CCRED
CENTRE FOR COMPETITION, REGULATION AND ECONOMIC DEVELOPMENT

the dtic
Department:
Trade, Industry and Competition
REPUBLIC OF SOUTH AFRICA

SARChI Chair:
Industrial Development

Discussion Paper 1

DATA GOVERNANCE: TOWARDS A POLICY FRAMEWORK

Rory Macmillan

Macmillan Keck Attorneys & Solicitors

rory@macmillankeck.pro

Discussion Paper prepared for Expert Panel on Regulating Digital Platforms for Economic Development

April 2020

This paper forms part of a series of studies on the challenges of industrialisation undertaken by the Industrial Development Think Tank (IDTT). Established in 2017, the IDTT is supported by the Department of Trade and Industry (the dti) and is housed in the Centre for Competition, Regulation and Economic Development (CCRED) in partnership with the SARChI Chair in Industrial Development at the University of Johannesburg.

Contents

Executive summary

Introduction

This paper is contributed to the second *Expert Panel on Regulating Digital Platforms for Economic Development* under the auspices of the Department of Trade & Industry and the Centre for Competition, Regulation and Economic Development of the University of Johannesburg.

The Sustainable Development Goals (SDG) depend on the effective exploitation of data across numerous sectors. The wide host of issues relating to how data is to be governed in society today, whether globally, regionally or nationally is referred to here as data governance – a framework of policies, laws, regulations and processes that enable, guide, sometimes limit, and hold market participants accountable for, the collection, use and sharing of data.

Information and trust

Numerous systems gather and organize particular data to increase confidence in its reliability, rendering it useful for economic decision-making. Identification systems, consumer and credit reporting agencies, financial markets and securities exchanges, healthcare systems, education institutions, media organizations, judicial and other dispute resolution systems, all rely to some degree on certain rules about the organization and sharing of information. Some of these involve making data available while others restrict the flow of data. Trust is central to many dimensions of data governance – and not trust in the accuracy of data, but in the systems that collect, use and share it.

Data, opportunities and risks

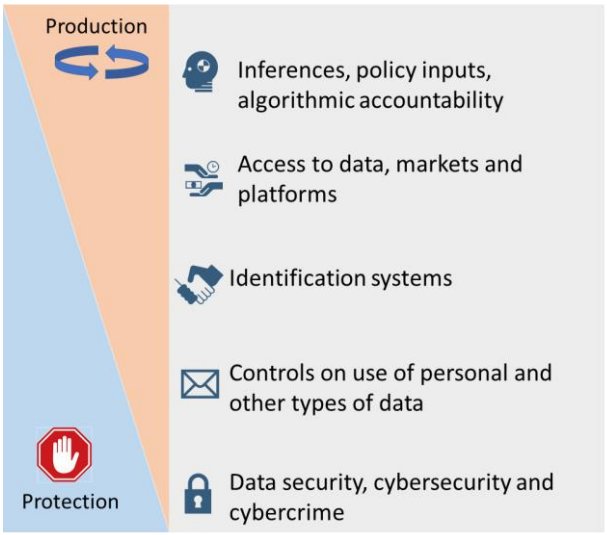
Data is non-rivalrous (it may be used by multiple persons for multiple purposes without depleting it), and is replicable and transferable at relatively low incremental cost. Technologies using it have the potential to achieve radical improvements in transport, health, financial services, energy, education and other key areas of economic and social life. Access to data can also reduce information asymmetries in business and between citizens and government, and can empower individuals.

The nature of data also introduces important risks. It is vulnerable to unauthorized access, theft and manipulation. Insecure processing, storage and transfer of data may weaken its value and usefulness, and even expose critically important infrastructure and services to serious risk. Access to data about individuals may increase a government's, a corporation's or others' power over them. Disclosure of personal data can leave individuals vulnerable to an array of potential harms, including discrimination, identity theft and violation of privacy. The opportunity and incentives to use big data for private profit or public welfare may put strain on public and private goods such as competition and privacy.

From protection to production

An effective data governance policy should address the risks inherent in its nature while ensuring that a wide variety of data will be used to its greatest economic and social potential. Good data governance recognizes data's economic and social utility and aims to deliver welfare benefits across the economy, doing so securely where security matters, all while respecting increasingly widely recognized norms of consumer protection and privacy.

Figure 1: Data governance seeks productivity founded on protection



Source: Rory Macmillan

The range of data governance issues is illustrated in Figure 1. At one end of the spectrum are foundational protective measures designed to prevent harm to systems and persons. These are prerequisites for the use of data in business, government and society, where data governance seeks to facilitate the economic productive use of data by establishing and maintaining incentives and minimizing barriers to such use. The key areas explored in this paper include:

- the security of the computing systems and telecommunications networks on which data is processed, stored and transferred to protect against unauthorized access to, or modification or theft of data (**data security, cybersecurity and cybercrime**);
- the conditions on which data about individual persons and other data may be collected, processed, retained and shared with a view to ensuring a digital environment in which populations are confident to participate (**controls on use of personal and other types of data**);
- the organization and processing of data on select attributes of individuals, legal entities and objects to enable their identification for the purpose of transacting and otherwise interacting with one another on a trusted basis (**identification systems**);
- access to data and its free flow across national and organizational boundaries, and the effective functioning of markets and business processes in which data is a central feature, with the goal of ensuring that data supports innovation, competition and trade (**access to data, markets and platforms**); and
- organization of and accountability for the use of outputs from data processing that produce insights for commercial and policy making, automated processes and decisions (**inferences, policy inputs and algorithmic accountability**).

Data security, cybersecurity and cybercrime

Many public and private infrastructure systems of vital importance to the economy depend increasingly on digital systems that host, process and transmit their data. These include

payment systems, banking networks, defence systems, electricity networks, hospitals, data centres and telecommunications networks for example. These are threatened by intentional breaches of data systems to extort money, disrupt government or business, influence political processes, and cause personal harm. Other threats include natural disasters and human error.

Data security involves sets of practices and techniques to limit the risk from these threats and allow for recovery of any lost or altered data. A high priority for a country like South Africa, which can host data processing and export data-driven software and infrastructure services that use its data processing capabilities, must be to ensure a stellar reputational level for data security.

This requires ensuring effective cyber security readiness procedures and expertise that traverse public and private sectors. It also depends on building the human capacity of policy-makers, legislators, judges, lawyers, prosecutors, investigators and civil society with regard to legal issues relating to cybercrime. This requires a multidisciplinary, multi-stakeholder, public-private approach and assessment to prioritize how it allocates scarce, capacity-building resources.

