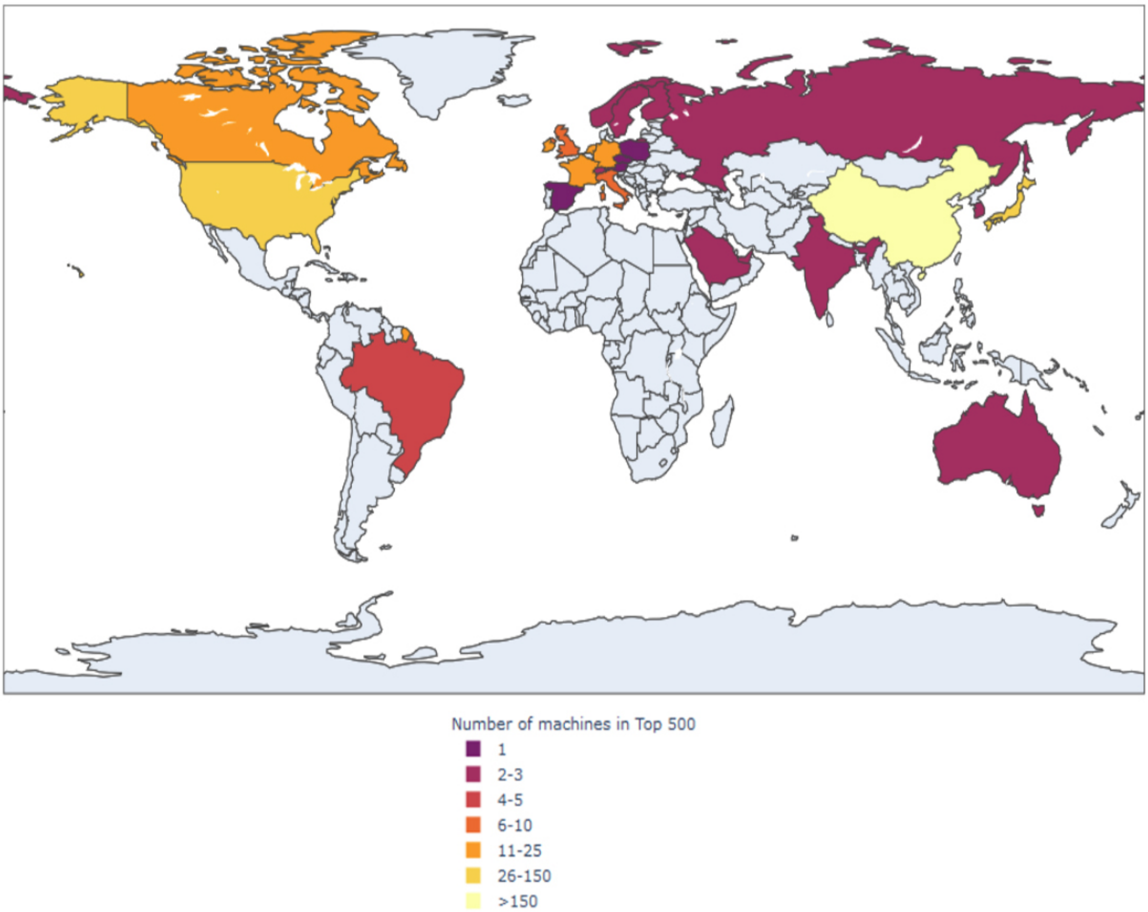
c) Data Processing and Analytics

Data analytics refers to a broad range of activities that are used to transform data into economically valuable products and insights (Turban et al. 2008, Raghupathi & Raghupathi 2014, Provost & Fawcett 2013). This transformation relies on the ability to combine and analyse large amounts of data to generate patterns and inform decisions. The ability to process and analyse large datasets has been used to identify new trends and, in some cases, paradigm shifts toward new goods, services or technologies (Abd Rabuh et al. 2016). As the economic value of data is dependent on the ability to integrate it with data collected from other sources, economic actors are likely to combine their own data with other external datasets, whether that be commercial or publicly available data. For example, a financial institution might combine its data on customer's spending habits and financial status with external data on lifestyle, family status, and other metrics to create personalised financial products (Han et al. 2011). Similarly, a clothing retailer can combine the data obtained from internet browsing with location data to send personalised offers to potential customers as they walk outside the store in which that product is available (Herhausen et al. 2015).

Data processing and analytics require both human and architectural capabilities. In terms of human capabilities, data processing and analytics skills are closely associated with university-educated data science and computer science professionals. Within this area, the range of skills that are needed include data mining & statistical analysis, database management & architecture, and machine learning and cognitive computing development, amongst others (Alpaydin 2020, Chen et al. 2012, Bakshi 2012, James et al. 2013, Hand & Adams 2014). The demand for such roles is increasing rapidly with economies competing to build capacities and attract skilled people in this area (Huang & Arnold 2020).

In addition to human capabilities, data analytics require advanced computer architecture capabilities. The growing amounts of data and the higher requirements for speed are creating a major challenge for computing (Asch 2018), particularly with the emergence of advanced analytics linked to machine learning. To deal with such requirements, key models such as deep learning and rendering that draw on large datasets require high performance computing, or supercomputing (Mercier et al. 2017). Supercomputers have the ability to conduct parallel processing of data through thousands of processors working at the same time performing calculations that are thousand times faster than normal computers. As such, major economic powers have been competing to develop such capabilities (Skordas 2019). In terms of geographies, high performance computing capabilities are almost exclusively located in the advanced economies with countries such as Japan, the United States, and China leading the world in such capabilities (figure five).

Figure 5: Top 500 supercomputers in the world by country, 20



Source: based on data from the Top500 project

d) Application in Data-Driven Products

The final stage in the data value chain is using the insights of the previous stages in developing data-driven products. As mentioned earlier, this stage influences previous stages as the anticipated product will shape the types of data to collect and what data processing are undertaken. Applications includes the development of new products that are only possible due to data-intensive technologies. It can also include the adaptation of existing goods and services offering additional features, to both consumers and business clients, that leverage the insights offered by data analytics to assist in better organisation and decision-making.

Box 1: The Data Value Chain in Autonomous Vehicles

Self-driving cars are an example of how the data value chain operates. In recent years, a number of automotive manufacturers and tech companies have been investing heavily in a race to develop this technology. The self-driving technology is based on the ability to collect, move, analyse, and apply huge volumes of data in order to “teach” the car how to behave in different circumstances. A car needs to be capable not only of understanding maps and the rules of the road but also to identify different objects and to react to their movements including the unpredictable behaviour of pedestrians and cyclists.