Controls on use of personal and other data

Personal data protection and privacy concern limits on who should be authorized to have access to or to alter or share personal data (which typically relates to attributes of identified or identifiable living individuals), and the conditions on which they may do so.

Privacy and personal data protection may concern competing claims to information, and may influence relationships of power, in both commercial and political contexts. Controversy over these ideas often concerns political and economic ideologies. Notwithstanding these debates, privacy and personal data protection are vital to inclusive growth of the digital economy, which depends fundamentally on achieving and sustaining widespread trust in access to and use of personal data.

South Africa's POPI Act is still not yet fully in force (and will have a 12 month grace period after it takes force before applying in full), but will position the country in the mainstream of relatively well-developed data protection laws, restricting the collection, processing and sharing of personal information. Increasing engagement to introduce measures that implement privacy protections in effect in organizations will be necessary.

The following might be considered to bolster the data protection regime being established in the POPI Act in South Africa:

- requiring privacy by design and default;
- introducing required data protection impact assessments;
- taking the reasonable expectation of the data subject into account with regard to privacy and controls on personal data;
- permitting data to be used only for legitimate purpose that are compatible, consistent, and beneficial to consumers;
- introducing information fiduciary responsibilities for certain data controllers;
- encouraging the use of data intermediaries or personal data management service providers to act as an agent or guardian on behalf of the consumer;
- introducing personal data management tools conferring on the consumer greater control over data about him or her;

- treating notions of property ownership right in personal data with caution;
- forcing a change in the underlying business models that rely so extensively on personal data; and
- improving the consumer's control over use of data about him or her.

Digital identification

Privately-operated and state-operated identification systems all present data governance issues. In addition to data security and privacy and personal data protection measures required to ensure trust in the system, a digital identification system depends on an effective 'trust framework.' This combines generally applicable laws of contract and liability, data-specific laws and standards, and identification scheme rules and protocols. A trust framework that works ensures that rights and duties of participants of the scheme are clear and ensures that there are sufficient economic incentives for them to play their respective roles, such as scheme operator, enrolment agent, consumer or authentication agent.

Other issues important to the design and operation of such schemes include non-discriminatory inclusion of the population at large, interoperability with other public agencies' and private firms' systems to leverage digital identification for multiple functional purposes, the use of open standards, and ensuring competition among scheme vendors (avoiding lock-in). Recognition of digital identification systems across borders is increasingly important.

Development of such mutual recognition standards in the SADC area and across the continent may be a valuable part of the development of regional markets in digital services in which South Africa could be a leader.

Access to data, markets and platforms

The use of data for social and economic good depends on access to it across national and organizational boundaries.

Opening up data

Proprietary data

Access to proprietary data of private organizations may have immense economic and social opportunity. A variety of institutional forms may be used to exploit the opportunity of proprietary data developed by organizations.

Interoperability, protocols and standards are required for much data sharing to be useful at all. This involves labour intensive IT work to establish, and so costs on the organization. Concerns about expropriation of a commercial asset and weakening of competitive incentives must be weighed against the benefit to society at large. Data trusts, data cooperatives and other models are mechanisms that could be deployed to allow collaboration and sharing of data for public good in a trusted manner, whether among private entities, between public sector and private sector, and across-borders.

This will require city authorities or vertical ministries to take a lead, alongside information regulators, competition authorities and private entities involved. Consumer and other civil society organisations may be able to contribute support that also builds trust among potential participants where these are individuals. They might monitor performance, and formal auditors may be required to provide reports assessing conduct against pre-agreed criteria.

Open public data

Making data held by Government and other public institutions freely available and redistributable offers important opportunities for medical, climate change and other scientific research, improved organization and regulation of public and private services, and development of new digital applications and services.

Concentrated data, competition, data portability and access

Data concentration and competition

Regulatory policy makers are today vigorously re-examining their competition laws and enforcement practices as they relate to big data, the platform economy and artificial intelligence. Data may increase information asymmetry between consumers and firms able to extract systemic information from large datasets. It may also increase information asymmetry between successful platforms and other firms. Much of the concern about market power arises from aggregation of data through vertical and horizontal consolidation, and leverage of market power from one market to another. South Africa will have to engage with these issues as its digital economy develops.

Data portability and access

Data portability and access to data have been proposed as a solution for competition problems arising from concentrated data. They can reduce switching costs by enabling the consumer to make relevant data available to an alternative service provider. However, they involve challenges as data is often unstructured, or structured in different ways in different organizations, and its organisation is often sector specific. Such remedies need to be deployed only where the benefits are likely to exceed the costs, both financial and administrative. They may be more feasible in the case of some vertical sectors, such as open banking. South Africa will need to be ready to deal with these sorts of issues as platform economy grows, and is already confronting them in the areas of healthcare, financial services and ecommerce.

Cross-border data flows – data sovereignty, ownership and localization

Access to services across borders offers huge opportunities, particularly for export to countries whose domestic tech industries may take time to develop such services. Cross-border trade in goods in the physical economy depends on cross-border flows of information to communicate demand and ability to supply, and to manage logistics and process of transport and delivery.

The cross-border dimension presents questions for policy makers and regulators. It will be vital in particular to examine whether excessive requirements to keep data within the country may undermine the efficiency and innovation opportunities of big data, the cross-border provision of cloud services, customer relationship management and regional and global value chains. Data localization may become a tool for protectionism if it effectively prevents foreign providers from offering services in a country, and so is today a key component of trade policy. At the same time, bilateral trade negotiations with leading economies that seek to minimise data localization are resulting in a patchwork approach.

South Africa may have an opportunity for export of digital services and data processing, where the rest of the continent badly lags behind. It appears likely to benefit from a relatively liberal regime for cross-border transfers. This would be supported by greater multilateral efforts to find common ground among countries on data regulation and common rules for trade in e-commerce and digital services.

Inferences, policy inputs and algorithmic accountability

Governments and businesses can use vast data troves to build a detailed personal profile of an individual and their behaviour (preferences, activities and movements) which may be used for commercial offers, State and private surveillance. They may be used to identify an individual and to determine their eligibility for a service or product. They may bring benefits to healthcare provision, medical research, transport, education, advertising, policing and the justice system.