To be able to do so, the self-driving car need to be equipped with a sensor system that enable the car to see its environment. Such system includes sensors, radars, GPS systems, and computer vision systems. This system generates a staggering amount of data. Intel estimated the daily data generation of a self-driving car at 4000 gigabytes. The data generated from this system will be fed into the vehicle’s machine learning algorithm which interprets this data and makes a decision on the movement of the car. In addition, this data will be transferred to cloud infrastructure supporting the car, to be fed into training the machine learning algorithm to improve the system. The machine learning system needs to be trained in advance using huge amounts of data. This training process however does not end when a self-driving car is operating on the roads as the cycle of collecting data, feeding it into the machine learning system to improve the system is endless.

In this section, we offered a simplified version of the data value chain. As discussed in the outset of the section, the reality of the data value chain is more complex and the separation of these different stages is often difficult (as we illustrate in the example of autonomous vehicles in Box 1). However, this conceptualisation enables us to develop a better understanding of the data economy and to situate the different policies that nation states are adopting along this chain. We now move to focus on the issue of digital policy.

3. National Policies and the Data Value Chain[5]

The previous section has detailed the value chain of activities linked to the collection, storage, analysis and applications of data. Given this chain of diverse processes, data policy should therefore not be seen narrowly, but including a range of approaches unique to each step. The different processes of data transformation – collection, cleaning, storing, sharing, integration and analysis each involves different costs, work, skills and attention (Bates et al. 2016). The notion of "data friction" has been used to describe the resistances that data has to these transformations, and hence the work that needs to be done in each transformation. Data frictions are not constant and will vary according to the task, the form of data involved, size of data, types of transformation, complexity, organisations involved, amongst other factors. Crucially, policy will play an important role in determining data frictions: for instance, national data sharing frameworks may allow data to be shared more or less frictionlessly; but data restrictions or requirements increase the work required to transfer of transform data into useful insights (Bates 2017). At a macro-level, the extent of data friction across the data value chain will play a key role in determining which data transformations are viable and economic. This notion of data friction is important because it moves thinking of data policy beyond the binary of data blocking vs open data outlined in the introduction. This will therefore be a focus in this section, covering a broader range of policy relevant to data that can potentially shape data frictions making data transformations easier or more difficult.

From this perspective, there are a broad range of rules and strategies which shape the data value chain. Firstly, one can identify a range of different rules around data export, localisation, permissions (Kaplan & Rowshankish 2015). Nation states may also integrate rules which shape the governance, ownership and stewardship of national or sectoral data (Sampath 2018). Secondly, given the growing diversity of data flows, increasingly it is important to think about data policy "horizontally" and the way that policies are applied across specific categories of data, or focussing on specific sectors. Categorisations around personal, sensitive, non-personal and open data have been one way to differentiate data policies (Ciuriak & Ptashkina 2018, López González & Jouanjean 2017). In the following sections we examine such policies, with a focus on developing and emerging economies.

[5] This paper mainly focusses on policy that links to cross-border data flows and domestic data economies. Therefore, the scope is mainly on national policy agendas and how these rules link to economic outcomes. It is important, however, to highlight that this may underplay broader tensions around development agendas that will be important in data policy including considerations around ethics and justice, inclusion, and digital surveillance. Some of the sources for these discussions include Heeks & Renken 2018, Zuboff 2018, Arora 2016, Flyverbom et al. 2017, Mann 2018, Liu 2011, Drake et al. 2016, GCIG 2016, Thatcher et al. 2016, Foster & Azmeh 2020, Singh & Vipra 2019, IDC & OpenEvidence 2017.

3.1 Data policies and the value chain

The data value chain (Figure 1) is used to highlight the different policies that are relevant to data across the different steps of data transformation.[6] In order to provide a more systematic perspective, we also identify four important elements of data policy drawing upon previous analysis (Azmeh et al. 2020, Azmeh & Foster 2018a, Foster & Azmeh 2020). At its heart, data policy is likely to reflect the balance between domestic-orientated policy (that shapes domestic firms, secures domestic data flows and resources) and outwards-orientated policy (that shapes the conditions of foreign firms within a country and their data flows). In line with the ideas of data frictions, viable data activities and transformations will also closely depend on the underlying nature of technology and skills (that were discussed earlier) and the ability for developing countries to transfer skills and technologies in each stage of the data value chain. The extent of data activities in specific countries will also be shaped by the levels of local demand, and therefore considering "demand-side data policies" as an important policy aspect. Policies such as government demand for data, or other digitalisation initiatives can push increased use or more sophisticated demand of data locally and support the domestic data economy.

These four elements: domestic-orientated policy, outward-orientated policy, demand-side policies and the nature of transfer of skills and technology are explored for each of the four stages of the value chain in the sections below, with an overall summary of different of policies shown in Table 1. It should be emphasised that the goal is not to contend that all policies will be viable or successful, indeed many are in their very early stages and require more research. Rather, we look to build a richer knowledge of how data interacts with policy in developing countries which will be important to discuss specific national pathways that nation states are taking.

[6] Clearly many policies will be relevant to multiple or all steps in the value chain, but can particularly be important to specific steps

Table 1: Data policies and the data value chain

	Nature of skills and learning	Supporting local content & production	Shaping external actors	Demand-side
Collection	Skills: IoT, integration of data and machines Tech: data standards, dataset availability, government data	Clear regulation on data sharing Strategies to push sectoral data sharing from private or public sector	Non-personal data rules Expanding personal data protection Conditionalities associated with data protection rules	Opening up government, sharing of national data (e.g. health, government data)
Storage/ Infrastructure	Skills: Distributed systems, databases, network admin, cybersecurity Tech: internet infrastructure, networks databases	Support to integrate local firms in infrastructure	Data localisation rules	Encouraging businesses to localise given growing commercial advantages of local data
Analytics	Skills: Low end data management, Data science, AI/ML Tech: Data science libraries, AI, computer vision, voice	Supporting local labs, demonstrator firms or key projects with spillovers or links		Public procurement and integration of local analytics firms
Applications	Skills: App development, business integration, services Tech: Open app/client server platforms, open source, sectoral specific tools	Subsides or support for local firms for app development	Licencing for apps, blocking of applications, Physical presence rules of firms in China, Brazil Local content and partnership requirements on digital firms	Procurement of local firms Broader connectivity/ infrastructure that pushes demand and sophistication

Data Generation and Collection

The first stage of the data value chain is data generation and collection. In terms of skills and technologies, norms and technologies linked to data collection are becoming standardised in key sectors and this provides easier use and transformation (e.g. health data, goods import/export metadata). In terms of data availability, where data overlaps with public good there has also been strong push and support for digitalisation and sharing of key data. Moreover, we have also seen trends of global technology firms opening up their datasets to drive adoption, including a growing range of open labelled, cleaned and categorised datasets that can serve as the basis for machine learning algorithms and other processes (Perrier 2018).

Notwithstanding these positive actions, core processes of data collection often still require technologies and standards that remains proprietary (CNNum 2017). For example many forms of industrial machine and devices collect and communicate using vendor specific protocols (Liao et al. 2017). More broadly, the majority of data generated is owned by large tech firms who use it to shape their algorithms. Such data may include personal data and as such it may be difficult to share in terms of regulation.

Although in some cases, policy makers may negotiate data access for public good, for example tech firms as diverse as Uber, Waze, Airbnb and Orange have shared limited data to support urban planning and other social projects, this is often after long negotiation and under terms that would limit its use in the private sector (Villani 2018). Private sector firms may be able to gather such data through data brokers, but this is likely to be expensive and data structures are inconsistent. In sum, data is becoming more standardised, and is more collectable and available than before but this is up to a limit, particularly in accessing large relevant datasets that are used for advanced analytics.

Policies have looked to support the collection and sharing of data (including industry, societal and personal data) to expand availability. A focus on producing relevant local data is likely to support domestic data firms. For public sector data, initiatives around collection, standardisation, digitalisation and sharing of government data can link to the goal of supporting more data driven industries, in areas such as health and agriculture (Magalhaes et al. 2014). Governments are also putting in place systems to allow them to act as intermediary in terms of firms accessing more sensitive and personal data in areas such as national identity and financial sectors. This is most vividly shown in India where government systems associated with the Aadhaar identity act as an intermediary between data and private firms. Such data is increasingly central to a range of sectors such as banking and for verification (Singh 2018). Such approaches are controversial due to risks that they result in data profiling and increased government surveillance, but they can play an important economic role in spurring new types of verifiable data and data-driven interactions (Singh 2018).

Beyond the public sector, the emergence of data regulations (both in terms of personal and non-personal data) are providing clearer guidance around how data is collected and shared. This is important to spurring data value chains (Villani 2018). For example, well established regulations are vital in sectors such as health, urban and mobile communication where there is a complexity of data and firms. The clarity of regulation can reduce grey areas, support better data flows and clarify which data is considered personal and what can be shared (Ciuriak 2018a). Beyond regulation, government may push more active strategies that motivate actors to share data, in order to support the interconnection of datasets (Ciuriak 2018b, CNNum 2017). For example, health regulation may include requirements for actors to share subsets of data for health monitoring purposes. In China, the government has pushed policies around mandatory sharing of vehicle data as an approach to accelerate data-driven development of EV and self-driving cars (Yang 2018).

In addition to enabling data sharing and use, data regulation often embeds rules which set the boundaries for data collection and sharing. For personal data, most countries have long established data protection rules which define personal data and determine how this data can be used (Berry & Reisman 2012). Other data protection rules may define how non-personal data can be created and shared (CCIA 2015, Chander & Le 2014). For example, a number of countries have adopted "tiered systems" of cyber-security or data rules, such as China, Indonesia and India. Such tiered rules define sensitive data that is seen as crucial to national security or sensitive (e.g. military data, health data). Often these rules expand into potentially additional categories (e.g enterprise data) and therefore have economic implications. In the case of China this includes commercial sensitive firm data where firms require licences and to fulfil additional requirements on how data is shared or stored (Greenleaf & Tian 2014).

Personal data is becoming central in shaping business models of large digital firms (Zuboff 2018). Responding to this, recent years have seen "analogue" personal data regulations being updated for the internet age. The flagship regulation in this area has been the EUs General Data Protection Regulation (GDPR) which looks to expand the types of data deemed personal and the scope of jurisdiction (Bird & Bird 2017). It has been a common observation that EU regulations tend to diffuse more globally (Bradford 2020). This is true of GDPR which is increasingly being replicated in developing countries and therefore the key statutes, as well as the extent of implementation in each country, will shape data industries (Greenleaf 2017). In both personal and non-personal data regulation, such rules determine if, and how, data can be moved abroad. Therefore data protection regimes have sometimes been discussed as 'disguised data localisation' rules by industry in that they can encourage data localisation by making cross-border data flows less viable (Azmeh et al. 2020). In the medium term, by providing a well-defined framework for data use, such rules may lead to a common standard of data sharing across many nations.

Storage and infrastructure

Rules linked to the local storage of data are becoming a key topic of policy discussions. A number of countries, such as Nigeria, India, Vietnam and Indonesia, are looking to institute data localisation rules that require categories of locally produced data (particularly personal data) to be stored domestically (USITC 2017). Such rules may be embedded as part of personal data protection and tiered data protection policies outlined in the previous section, but there has also been a growth in specific sectoral regulations that institute such rules, in areas such as financial, and internet of things (IoT) (Cory 2016). Initially the goal of data localisation rules was national data security and to ensure that domestic data was accessible to law enforcement by ensuring that a copy of data remained under domestic jurisdiction (Azmeh et al. 2020). Such motivations remain a factor for data localisation, but such rules can also support economic agendas such as the emergence of local data infrastructure or supporting domestic capabilities (Azmeh et al. 2020).

One argument against developing countries implementing such rules links to the demands of technology, infrastructure and skills (as outlined in the previous sections). It has been argued that local storage is less efficient and riskier due to the lack of scale and capabilities (Ezell et al. 2013). These arguments may, however, be declining in recent years as the locally required skills needed for data centres - database administration, network administration and cybersecurity have become more widespread. As internet use is growing, we are beginning to see a growing number of countries hosting local datacentres and with growing backbone infrastructure outside the EU and US (Rossiter 2017).

One can also argue that localising data and development of local infrastructure is becoming more desirable for all economic actors. As data use has grown throughout the globe, there is a need for low latency and reduced transit times for real-time and high bandwidth applications. Therefore, governments may look to policy that supports partnering and developing local firms by making local hosting more viable economically (and as an alternative to data localisations rules). A good example of this is the success of local internet exchange points (IXP) in Nigeria, Ghana and Rwanda and the spillover effects into data sectors (Kende 2020, Kende & Quast 2017). Partnerships and investments of private, public and non-profits actors supported the growth of IXPs, a local infrastructure node that allows local internet traffic to be exchanged locally without international transit, reducing overall costs. Once they were established, the presence of IXPs has created spillover effects with, for example, Google investing in local caching services within the IXPs and building new demand for local data centres, cloud computing and other infrastructure. In addition, IXPs have provided a focal point for the development of local skills in network administration, configuration, and cybersecurity (ibid).

Analytics

As outlined in earlier sections, analytics outlines a broad area of potential data activities. For developing countries it is useful to think about data analytics in terms of a ladder of different capabilities: low value data extraction and labelling that is increasingly outsourced through distributed work platforms (Gray & Suri 2019); data cleaning and simple visualisation outsourced to professional service firms; and skilled engineering of unstructured data such as voice or computer vision tasks, and training of machine learning (ML) systems.

Some of these areas are highly emergent, but in terms of skills, platforms and technologies there is potential for skills development and learning. Analytics platforms, key data science and ML libraries and deep learning models are openly available and generally with permissive licences. The solidifying of curriculums in data science and AI/ML and vibrant online communities, highlights that analytics offers a growing number of entry points in developing countries and potential for skills building. However, the challenge is that data platforms, deep learning models and high-end computing resources often reside within dominant technology firms and this might limit the trajectories of smaller or domestic firms in developing countries.