However, use of algorithms to make decisions based on these datasets presents a new set of risks. Big datasets drawn from structured and unstructured data gathered from multiple direct and indirect sources over time risk being inaccurate or out of date. Inaccuracies may lead to erroneous inferences and decisions. Algorithms trained on data from past experience may reflect and perpetuate the biases embedded in historical treatment of ethnic, religious or gender groups even where efforts are made to avoid using special or protected categories of data about a person (such as ethnicity, religion or gender).

Initiatives are underway in many countries from several angles to address such problems. Various ideas are being developed and could be considered for introduction in South Africa's legal and regulatory framework. These include legislating for a right to receive an explanation for automated decisions, the right to appeal to a human, efforts to address algorithmic bias, and development of ethical frameworks for the use of artificial intelligence.

Conclusions

Data governance is not only important to protect the population but is strategically central to economic success and social cohesion in the future – near and far. South Africa has some strong measures in place, including a relatively well-developed data protection law in the POPI Act, even if it is not yet properly in force. But there are numerous steps that it should be considering taking in order to capture the opportunity presented by data. The protections are prerequisites to building the trust necessary to assure the growth potential, and such trust will only enhance the South African brand as the country pursues the realistic opportunity to be a regional and even international hub for data-centric services.

I. Introduction

Digital technologies are transforming economic, social and political activities – individual and social behavior, business models and processes, and the administration of government. Such technologies play an increasingly important part of every one of any government's economic development interventions – in energy, transport, health, financial and all other areas.

These technologies, and the commercial business models and governmental processes that use them, rely on the collection, processing and sharing of a variety of data in multiple ways. The Sustainable Development Goals (SDG) depend on the effective exploitation of data across numerous sectors. As data becomes more useful and cheaper to generate, process and transmit, technologies are also producing far greater volumes of it. This trend appears likely only to accelerate as the internet of things (IoT), machine-to-machine (M2M) communications and smart cities grow.

These trends give rise to a wide host of issues relating to how data is to be governed in society today, whether globally, regionally or nationally. This paper explores these themes of data as a contribution to South Africa's discussion of how to approach various issues relating to data. It is contributed to the second *Expert Panel on Regulating Digital Platforms for Economic Development* under the auspices of the Department of Trade & Industry and the Centre for Competition, Regulation and Economic Development of the University of Johannesburg.

The constellation of issues is referred to here as data governance – a framework of policies, laws, regulations and processes that enable, guide, sometimes limit, and hold market participants accountable for, the collection, use and sharing of data.

Like data itself and the opportunities and risks it presents, data governance is not limited by national boundaries but extends to transnational economic activity. The nature of data, the uses for which it may be deployed, and the challenges to which these give rise, now make data governance a vital dimension of economic development policy. For example, Japan placed data governance squarely on the international agenda for the 2019 G20 summit as it sought to ensure a free flow of trusted data.

More than ever, leadership and guidance are required to build robust and enabling data governance systems that are conducive to secure, innovative and competitive uses of data. This concept note discusses key dimensions of data governance and its relevance in particular to the Government of South Africa's work across economic development, regulation of data-intensive sectors, and application of competition policy to data.

A. Information and trust

Data governance must be considered in the wider context of the importance of trusted information to human society and economic life. At a most basic level, this may be seen in longstanding laws which establish responsibilities for inaccuracy of information about who a person is, or what he or she is selling and buying. These might include prohibitions on fraud or other basic consumer protection elements.

At a more sophisticated level, numerous systems gather and organize particular data to increase confidence in its reliability, rendering it useful for economic decision-making. Identification systems, consumer and credit reporting agencies, financial markets and securities exchanges, healthcare systems, education institutions, media organizations, judicial and other dispute resolution systems, all rely to some degree on certain rules about the

organization and sharing of information. They seek to ensure that it can be trusted for the purpose to which it will be put.

For example, identification systems gather data on verifiable attributes of individuals to enable commercial entities, governmental agencies and other individuals to have confidence that they are transacting (whether for a financial, health or basic commercial service) with the correct individual. Securities regulators and stock exchanges require the organization and public disclosure of particular data to enable investors to assess companies and make informed decisions in financial markets. They do not verify the data but those publishing it are responsible for its accuracy.

Sometimes, achieving the desired level of trust has relied on institutions and processes that validate, or reduce the transaction costs of verifying, claims about the accuracy of information. Passport agencies issue passports according to agreed standards, and other agencies issue a wide range of identification credentials to certify or otherwise provide evidence as to the identity of the bearer.

Many such institutions and processes do not merely require the generation and disclosure of such information to enable a purchaser or investor to make its own assessment, but go further in making an assessment of relevant attributes that may be relied on by others. Consumer and credit reporting agencies collect and publish particular types of personal data on individuals, corporate entities and governments to generate reputational data outputs (e.g., a credit score or ranking). These reduce the transaction costs of lenders and other commercial counterparts in assessing credit risk.

At the same time, there is considerable economic opportunity in making data available even without substantial institutional organization, as open data may be used for large scale analytics to identify trends useful for healthcare, environmental research, traffic management and numerous other purposes.

These uses of information not only concern making data available for use, reuse and sharing, but also depend on imposing (and sometimes allowing) restrictions on the flow of data. For instance, the corollary of disclosure rules in securities markets (which require publishing data) is insider trading laws (which prohibit disclosure of non-public price-sensitive information). These prevent information flows that would enable unfair arbitrage opportunities arising from information asymmetries. Companies are encouraged by legal protections to keep pricing strategies, investment plans and trade secrets confidential because the very secrecy preserves the competitive process of pursing and exploiting commercial advantage through innovation and investment. Certain professions that require particular levels of trust to provide their services are required to honour the confidentiality of their patients and clients.

Data governance does not comprise this vast realm of ways in which information is regulated – these are each subject to general and sector-specific laws and rules. Data governance concerns clusters of issues that are especially pertinent to the nature of data itself, particularly how it is used in the world of intense datacentric processes relying on computing systems and telecommunications networks. Nevertheless, as will be seen, the theme of trust that is outlined above is central to many dimensions of data governance – and not trust in the accuracy of data, but in the systems that collect, use and share it.

B. Data, opportunities and risks

Data can take many forms. It includes any raw or processed information about persons, entities, things, processes and anything else that can be represented in digital form. It may be compiled and analysed statistics, or inferences about people drawn from other data (creditworthiness, eligibility for a job, probability of recidivism) through machine learning, profiling and other data processing. Data can be uninteresting in itself, or it may be financially valuable, whether because it may be used to identify a person or entity uniquely for a commercial transaction or because it is a trade secret. It may also concern profoundly personal facts that for psychological and cultural reasons individuals or communities prefer to keep private.