There is relatively limited policy in this area in developing countries, but more globally some nations that see themselves as behind the curve have attempted to undertake policy. This has led to strategies that institute or support data or AI labs, demonstrators or other projects created with a goal of developing national capacities in analytics. Such supported activities include cloud computing demonstrators in China and AI centres across Europe aligning with paradigms of R&D (Berry & Reisman 2012, Villani 2018). In a handful of developing countries such as India and Thailand there have been moves to follow such models. There is yet little research of how effective such actions are, and what the spillover effects are, if any, particularly in developing counties.

Approaches to push national development in data analytics from a demand side are a potential complimentary form of policy. Public sector use of analytics and AI across a range of sectors in predicting demands, automating services and exploiting data is at the heart of the future of the public sector. Therefore if products and services are procured prudently with inclusion of domestic firms, procurement rules might push innovation in the analytics area and provide support for domestic actors (Villani 2018). In India, for example, public sector procurement in areas such as education and transport have begun to look to procure big data systems and nascent AI that includes domestic vendors (Rakesh et al. 2018, Sengupta et al. 2016). The challenge for successful policy in the contexts of analytics is that developing countries face a risk of 'brain drain', as individuals migrate or nurtured firms are bought up by international giants.

Application

The use of data-driven applications in the economy is an important final step in the data value chain. In this area, data policy is liable to overlap with broader policies around the digital economy, and in developing countries there is also likely to be close links between the state of ICT regulation and extent of internet access in a country. Evidence from analysis of countries such as Brazil and Indonesia, for example, has suggested that where broader connectivity has been strongly advocated for in policy, this has led to a more vibrant digital economy and resulted in enhanced demands for data over time (Azmeh & Foster 2018b). Clearer regulation and connectivity therefore enable more firms and users to go online, and over time undertake more advanced activities (Kelly & Rossotto 2011). This is liable to be a basis for growing collection and storage of data and a demand for more sophisticated data analytics.

As outlined previously, use of data within applications will encompass a very broad range of potential activities: standalone applications, cloud systems, integration of analytics and AI products, as well as a range of services. Therefore on the supply side, skills and technologies demands in this area will vary across different types of application. Skills will align with mainstream software development skills, such as application development, programming, business logic and information systems design. In many of these areas, open source technology stacks provide the basis for applications and are accompanied by rich availability of education and learning material including in developing countries. Thus, even if specific sectors utilise proprietary tools, there are clear entry points, as shown by the expanse of "app entrepreneurship" emerging in developing countries, often involved developing data-rich applications (Friederici 2016).

For supply-side policy, policy might have the goal to accelerate the local creation and use of such applications, and therefore drive domestic data industries upstream. Countries, regions or sectors often build strategies that either subsidise or build platforms. For example in Thailand e-commerce and payment resources have been built to support local sellers of goods online, and there are also examples of attempts to create domestic agriculture and digital work platforms with government support (Krishnan et al. 2020, Parthasarathy & Matilal 2019).[7] While there are indications that a large number of these applications may fail to reach critical mass, they are relatively cheap to support and with potentially huge payoff.

[7] For example, in Thailand, the successful state-driven platform ThaiTrade platform has supported cross-border sellers. There are now growing discussions about how to use the data generated. In Uganda, the e-voucher platform used by over 800,000 framers provides a way for farmers to purchase high quality inputs. The Nigerian platform NaijaCloud was a less successful attempt to create a platform for digital work with the country supported by the World Bank.

The area of applications is also subject to policies which look to shape how international firms operate by filtering data flows associated with applications. At their most strict, a small number of countries have implemented strict filtering of data flows of specific firms or sectors, the most well-known being China's great firewall. While this is principally discussed in terms of censorship, it also has an economic component that policy makers are aware of and have used to protect local sectors from international competition (Liu 2011). It has been argued that the largest tech firms in China have rapidly grown due to the protection that the great firewall and other policies has provided them with (Foster & Azmeh 2020). Such application blocking rules are less common elsewhere and are often associated with applications less related to the data value chain such as digital telecommunication and media (Cohen & Southwood 2004).

Policies around data applications have also focussed on shaping the actions of international firms as they enter developing countries. Rules around licencing or local requirements may shape how international data firms operate and how they use data. Licencing with conditions for firms to operate are emerging in areas such as FinTech and online transport, in countries such as China, Brazil, India (Mozur 2016, USTR 2018, Zanatta & Kira 2018). Such rules focus on building an orderly sector and reducing unethical behaviours, but they can also have stronger conditions. One example of this is physical presence rules that require platform firms to be registered within the country in order to operate (Aaronson 2016). In Nigeria and Indonesia, local content rules require that international firms source local digital resources or content as part of entering markets (Cory 2016). Partnership rules may also exist, which require foreign firms to partner with local ones for market access (USITC 2017).

Such rules for developing countries might allow them to accelerate the spillovers of technologies and skills from international to domestic actors, while still allowing international firms to gain market access. However, these rules may only be viable in developing countries with large and lucrative domestic markets. Imposing strong conditions on international firms operating in small countries could just dissuade them from operating (Foster & Azmeh 2020). There are also future challenges for small countries if trade agreements begin to restrict such policies (Azmeh et al. 2020).

4. Policy strategies in the data value chain

We have outlined a range of policies, and how they fit with specific stages of the data value chain. This discussion reinforces that building data economies is likely to go into a broad set of policies which reduce and expand data frictions across different stages and actors. In this section, based on this framework, we analyse how states are looking to operationalise different policies into more coherent national "data pathways" – that is, strategies or agendas that drive broader goals around data and the digital economy.

Some evidence suggests that larger nations will look to undertake policy initiatives across all stages of the data value chain in order to catch up. China and the EU, for example, have relatively broad policies across all stages, and while India and Brazil do have some specific policy emphasis (see below) they also have policies across many steps of the value chain (Azmeh & Foster 2018b, Foster & Azmeh 2020, Gruber 2017).

However, such an approach may only be suited to higher income or larger emerging countries, with existing capabilities across the data value chain. In contrast, as discussed in this section, there is evidence that other nations, particularly lower income ones, are trying to focus on specific areas or stages where they might specialise and with this potentially leading to broader spillover in the future. Unpacking some of the key pathways and visions of catch-up can provide an important first step to understand how data policies might be strategically implemented for broader development. Based on our analysis, we discuss four pathways of policy that are emerging globally: 1) active data localisation, 2) strategic data sharing, 3) opportunities in low income data processes and 4) building sectoral specific application linked to data.

Pathway one: Active use of data localisation rules as a foundation for local ecosystem

For a number of countries, particularly in neighbouring nations in Asia, China's model of interventionist digital policy is seen as a successful approach in supporting the digital economy (Foster & Azmeh 2020). Therefore a potential pathway centres around nurturing local data industries through a range of policies, especially data localisation.

Indonesia is an example of a nation where a push towards localisation of data is an important part of a national pathway to developing domestic data-driven firms. Digital rules in Indonesia use data localisation measures that mandate local storage for various types and categories of data (Azmeh & Foster 2018b). Economic research has suggested that such policies might lead to negative economic impacts over nations that allow "free flows of data" (e.g. Bauer et al. 2014), but such findings often come into tension with empirical cases such as China and Indonesia, where such rules appear to be one aspect of supporting domestic firms. To illustrate the different arguments, Indonesia is notable in the Asian region for the vibrancy of digital and data firms emerging in recent years, including a number of so-called "Unicorn start-ups" that have expanded into the region and even globally (Azmeh & Foster 2018). In contrast to this potentially more optimistic story is that of Nigeria. While Nigeria has attempted to push data localisation and local content rules (USTR 2018), these have not been well implemented or enforced,[8] and there is little evidence to indicate that they have sparked a broader emergence of data industries.

Pathway two: Facilitating rich data to support upstream data economies

One important pathway observed in a number of developing country cases revolves around national strategies that see encouraging richer use of data locally as the basis to enhance data economies. This is manifested in different ways across different countries. The emergence of initiatives that push expansion of data availability in India have been well documented and outline how these play a potentially important role in supporting data-driven sectors (Singh 2018). Much interest has centred on the so-called IndiaStack, and the set of services, interfaces and clear regulation around identity, universal payments and other document and certificate services (Singh 2018). Alongside this, public initiatives have looked to open up data availability across public sector in areas such as transport (Heeks et al. In press, Rakesh et al. 2018, Sengupta et al. 2016). These trajectories require further analysis, but the expansion of data availability offers opportunities for the emergence of new data-driven firms in areas such as finance and public data analytics.

Elsewhere similar attempts to build data economies through leveraging rich data can be seen. In Chile, as well as some other Latin America counties, a strong focus on open government data was mainly orientated towards increasing government transparency, but these can push opportunities for data driven firms, although it is unclear if these will lead to broader impacts (Gonzalez-Zapata & Heeks 2015). Thailand has some parallels to India in pushing data flows linked to financial data, for example a relatively low profile inter-banking data initiative called Promptpay (and follow up projects) have resulted in strong spillover outcomes around inter-bank and data driven finance (BoT 2017, BoT 2015).

Pathway three: opportunities in low income data processes

In the area of information processing and analytics, a number of nations have stood out as being centres of low-value data work, including India, Pakistan and the Philippines (and to a lesser extent some better connected cities in Africa) (Graham et al. 2017, Malik et al. 2016). An important question is if these "low value" roles can act as a stepping stone for upgrading of digital workers to "high value" roles. In East Africa, there is some evidence this may occur. In research detailing the unstable environments of higher skilled digital and data workers, digital work platforms such as Upwork were used within firms as a fall-back measure for gaining work when these firms faced challenges in attracting businesses. This suggests that there are important connections of lower

[8] Authors discussions with policy experts in West Africa.

value and higher value digital analytics (Mann et al. 2015). Notwithstanding such examples, overall in Africa there appears to be limited evidence of dynamic upgrade paths in this area (Graham et al. 2017).

In locations such as the Philippines and India where data processing tasks have been well established for a number of years within established BPO/ITES industries, there does appear to be some trends that well established IT firms are climbing the ladder towards higher value roles (Aldaba 2019). These industries are well supported with strong industry-government links, and adaptive policy over time as the industry evolves (Parthasarathy 2010).

Pathway four: building sectoral specific application linked to data

Data economies might also be driven from the application side of the data value chain - that is, policy which drives adoption of data rich applications leading to new opportunities and demands across the data value chains. One area that has received significant policy attention in developing countries is building strategies linked to the so-called 'fourth industrial revolution.' Key to these visions is the idea of data integration in key industrial sectors will allow more dynamic and agile production and so these initiatives often have a strong data component.

For countries already with more advanced production, such as Italy (MdSE 2018, Miragliotta 2018) and France (EC 2017), industrial applications involving data have been seen as an area where domestic firms can build competitive advantage. They offer opportunities for data firms to expand in the digital economy where there is perceived lower dominance of the US (EC 2017). Countries with smaller production sectors, but with pockets of more advanced activity such as South Africa, Mexico and Malaysia are also looking to Industry 4.0 strategies to use data in manufacturing as a path to upgrade other less advanced sectors in these countries (DTI 2017, MITI 2018, MoE 2016, Santiago 2018).

Potentially there might be other similar pathways from applications to enhanced data sectors, although evidence is more anecdotal in terms of explicit policies. Kenya, for example, through early adoption of mobile money and enlightened regulations has become a hub for data rich financial and fintech products (Sy et al. 2019). Brazil has seen a growth in domestic e-commerce firms through a combination of the specific consumer needs (e.g. language, payment approaches) and the regulatory environment (e.g. tariffs of international payments) leading to larger domestic firms who are increasingly using customer data more widely (de Lima 2017). These pathways are less clear than those based around the Fourth Industrial Revolution, but together they do highlight that there is likely to be a range of pathways stemming from appropriate applications across different sectors.

5. Conclusions

The rapid advancement of digital technology is driving structural economic and technological changes in the world. The result of these changes can be seen in the rise of new modes of production, exchange, and consumption, the emergence of new economic actors, and the growing challenges to the policy landscape regulating the world economy. Data is central to these changes as the ability to collect, move, store, and analyse data is fundamental to new business models and actors. As a result, debates around data and how flows of data can be regulated globally have increased in recent years as states work to develop tools to extract more value from the data produced within their borders and aim to limit the potential economic and social negative impact of new modes of production and trade.

This paper has aimed at developing this discussion through thinking about the collection, storage, processing, and application of data as a value chain. While these stages do not always happen sequentially and while it is often difficult to separate those stages, this conceptualisation enables us to develop a better understanding of the data economy and to examine issues such as barriers to entry, and the role of different types of capacities in different stages.

This conceptualisation enables us to examine the different policies with relation to data and the data economy, and how these policies could translate into different pathways to achieve digital development. We have identified four pathways to digital development that focus on different stages in the data value chain: 1) active data localisation, 2) strategic data sharing, 3) opportunities in low income data processes and 4) building sectoral specific application linked to data.

An important research direction forward is to understand the viability of these pathways in more detail and what approach works for different countries and economies. Policies that support data localisation, for example, can be useful for larger economies with potential to develop the broader data infrastructure whether through domestic investments or through offering enough market for leading digital firms to invest in those economies. While data storage, as discussed earlier, is a substantial economic sector in terms of capital investments, it is not as significant on its own in terms of job creation and it is not yet clear to what extent investments in storage have enough linkages to drive investments in other stages of the value chain.