Data has several important features that afford opportunities for economic development.

Unlike physical goods, but like ideas (Romer, 1990), data is non-rivalrous, replicable and transferable. Unlike a bag of rice which once eaten is gone, data (e.g., a person's location, identity or preferences) is non-rivalrous: it may be used by multiple persons for multiple purposes without depleting it. Today, with high speed connectivity, data may also be copied at low cost (except where controls act to prevent such replication, in which case it may become scarce). Data can be transferred at relatively low incremental cost, enabling unprecedented social, cultural and economic production (Benkler, 2007; Mayer-Schönberger & Cukier, 2013).

High velocity data processing applied to large volumes and varieties of data (referred to as big data) can yield disruptive innovation and substantial welfare returns. Use of big data has the potential to achieve radical improvements in transport, health, financial services, energy, education and other key areas of economic and social life.

Access to data can also reduce information asymmetries between provider and consumers, improving the provider's ability to identify consumers and tailor goods and services as well as strengthening the consumer's ability to assess and compare providers. The ability to access and use data can also empower the individual by providing greater access to services it offers, more information about technology, greater price and quality transparency, improved ability to communicate. These improvements can increase his or her capabilities and freedom to realize a fulfilled life, a core objective of development policy (Sen, 1999). Access to data about the conduct of government (revenues, budget, procurement, delivery of services) may reduce information asymmetries between citizens and government, with greater transparency potentially enabling greater accountability.

The nature of data also introduces important risks.

Data is vulnerable to unauthorized access, theft and manipulation by malicious parties. Insecure processing, storage and transfer of data may weaken its value and uses to which it may be put, and even expose critically important infrastructure and services to serious risk. A government's access to data on citizens and residents (identities, opinions, relationships, affiliations, locations) may increase its power over them. Disclosure of personal data can leave individuals vulnerable to an array of potential harms, including discrimination and identity theft. Even in the absence of a quantifiable, economic harm, such disclosure is often seen as a violation of an individual's right to privacy.

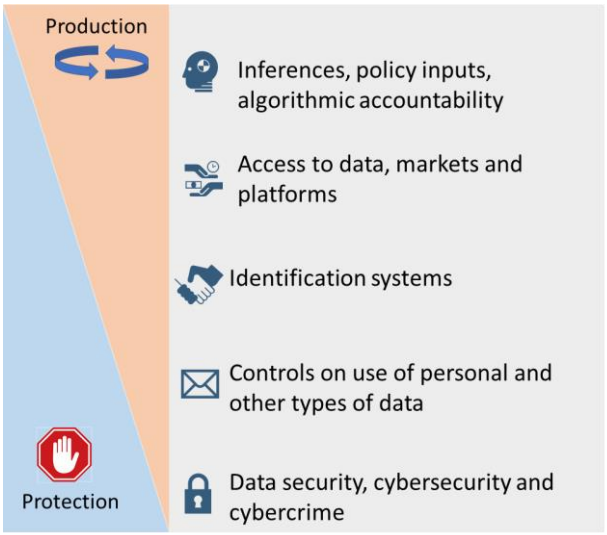
When it comes to big data, machine learning algorithms depend on a vast data catchment and substantial capital investment. The opportunity and incentives to use data for private profit or public welfare may neglect or put strain on other public and private goods, such as competition

and privacy. It may increase incentives to hoard data to create scarcity, increasing the value of data and foreclosing markets to competitors. These can lead to competition, consumer protection and other market failure concerns.

C. From protection to production

An effective policy encompassing a wide range of governance issues relating to data should address the risks inherent in its nature while ensuring that a wide variety of data will be used to its greatest economic and social potential. Data governance recognizes data's economic and social utility and aims to deliver welfare benefits across the economy, doing so securely where security matters, all while respecting increasingly widely recognized norms of consumer protection and privacy.

Figure 2: Data governance seeks productivity founded on protection



Source: Rory Macmillan

The range of data governance issues is illustrated in Figure 2. At one end of the spectrum of data governance issues are foundational protective measures designed to prevent harm to systems and persons. These are prerequisites for the use of data in business, government and society, where data governance seeks to facilitate the economic productive use of data by establishing and maintaining incentives and minimizing barriers to such use. The key areas explored in this paper include:

- the security of the computing systems and telecommunications networks on which data is processed, stored and transferred to protect against unauthorized access to, or modification or theft of data (**data security, cybersecurity and cybercrime**);
- the conditions on which data about individual persons and other data may be collected, processed, retained and shared with a view to ensuring a digital environment in which populations are confident to participate (**controls on use of personal and other types of data**);
- the organization and processing of data on select attributes of individuals, legal entities and objects to enable their identification for the purpose of transacting and otherwise interacting with one another on a trusted basis (**identification systems**);

- access to data and its free flow across national and organizational boundaries, and the effective functioning of markets and business processes in which data is a central feature, with the goal of ensuring that data supports innovation, competition and trade (**access to data, markets and platforms**); and
- organization of and accountability for the use of outputs from data processing that produce insights for commercial and policy making, automated processes and decisions (**inferences, policy inputs and algorithmic accountability**).

Beyond the areas listed above are other important vital areas of policy and law for the development of the digital economy. At the infrastructure and networks level, these include the regulation of telecommunications networks and services across which data travels, as well as competition in and standards for computer technologies used to process data. At the service and application level, other important areas include regulation of data-intensive applications in vertical sectors such as transport, finance, health, education, media (and many more), as well as the laws of electronic commerce. Government is of course involved in all of these areas, but they are not discussed here in terms of data governance.

Good data governance does not mean imposing a major set of restrictions on data generally, which might suffocate the emergence of data-driven innovation in countries that might benefit from it. It is vital to ensure proportionate intervention that does not over-architect the data ecosystem. The tools of data governance may range from 'soft' guidance in ethical principles and self-regulatory systems to 'hard' rules backed by legally enforceable processes and remedies that provide accountability where the use of data has a significant impact on peoples' lives.

The remainder of this paper discusses the five areas of data governance mentioned above. It traverses these at an overview level, but exploring in more detail governance of use of personal data, cross-border transfers of data, and artificial intelligence and automated decision-making, including options for improving South Africa's approach to these.