Smaller countries, in terms of population and size of economy face a far greater challenge in terms of following a similar approach. For many of those countries, pursuing a "storage-led" approach in the hope that localising data would drive other investments in the value chain is unlikely to succeed due to the large capital investments and technical capacities needed for this stage. In some cases, pursuing data localisation might harm these economies by limiting their ability to benefit from the global data storage infrastructure to focus on developing capacities in other areas. Alternatively, focusing on developing capabilities in data analytics or application using the global cloud computing infrastructure can prove to be more successful for some economies through approaches such as low-income data processes and building sectoral specific application linked to data. Such an approach is, however, also faced by a range of challenges including lack of human capabilities in data analytics and application and the challenges facing upgrading from low-value added to high value-added activities. Furthermore, merely integrating in the "global" digital economy and liberalising digital trade and data flow will not necessarily drive advancement in those other activities leaving the question of what education, infrastructure, and start-up policies are needed to drive those areas.

This paper aimed to provide an initial identification of these different digital development pathways by bringing the discussion of policies and regulation closer to the discussion of the data economy and a stronger understanding of data processes. This effort, however, is only a first step in developing a better understanding of the different pathways countries could adopt to achieve digital development and what policies are more likely to succeed in driving each of these pathways. Further work is needed to unpack those issues further.

References

Aaronson, S.A. (2016) *The Digital Trade Imbalance and Its Implications for Internet Governance*, 25, Centre for International Governance Innovation, Waterloo, Canada.

Abd Rabuh, A., El-Masri, M., & Al-Yafi, K. A Text Analytics Approach to Evaluate Paradigm Shifts in Information Systems Research.

Aldaba, R.M. (2019) Mapping the Philippines in the Offshoring Services Global Value Chain. *Journal of Southeast Asian Economies*, 36(2), pp. 153–182.

Alshamaila, Y., Papagiannidis, S., & Li, F. (2013). Cloud computing adoption by SMEs in the north east of England: A multi-perspective framework. Journal of Enterprise Information Management, 26(3), 250-275

Andreosso-O'Callaghan, B., Lenihan, H. & Reidy, P. (2015) The Development and Growth of the Software Industry in Ireland: An Institutionalized Relationship Approach. *European Planning Studies*, 23(5), pp. 922–943.

Arora, P. (2016) Bottom of the Data Pyramid: Big Data and the Global South. *International Journal of Communication*, 10, pp. 1681–1699.

Azmeh, S. & Foster, C. (2018a) *Bridging the Digital Divide and Supporting Increased Digital Trade - Scoping Study*, GEG South Africa, Pretoria, South Africa.

Azmeh, S. & Foster, C. (2018b) *Bridging the Digital Divide And Supporting Increased Digital Trade: Country Case Studies, Discussion Paper*, GEG Africa, Pretoria, South Africa.

Azmeh, S., Foster, C.G. & Echavarri, J. (2020) The International Trade Regime and the Quest for Free Digital Trade. *International Studies Review*, 22(3), pp. 671–692.

Basso, B., & Antle, J. (2020). Digital agriculture to design sustainable agricultural systems. Nature Sustainability, 3(4), 254-256.

Bates, J. (2017) The Politics of Data Friction. *Journal of Documentation*, 74(2), pp. 412–429.

Bates, J., Lin, Y.-W. & Goodale, P. (2016) Data Journeys: Capturing the Socio-Material Constitution of Data Objects and Flows. *Big Data & Society*, 3(2), pp. 1–12.

Bauer, M., Lee-Makiyama, H., van der Marel, E., et al. (2014) *The Costs of Data Localisation: Friendly Fire on Economic Recovery*, ECIPE Occasional paper, 3, European Centre for International Political Economy, Brussels.

Berry, R. & Reisman, M. (2012) Policy Challenges of Cross-Border Cloud Computing. *Journal of International Commerce and Economics*, 4(2), pp. 1–38.

Bloom, C., Tan, J., Ramjohn, J., & Bauer, L. (2017). Self-driving cars and data collection: Privacy perceptions of networked autonomous vehicles. In Thirteenth Symposium on Usable Privacy and Security ({SOUPS} 2017) (pp. 357-375).

Bird & Bird (2017) *Guide to the General Data Protection Regulation*, Bird & Bird, London, UK.

BoT (2017) *Payment Systems Annual Report 2017,* Bank of Thailand, Bangkok, Thailand.

BoT (2015) *Payment Systems Report 2015*, Bank of Thailand, Bangkok, Thailand.

Bradford, A. (2020) *The Brussels Effect: How the European Union Rules the World.* Oxford University Press, USA.

CCIA (2015) *Request for Public Comments to Compile the National Trade Estimate Report on Foreign Trade Barriers - Comments of Computer & Communications Industry Association*, Comments on National Trade Estimate Report on Foreign Trade Barriers, Docket No. 2015-0014, Office of the United States Trade Representative, Washington D.C.

Chander, A. & Le, U.P. (2014) *Breaking the Web: Data Localization vs. the Global Internet*, Legal Studies Research Paper, 378, UC Davis, Davis, CA.

Ciuriak, D. (2018a) *The Economics of Data: Implications for the Data-Driven Economy*, Data Governance in the Digital Age Essay Series, ID 3118022, Center for International Governance Innovation, Waterloo, Canada.

Ciuriak, D. (2018b) *Rethinking Industrial Policy for the Data-Driven Economy*, CIGI Papers, 192, Centre for International Governance Innovation, Waterloo, Canada.

Ciuriak, D. & Ptashkina, M. (2018) *The Digital Transformation and the Transformation of International Trade*, Issue Paper, International Centre for Trade and Sustainable Development (ICTSD).

CNNum (2017) *Opinion of the French Digital Council on the Free Flow of Data in the European Union,* French Digital Council, Paris, France.

Cohen, T. & Southwood, R. (2004) *An Overview of VoIP Regulation in Africa: Policy Responses and Proposals For the Commonwealth Telecommunications Organisation*. London.

Cory, N. (2016) *The Worst Innovation Mercantilist Policies of 2015*, Information Technology & Innovation Foundation, Washington D.C.

De Hert, P., Papakonstantinou, V., Malgieri, G., Beslay, L., & Sanchez, I. (2018). The right to data portability in the GDPR: Towards user-centric interoperability of digital services. Computer Law & Security Review, 34(2), 193-203.

Doelitzscher, F., Reich, C., & Sulistio, A. (2010, June). Designing cloud services adhering to government privacy laws. In 2010 10th IEEE International Conference on Computer and Information Technology (pp. 930-935). IEEE

Drake, Wi., Cerf, V. & Kleinwachter, W. (2016) *Internet Fragmentation: An Overview,* World Economic Forum, Geneva, Switzerland.

DTI (2017) *Industrial Policy Action Plan - 2017/18 - 2019/20*, Department: Trade and Industry, Pretoria, South Africa.

EC (2017) *Key Lessons from National Industry 4.0 Policy Initiatives in Europe*, Digital Transformation Monitor, European Commission, Brussels, Belgium.

Ezell, S.J., Atkinson, R. & Wein, M.A. (2013) *Localization Barriers to Trade: Threat to the Global Innovation Economy*, Information Technology & Innovation Foundation, Washington D.C.

Flyverbom, M., Madsen, A.K. & Rasche, A. (2017) Big Data as Governmentality in International Development: Digital Traces, Algorithms, and Altered Visibilities. *The Information Society*, 33(1), pp. 35–42.

Foster, C.G. & Azmeh, S. (2020) Latecomer Economies and National Digital Policy: An Industrial Policy Perspective. *Journal of Development Studies*, 56(7), pp. 1247–1262.

Foster, C., Graham, M., Mann, L., Waema, T., & Friederici, N. (2018). Digital control in value chains: Challenges of connectivity for East African firms. Economic Geography, 94(1), 68-86.

Friederici, N. (2016) *Innovation Hubs in Africa: Assemblers of Technology Entrepreneurs*. PhD Thesis. Oxford, UK.

GCIG (2016) *One Internet*, Global Commission on Internet Governance, Waterloo, Canada.

Gonzalez-Zapata, F. & Heeks, R. (2015) The Multiple Meanings of Open Government Data: Understanding Different Stakeholders and Their Perspectives. *Government Information Quarterly*, 32(4), pp. 441–452.

Graham, M., Hjorth, I. & Lehdonvirta, V. (2017) Digital Labour and Development: Impacts of Global Digital Labour Platforms and the Gig Economy on Worker Livelihoods. *Transfer: European Review of Labour and Research*, 23(2), pp. 135–162.

Gray, M.L. & Suri, S. (2019) *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*. Houghton Mifflin Harcourt, Boston.

Greenleaf, G. (2017) *'European' Data Privacy Standards Implemented in Laws Outside Europe*, Law Research Series, 21, University of New South Wales, New South Wales, Australia.

Greenleaf, G. & Tian, G. (2014) *China Expands Data Protection through 2013 Guidelines: A 'Third Line' for Personal Information Protection*, Legal Studies Research Paper Series, 2014/15, University of Technology Sydney, Sydney, Australia.

Gruber, H. (2017) Innovation, Skills and Investment: A Digital Industrial Policy for Europe. *Associazione Amici Di Economia e Politica Industriale*, 44(3), pp. 327–343.

Han, J., Pei, J., & Kamber, M. (2011). Data mining: concepts and techniques. Elsevier.

Heeks, R., Rakesh, V., Sengupta, R., et al. (In press) Datafication, Value and Power in Developing Countries: Big Data in Two Indian Public Service Organisations. *Development Policy Review*.

Heeks, R. & Renken, J. (2018) Data Justice for Development: What Would It Mean? *Information Development*, 34(1), pp. 90–102.

Herhausen, D., Binder, J., Schoegel, M., & Herrmann, A. (2015). Integrating bricks with clicks: retailer-level and channel-level outcomes of online–offline channel integration. Journal of retailing, 91(2), 309-325.

IDA (2018) A Study of the Economic Benefits of Data Centre Investment in Ireland, a study by GrantThornton for IDA Ireland.

IDC & OpenEvidence (2017) *European Data Market,* SMART 2013/0063, European Commission, Brussels, Belgium.

Kaplan, J.M. & Rowshankish, K. (2015) *Addressing the Impact of Data Location Regulation in Financial Services*, GCIG Paper, 14, Center for International Governance Innovation.

Kelly, T. & Rossotto, C.M. (2011) *Broadband Strategies Handbook*. World Bank Publications, Washington, D.C.

Kende, M. (2020) *Anchoring the African Internet Ecosystem: Lessons from Kenya and Nigeria's Internet Exchange Point Growth*, Internet Society, Geneva, Switzerland.

Kende, M. & Quast, B. (2017) *The Benefits of Local Content Hosting: A Case Study*, Internet Society, Geneva, Switzerland.

Kozul-Wright, R. & Banga, R. (2020) *Moratorium on Electronic Transmissions: Fiscal Implications and Way Forward*, UNCTAD Research Paper, 47, UNCTAD, Geneva, Switzerland.

Krishnan, A., Banga, K. & Feyertag, J. (2020) *Platforms in Agricultural Value Chains: Emergence of New Business Models*, ODI, London, UK.

Li, W. C., Nirei, M., & Yamana, K. (2019). Value of data: there's no such thing as a free lunch in the digital economy. US Bureau of Economic Analysis Working Paper, Washington, DC.

Lian, J. W., Yen, D. C., & Wang, Y. T. (2014). An exploratory study to understand the critical factors affecting the decision to adopt cloud computing in Taiwan hospital. International Journal of Information Management, 34(1), 28-36.

Liao, Y., Ramos, L.F.P., Saturno, M., et al. (2017) The Role of Interoperability in The Fourth Industrial Revolution Era. *IFAC-PapersOnLine*, 50(1), pp. 12434–12439.

Low, C., Chen, Y., & Wu, M. (2011). Understanding the determinants of cloud computing adoption. Industrial management & data systems.

de Lima, G.A. (2017) *National Report on E-Commerce Development in Brazil*, 14, UNIDO, Vienna, Austria.

Liu, C. (2011) Internet Censorship as a Trade Barrier: A Look at the WTO Consistency of the Great Firewall in the Wake of the China-Google Dispute. *Georgetown Journal of International Law,* 42(4), pp. 1199–1242.

López González, J. & Jouanjean, M.-A. (2017) *Digital Trade: Developing a Framework for Analysis,* OECD Trade Policy Papers, 205, OECD Publishing, Paris, France.

Magalhaes, G., Roseira, C. & Manley, L. (2014) Business Models for Open Government Data. *Proceedings of the 8th International Conference on Theory and Practice of Electronic Governance*, pp. 365–370.

Malik, F., Nicholson, B. & Morgan, S. (2016) Assessing the Social Development Potential of Impact Sourcing, in *Socially Responsible Outsourcing: Global Sourcing with Social Impact*, Technology, Work and Globalization, B. Nicholson, R. Babin, & M. C. Lacity (eds), Palgrave Macmillan UK, London, pp. 97–118.

Mann, L. (2018) Left to Other Peoples' Devices? A Political Economy Perspective on the Big Data Revolution in Development. *Development and Change*, 49(1), pp. 3–36.

Mann, L., Graham, M. & Frederici, N. (2015) *The Internet and Business Process Outsourcing in East Africa*, Project report, Oxford Internet Insitute, Oxford, UK.

MdSE (2018) *National Plan Impresa 4.0*, Minstero dello sviluppo economico, Rome, Italy.

Mejias, U. A., & Couldry, N. (2019). Datafication. Internet Policy Review, 8(4).

Meltzer, J.P. (2015) The Internet, Cross-Border Data Flows and International Trade. *Asia & the Pacific Policy Studies*, 2(1), pp. 90–102.

Mercier, M., Glesser, D., Georgiou, Y., & Richard, O. (2017). Big data and HPC collocation: Using HPC idle resources for Big Data analytics. In 2017 IEEE International Conference on Big Data (Big Data) (pp. 347-352). IEEE.