II. Data security, cybersecurity and cybercrime

Many public and private infrastructure systems are of vital importance to the economy, and these depend increasingly on digital systems that host, process and transmit their data, including in developing countries. These include payment systems, banking networks, defence systems, electricity networks, hospitals, data centres and telecommunications networks for example.

As the digital economy grows and electronic commerce expands, payment system data and personal data about individuals, companies and other entities will increasingly be at risk from a panoply of threats. Foremost among these are intentional breaches of data systems, which can have a range of motivations including extorting money, disrupting government or business, influencing political processes, and causing personal harm, among others. Data systems are also vulnerable to other threats, such as natural disasters and other catastrophes as well as human error. Data security involves sets of practices and techniques to limit the risk from these threats and allow for recovery of any lost or altered data.

Data security is integral to ensuring the robustness of data systems by providing protection and resilience (albeit imperfect in both cases) from these potential threats. It is particularly challenging because the threats are often unpredictable, do not respect national borders, are

technologically highly dynamic, constantly innovate with new business models, and can pose systemic risk. In the context of cyber and internet threats, data security is referred to as cybersecurity, and cybercrime is its corollary, outlawing the harmful conduct against which cybersecurity practices are designed to protect.

Countries and industry bodies are increasingly developing practices to minimize exposure to such threats and ensure readiness to cope with them. These involve sharing of information across public and private sectors, developing standards and procedures for assessing and addressing risk, coordinating contingency planning, and other practices.

A high priority for a country like South Africa, which can host data processing and export data-driven software and infrastructure services that use its data processing capabilities, must be to ensure a stellar reputational level for data security. This requires ensuring effective cyber security readiness procedures and expertise that traverse public and private sectors. It also depends on building the human capacity of policy-makers, legislators, judges, lawyers, prosecutors, investigators and civil society with regard to legal issues relating to cybercrime. This requires a multidisciplinary, multi-stakeholder, public-private approach and assessment to prioritize how it allocates scarce, capacity-building resources.

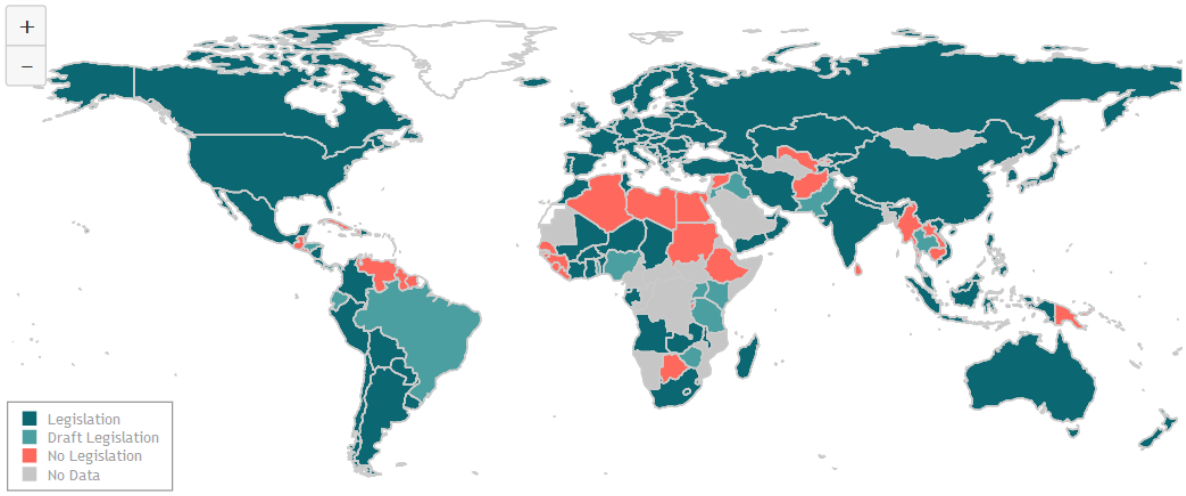
III. Controls on use of personal and other data

Personal data protection and privacy concern limits on who should be authorized to have access to or to alter or share personal data, and the conditions on which they may do so. The scope of what constitutes personal data varies by jurisdiction, but at their core it concerns attributes of identified or identifiable living individuals.

While data security is concerned with the mechanisms for protecting data, personal data protection and privacy are concerned with who should be allowed to access, disclose and alter personal data, in what circumstances and for what purposes. These decisions depend in part on the normative values of the society. Laws protecting personal data are typically undergirded by philosophical notions of individuality, autonomy, integrity, dignity, and legal notions of fairness and privacy. They increasingly seek to provide individuals with a degree of control over their personal data, a longstanding privacy concept (Westin 1962). These decisions are also likely to be contextually dependent (Nissenbaum, 2010). For example, some aspects of privacy are so intrinsic to the context of certain relationships or communications that privacy drives a need for high data security (e.g., electoral voting, legal advice or doctor-patient communications).

Privacy and personal data protection may concern competing claims to information, and may influence relationships of power, in both commercial and political contexts. Controversy over these ideas often concerns political and economic ideologies, such as about the nature of markets and government, and relationships between individuals on one hand and the State and commercial entities on the other. Some jurisdictions conceive of privacy and personal data protection as a fundamental right (EU, India), others treat it more as a matter of fairness, consumer protection and harm-based risk-management concern (USA; NIST). Section 14 of South Africa's Constitution makes privacy a fundamental right, implemented in the Protection of Personal Information Act No. 4 of 2013 (POPI Act).

Figure 3: Data protection and privacy legislation worldwide



Source: UNCTAD, 14/01/2020

Notwithstanding these debates, economic development privacy and personal data protection are vital to inclusive growth of the digital economy. Returning to our earlier theme, this depends fundamentally on achieving and sustaining widespread trust in access to and use of personal data by governmental agencies and service providers.

So, for instance, privacy and data protection inform work on digital identification systems (discussed below). The legitimacy – social and legal – of such systems, and so their uptake and usage, depend on trust, including in relation to protection of personal data. The relevance of such legitimacy has been starkly illustrated in recent court cases suspending aspects of national digital identification systems in India (Puttaswamy), Jamaica (Robinson) and Kenya (Nubian Rights).

Personal data protection and privacy make an important practical contribution to the effectiveness of commercial and public services that rely on personal data. Just as public and private services and goods that rely on personal data cannot be effective if they are hacked, they cannot work properly if they use erroneous data about individuals. Thus, personal data protection and privacy laws increasingly entitle individuals to access and rectify erroneous data about them held by providers. This in turn depends on access to mechanisms for complaint and redress. In some cases, laws require data to be kept accurate (this is particularly so in the case of medical and financial services), and to be kept only as long as it is needed.