Miragliotta, G. (2018) Nuovo Piano Impresa 4.0, Miragliotta: "Ecco Perché è Un Passo Indietro Pericoloso per Il Paese." *Agenda Digitale*. Available from: https://www.agendadigitale.eu/industry-4-0/nuovo-piano-impresa-4-0-polimi-ecco-perche-e-un-passo-indietro-pericoloso-per-il-paese/ [Accessed 21 January 2019].

MITI (2018) *National Industry 4.0 Policy Framework*, Ministry of International Trade and Industry, Kuala Lumpur, Malaysia.

MoE (2016) *Crafting the Future: A Roadmap for Industry 4.0 in Mexico*, Ministry of Economy, Mexico City, Mexico.

Mozur, P. (2016) Didi Chuxing and Uber, Popular in China, Are Now Legal, Too, *The New York Times*, 28th Jul. Available from: http://www.nytimes.com/2016/07/29/business/international/china-uber-didi-chuxing.html [Accessed 24 August 2016].

Mattoo, A., & Meltzer, J. P. (2018). International data flows and privacy: The conflict and its resolution. The World Bank.

Oliff, H., & Liu, Y. (2017). Towards industry 4.0 utilizing data-mining techniques: a case study on quality improvement. Procedia CIRP, 63, 167.

Parthasarathy, B. (2010) The Computer Software Industry as a Vehicle of Late Industrialization: Lessons from the Indian Case. *Journal of the Asia Pacific Economy,* 15(3), pp. 247–270.

Parthasarathy, B. & Matilal, O. (2019) *The Platform Economy and Digital Work: A Developmental State Perspective*, DIODE Working Papers, 9, University of Manchester, Manchester, UK.

Pathways for Prosperity Commission. (2018). *Charting Pathways for Inclusive Growth: From Paralysis to Preparation*. Oxford, UK: Pathways for Prosperity Commission.

Perrier, A. (2018) Baidu Release Huge Dataset "ApolloScape" for Autonomous Vehicle Research [Online], *InfoQ*. Available from: https://www.infoq.com/news/2018/03/baidu-dataset-autonomous-vehicle/ [Accessed 25 July 2020].

Pilloni, V. (2018). How data will transform industrial processes: Crowdsensing, crowdsourcing and big data as pillars of industry 4.0. Future Internet, 10(3), 24.

Provost, F., & Fawcett, T. (2013). Data Science for Business: What you need to know about data mining and data-analytic thinking. "O'Reilly Media, Inc."

Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. Health information science and systems, 2(1), 3.

Rakesh, V., Heeks, R., Chattapadhyay, S., et al. (2018) *Big Data and Urban Transportation in India: A Bengaluru Bus Corporation Case Study*, Development Informatics Working Paper, 72, University of Manchester, Manchester, UK.

Rossiter, N. (2017) FCJ-220 Imperial Infrastructures and Asia beyond Asia: Data Centres, State Formation and the Territoriality of Logistical Media. *The Fibreculture Journal,* (29).

Rossotto, C. M., Lal Das, P., Gasol Ramos, E., Clemente Miranda, E., Badran, M. F., Martinez Licetti, M., & Miralles Murciego, G. (2018). Digital platforms: A literature review and policy implications for development. Competition and Regulation in Network Industries, 19(1-2), 93-109.

Sampath, P.G. (2018) *Promoting Transformation in the Digital Economy*, Global Development and Environment Institute Working Paper, 18–04, Tufts University, Medford, MA.

Santiago, F. (2018) *You Say You Want a Revolution: Strategic Approaches to Industry 4.0 in Middle-Income Countries*, Inclusive and Sustainable Industrial Development Working Paper Series, WP 19, UNIDO, Vienna, Austria.

Sengupta, R., Heeks, R., Chattapadhyay, S., et al. (2016) *Exploring Big Data for Development: An Electricity Sector Case Study from India*, Development Informatics Working Paper, 66, University of Manchester, Manchester, UK.

Singh, P.J. (2018) *Digital Industrialisation in Developing Countries - A Review of the Business and Policy Landscape*, IT for Change, Delhi, India.

Singh, P.J. & Vipra, J. (2019) Economic Rights Over Data: A Framework for Community Data Ownership. *Development*, 62, pp. 53–57.

Skordas, T. (2019). Toward a European exascale ecosystem: the EuroHPC joint undertaking. Communications of the ACM, 62(4), 70-70.

Sy, A.N., Maino, R., Massara, A., et al. (2019) *FinTech in Sub-Saharan African Countries; A Game Changer?*, International Monetary Fund, Washington, D.C.

Talukder, A. K., & Zimmerman, L. (2010). Cloud economics: Principles, costs, and benefits. In Cloud computing (pp. 343-360). Springer, London.

Tan, L., & Wang, N. (2010, August). Future internet: The internet of things. In 2010 3rd international conference on advanced computer theory and engineering (ICACTE) (Vol. 5, pp. V5-376). IEEE.

Thatcher, J., O'Sullivan, D. & Mahmoudi, D. (2016) Data Colonialism through Accumulation by Dispossession: New Metaphors for Daily Data. *Environment and Planning D: Society and Space*, 34(6), pp. 990–1006.

Turban, E., Sharda, R., Aronson, J. E., & King, D. (2008). Business intelligence: A managerial approach (pp. 58-59). Corydon^ eIndiana Indiana: Pearson Prentice Hall.

USITC (2017) *Global Digital Trade 1: Market Opportunities and Key Foreign Trade Restrictions,* 4716, United States International Trade Commission, Washington D.C.

USTR (2018) *National Trade Estimate Report on Foreign Trade Barriers*, US Trade Representative, Washington D.C.

Villani, C. (2018) *For a Meaningful Artificial Intelligence Towards a French and European Strategy*, French parliamentary mission, Paris, France.

Weber, S. (2017). Data, development, and growth. Business and Politics, 19(3), 397-423.

Wolfert, S., Ge, L., Verdouw, C., & Bogaardt, M. J. (2017). Big data in smart farming – a review. Agricultural Systems, 153, 69-80.

WRC (2013) *Economic Impact of Data Centres on Central Washington,* Washington Research Council, Washington, D.C.

Yang, Y. (2018) China Demand for Electric Vehicle Data Raises Privacy Issues [Online]. Available from: https://www.ft.com/content/a868e3dc-f3b9-11e8-ae55-df4bf40f9d0d [Accessed 25 July 2020].

Zanatta, R.A.F. & Kira, B. (2018) Regulation of Uber in São Paulo: From Conflict to Regulatory Experimentation. *International Journal of Private Law*, 9(1/2), p. 83.

Zuboff, S. (2018) *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. PublicAffairs, New York.

Zuboff, S. (2015). Big other: surveillance capitalism and the prospects of an information civilization. Journal of Information Technology, 30(1), 75-89.

BLAVATNIK SCHOOL OF GOVERNMENT

UNIVERSITY OF OXFORD

Digital Pathways at Oxford