Well-founded confidence in a system also depends on the ability to assess risk, and so requires some level of transparency about failures, such as notification and reporting of data breaches. A personal data protection regime (e.g., Europe's General Data Protection Regulation 2018, or GDPR) might also stipulate that the very collection, use and retention of data must in the first place have a lawful basis, whether on the basis of informed consent, to fulfil a contract, for certain public purposes. It might also allow constraints to be placed on the transfer of data (e.g., under the new California Consumer Privacy Act, or CCPA).

Privacy and personal data protection are closely related to all of the other aspects of data governance. Effective competition in data-driven markets (discussed below) may also align

with efforts to give consumers greater control over their data through data portability measures. Privacy enhancing technologies (PETs) are effectively data security mechanisms employed to implement privacy values, and the types and design of security mechanisms affect the potential and limits of privacy regimes.

The leading mechanism to protect personal data about individuals has been improving notification to the consumer of how his or her personal data is being gathered, used and shared, and improving his or her control over this. This has been used to address this is to reduce the asymmetry of information and bargaining power facing the consumer (ACCC 2018; Furman 2019; Cremer 2019).

Many jurisdictions apply more stringent security controls where the personal attributes are particularly sensitive (in South Africa, see POPI Act section 26). This may arise from the highly personal nature of the data (sexuality, religion, health), the importance of privacy as a feature of liberty or a functioning electoral system (political opinions), or vulnerability to use with potentially harmful consequences such as discriminatory treatment (race, ethnicity and religion).

Such an approach is coming under strain, however, as it is increasingly possible for algorithms to reverse engineer identity and discover sensitive data about a person from non-sensitive data, rendering all personal data potentially sensitive and then the additional protections ineffective. These issues relate to algorithmic accountability, where artificial intelligence may infer sensitive attributes from others (as discussed below).

There is increasing recognition that the 'notice and control' solution to empower individuals is generally inadequate. Consumers simply cannot keep up with the volume, complexity and uncertainty of information about how data about them is used, what the risks are, and what trade-offs they should consider when invited to click a consent button. It is not clear that the broken model of 'notice and control' can be solved by attempting to improve notices and ensure even more explicit consent, i.e., more of the same.

South Africa's POPI Act is still not yet fully in force (and will have a 12 month grace period after it takes force before applying in full), but will position the country in the mainstream of relatively well-developed data protection laws, restricting the collection, processing and sharing of personal information. Increasing engagement to introduce measures that implement privacy protections in effect in organizations will be necessary.

Other countries are only now grappling with how better to achieve data protection effectively rather than conceiving of it as primarily a matter of consumer control over data about him or her when this is not realistic. South Africa may wish to consider enhancing the POPI Act and taking other measures including exploring the following ways of shifting more of the burden from the individual to service providers.

The following might be considered to bolster the data protection regime being established in the POPI Act in South Africa:

A. Privacy by design and default

The POPI Act establishes strong legal bases for processing personal data, but could have stronger provisions requiring the incorporation of privacy principles into the design of products and services, ensuring that privacy is the default setting for people interacting with them (Cavoukian, 2011). This could be done in a manner whereby the simple obtaining of consent

would not override the privacy protections. Regulation, such as through codes of conduct developed by the information regulator, can require organizations handling larger amounts of sensitive data to implement technical and administrative measures, such as pseudonymization, to apply data-protection principles such as data minimization effectively.

B. Data protection impact assessments

Another way of deepening and accelerating the incorporation of data protection into the fabric of organizations is to require them to carry out impact assessments at every level of the organization. This might be begun with organizations that process larger amounts of personal data, particularly where it is sensitive or strategically important, such as health and financial data. Organizations can be required to report to the information regulator, and engage in a process of learning and reform, which can be shared with a broader range of entities across the economy.

C. Reasonable expectation of the data subject

A trade-off needs to be struck between the benefit of data processing, sharing and reuse of data on the one hand, and the importance of privacy and controls on personal data on the other. Instead of allowing unlimited reuse on the basis of consent, limits may be applied taking into account the reasonable expectation of the data subject in context. This is how US privacy law evolved before the Fair Trade Commission. It suffers from the weakness that it is particularly context specific, and consumer expectation and awareness are increasingly difficult to ascertain in an ever-more complex data environment. But it may nevertheless be a useful dimension to add to the regulatory framework.[1]

D. Legitimate purpose

Related to the reasonable expectation requirement above is a proposal (Medine & Murthy, 2020) to limit permitted data use to 'legitimate purposes,' i.e., uses that are 'compatible, consistent, and beneficial to consumers.' Such a basis would be more important than consent as a justification for processing a person's personal data, as in New Zealand (Edwards 2019). Firms would be permitted, however, to use robustly de-identified data to develop new and innovative products and services. This would not be overridden by consent (unlike GDPR and many other countries' laws), so that even individuals who consent to use of data about them would be protected by this test. The data could still be used for a wide variety of purposes related to those the individual signed up for with the associated service – e.g., servicing accounts, fulfilling orders, processing payments, collecting debts, controlling for quality, enforcing security measures, or conducting audits. Additionally, data could be used innovatively if consistent with the service in relation to which the data were initially collected.

E. Information fiduciaries

An obligation could be introduced for data controllers to owe a sort of fiduciary duty of loyalty to data subjects regarding treatment of personal data about them (Balkin 2018; Zittrain, 2018; Balkin & Zittrain, 2018). Doctors and lawyers owe duties of confidentiality to their clients and are not permitted use the information they collect about them against their interests. Due to the amount of information they know about their clients and the client's asymmetric dependence on them, they are required to act in good faith at risk losing their licence to practice and claims from their clients.

[1] The Consumer Privacy Bill of Rights proposed by President Obama's White House in 2012 took this approach. https://obamawhitehouse.archives.gov/sites/default/files/privacy-final.pdf.

Similarly, online service providers that have become indispensable might be subject to such obligations. These might prevent the service provider from distracting and guiding the data subject towards services he or she doesn't need. The fiduciary model has been proposed in India and the US. It is potentially powerful for the service provider that collects data from the subject. It is less likely to succeed with respect to third parties which acquire such data in a secondary market and have no relationship with the individual concerned.

F. Data intermediaries

A market-based model worth exploring and encouraging is the use of service providers, even if automated, to act as an agent or guardian "algorithmic angels" – on behalf of the consumer (Koponen, 2014). Personal data management service providers could carry out automated negotiations on behalf of individuals regarding authorization of use of personal data about them by service providers, for example according to pre-agreed profile, monitor compliance, retract consent if violated, and other measures (Orcutt, 2017).

A commercial model for such consent management services is evolving, for instance with firms like Sudo providing a pseudonym for subscription for services (e.g., e-commerce and online dating), or Apple's provision of anonymous sign-in for mobile apps using randomly generated email addresses (Herrera & Haggin, 2019). Such a commercial model can be encouraged through regulatory nudges, by creating the category of service provider, providing certification, safe harbour for liabilities, and even fiscal incentives.

G. Personal data management tools

Related ideas involve the consumer generally having greater control over their data. India's "Digital Locker" allows people to store their data and then control who may access it, including producing an auditable record of when their records are accessed. Similar approaches have been pursued in Estonia, for example.

H. Property ownership right in personal data

Other ideas include conceiving of a property right of ownership over personal data, although this has approach not yet gathered steam. One theoretical benefit is the greater agency it might confer on a data subject in the management of data about him or her, and aligning the incentive to exploit it with the person most concerned, and countering the risk of hoarding by organizations that have little or no opportunity cost to doing so. In particular, people can weight their privacy concerns against the economic gains from selling data about them (Jones, C. & Tonetti, C., 2018).

One concern about 'propertisation' of personal data is the risk that it become yet another resource subject to the logic of the market, and likely to be sold, with lower income, poorer educated individuals likely to face greater exploitation from large organizations, perpetuating inequality (Kerry & Morris, 1999). There are also basic challenges in making a non-rivalrous resource rivalrous through the exclusivity of property where data about one person may also include data about another, such as a contact list, picture of a family including pictures of more than one person.

I. Force change in the business model

The core problem with privacy is that the predominant internet-based business model that has emerged depends on attracting the attention and participation of users, particularly to services that are free to them, and the associated capture of their personal data. So long as this

business model exists, data protection will face an uphill struggle against commercial incentives. Even breaking up tech companies will not reduce this problem (Zuboff 2019).

One approach suggested has been to encourage businesses away from the advertising revenue model and towards subscription-based models, as has been suggested by Margrethe Vestager, the EU competition commissioner who is also responsible for ensuring Europe's readiness for the digital age.[2] Banning or limiting microtargeting of advertising, recommendations and other behavioural nudges would also be a means of reducing the strength of a business model that elevates the commercial value of capturing personal data. This is, of course, a challenge for any one country or even regional regulator on its own, but engagement with authorities in other regions, particularly Europe, may lead to measures in this direction.

J. Improved consent management

While consent has important limitations, some improvements may yet be made to the means by which it is obtained. Beyond improving the language of notifications, tiered consent might differentiate between types of data according to the types of purpose for which it may be used or which types of organisation may use it. Sunset clauses could be used by which consent will expire (Custers, 2016).

IV. Digital identification

Individuals, entities and objects have a variety of attributes and behaviours. We focus here on individuals, where the development of identification systems presents major economic development objectives, including towards the Sustainable Development Goals' target of achieving legal identification for every individual by 2030 (SDG Target 16.9).

Individuals can be identified using a wide range of data points as their attributes and behaviours are widely spread across their interactions in the digital economy, whether through communications, online searches, electronic commerce, social media, use of apps or other online activity. Just as when a person introduces him or herself at the door with an oral statement of their name, these can be used to identify a person.

A formal digital identification system, however, involves a carefully designed set of databases, processes, technology, credentials, and legal frameworks to capture, manage, and use specified personal data for the purpose of identification. It may also include or be used for authentication (e.g., to confirm he or she is the person to whom a credential was issued), authorization (e.g., to access a service) or attribution (e.g., to bind a person with a legal commitment). In a traditional, formal sense, a digital identity is something issued to an individual – something he or she receives from an organization.

Privately-operated digital identification systems include highly secure systems made available to consumers by banking, health and telecommunications service providers and other product and service vendors and to employees by their employers. In each case, these are identities that are issued to the individual by a private identity provider. Some of these rely on extensive know-your-customer (KYC) due diligence and others merely on providing a basic user name and password or a small number of specific attributes.

[2] https://www.euractiv.com/section/competition/interview/vestager-id-like-a-facebook-that-i-pay-with-full-privacy/

State-operated identification systems are also employed on a national or regional basis by governments for both foundational and functional purposes. Foundational identification systems provide general identification and credentials to the population for public administration and a wide variety of public and private sector transactions and services, or to establish and provide recognition and proof of citizenship and/or residency status. By contrast, functional identification systems manage the identity lifecycle for a particular service or transaction (e.g., voting, tax administration, social programs and transfers, financial services). State-operated identification systems have traditionally issued non-digital credentials such as national identity cards, employment cards and driving licenses. These credentials are increasingly integrated with digital systems, including a digital element such as a unique identifier number or chip on a card which may be used for electronic access to related services.

All such systems present data governance issues. These relate, as indicated earlier, to data security and privacy and personal data protection measures required to ensure trust in the system. Important scheme design and governance controls are required to prevent access to data about persons' use of identification systems in their daily lives that would enable development of a profile that may be used adversely against them (e.g., for political, religious, ethnic or other form of discrimination). This involves preventing retention of transaction data beyond what is necessary for the scheme's operations, and administrative controls on who has access to personal data.

In addition, a digital identification system depends on an effective 'trust framework.' This combines generally applicable laws of contract and liability, data-specific laws and standards, and identification scheme rules and protocols. A trust framework that works ensures that rights and duties of participants of the scheme are clear and ensures that there are sufficient economic incentives for them to play their respective roles, such as scheme operator, enrolment agent, consumer or authentication agent.

Other issues important to the design and operation of such schemes include non-discriminatory inclusion of the population at large, interoperability with other public agencies' and private firms' systems to leverage digital identification for multiple functional purposes, the use of open standards, and ensuring competition among scheme vendors (avoiding lock-in).

As people travel, trade and communicate across borders, there are increasingly movements to allow recognize of digital identification systems across borders. Europe's eIDAS regulation sets out common standards for mutual recognition of other EU countries' systems.

Development of such mutual recognition standards in the SADC area and across the continent may be a valuable part of the development of regional markets in digital services in which South Africa could be a leader. Opportunities to deploy its expertise in digital identification security in other countries will depend not merely on South African companies producing innovative products, but on South Africa maintaining an ecosystem of standards and protocols that frame and support the development of such products, and strengthen the reputation for trusted systems.

V. Access to data, markets and platforms

The use of data for social and economic good depends on access to it. Data may be shared and gathered in the context of a transaction or other form of interaction, by monitoring of behaviour, or by acquisition from third parties. The ability to transfer data across national and organizational boundaries is vital to its potential exploitation.

Data governance is directly relevant to how data is made available, the ability of those who can use data to access it as well as services that depend on it within and beyond borders, as well as the dynamics of data-driven markets. The purpose is not to explore regulatory frameworks relevant to use of data in the numerous various vertical sectors in which data is used, but to focus on key data governance issues that are generically applicable.

A. Opening up data

1. Proprietary data

Access to proprietary data of private organizations may have immense economic and social opportunity. For instance, access to call detail records (CDRs) of mobile telecom operators, and geolocation data about customers of mobile telecom operators and ride sharing applications, can offer vital information for the placement of health clinics, traffic management and public transport policy.

A variety of institutional forms may be used to exploit the opportunity of proprietary data developed by organizations. Under one long standing mechanism, banks may be obligated to share data about credit performance of their customers with regulated credit reporting agencies, which make the information available to other commercial actors.

Private healthcare providers and pharmaceutical companies will hold information that may be extremely valuable for medical research purposes. These may be encouraged to combine their data in a manner that allows collaboration on secure large data pools, while not resulting in anticompetitive coordination. Such initiatives already exist, and can be encouraged in South Africa. They may require exemptions from the competition authority for horizontal coordination (i.e., collaboration among competitors), which may have to work alongside the information regulator and sector regulator to ensure the initiative achieves its purposes while mitigating risks.

Interoperability is required for much data sharing to be useful at all. This may be the subject of ‘soft’ rules, such as industry standards, but these can be provided for in regulatory frameworks that convene sector participants in fair, reasonable and non-discriminatory manner.

Reaching into proprietary data generated and held by private companies and requiring it to be made available presents opportunities and challenges. It may be viewed as a form of expropriation of commercial asset developed by the firm, and potentially weakening competition by undermining first mover advantages. However, the benefit to society at large might sometimes outweigh these considerations, particularly in key areas of health for example, or development of public policy.

Often, data is a mix of personal data about identifiable individuals and proprietary data of the organization. So, for example, a firm’s accounting, ordering and sales data may also include information about customers. Untangling the application of obligations to protect personal data yet make commercially relevant information available is challenging.

Effective policy on arranging access to such data will not be achieved by a mere obligation to make data available. Complex requirements on protocols and standards are required for data to be useful. This involves labour intensive IT work to establish, and so costs on the organization. A trusted vision is thus important, and so the robustness of the mechanisms and their governance is vital. Several structures can be envisaged, as illustrated in Figure 4.

Figure 4 Mechanisms for data sharing

Approach	Distinguishing feature
Data trusts	Takes what has been learned from the use of legal trusts. Trustees of a data trust will take on responsibility (with some liabilities) to steward data for an agreed purpose.
Data cooperatives	Takes what has been learned from cooperatives. A mutual organisation owned and democratically controlled by members, who delegate control over data about them.
Data commons	Takes what has been learned from managing common pool resources – such as forests and fisheries – and applies the principles to data.
Personal data stores	Stores data provided by a single individual on their behalf and provides access to that data to third-parties when directed to by the individual.
Research partnerships	When data holders provide access to data to universities and other research organisations.

Source: Open Data Institute

One example of a mechanism that might be explored is the data trust. A data trust is a legal structure that provides independent third-party stewardship of data for the defined purpose. This can be established in a manner whereby the type of data and its governance, including conditions on which others can access it, are set out and agreed between the provider of the data and the agency seeking to ensure its exploitation for wider social benefit. Data trusts depend on having a clear purpose, a legal structure, constitution and trustees, rights and duties over stewarded data, defined decision-making processes, rules on how benefits are shared, and sustainable funding. Pilot data trusts have been used in the UK in collaboration with the UK Open Data Institute.

The benefit of this model is that it can allow collaboration and sharing of data for public good in a trusted manner, whether among private entities, between public sector and private sector, and across-borders. While any one data trust will require significant transaction costs to establish the institutional mechanisms, the marginal cost of reusing the model will decline as legal, financial and data experts build their understanding.

Data co-operatives also offer opportunities, under which people could contribute data towards collective social benefit. Medical data trusts might enable patients with specific health conditions to contribute their health records so that it is available for medical research. Imbalances of market power between organizations and workers in the gig economy can be redressed if the latter share data about their contracting and other working conditions. A data trust could collect data within a city with a view to improved transport, energy distribution, waste collection, or other benefits to citizens.

For these mechanisms to be effective, it will be necessary to build trust in them, ensuring that they are not somehow exploitative of those who contribute data, do not facilitate anticompetitive conduct or build significant market power through accumulating valuable data over which the mechanism maintains exclusive control. This will require city authorities or vertical ministries to take a lead, alongside information regulators, competition authorities and private entities involved. Consumer organisations may be able to contribute support that also builds trust among potential participants where these are individuals. They might monitor performance, and formal auditors may be required to provide reports assessing conduct against pre-agreed criteria.

Such mechanisms can be used to bridge trust gaps between private sector (which may hold the valuable data) and the public sector (which wants to make the data accessible for research or policy making). They can also be used across borders. So, for instance, where data might need to be removed from South Africa in order to be aggregated with data from another country for machine learning or other data analytics purposes, a data trust can be used to ensure governance rules that satisfy the data protection requirements of POPI Act in South Africa and other countries involved.

2. Open public data

Making data held by Government and other public institutions available offers important opportunities for medical, climate change and other scientific research, improved organization and regulation of public and private services, and development of new digital applications and services. Access to data, whether on an open basis or pursuant to freedom of information request laws, can also improve the transparency and accountability of government, and increase public participation.

Open data policies seek to make data freely available, usable and redistributable, subject to certain limited conditions of use and attribution. It may comprise data with commercial value or potential to be aggregated or processed to become valuable.

Many governments, including developing countries, have committed through the Open Government Partnership (OGP) and Open Data Charter to make their data open, subject to data governance controls. Effective use of such data depends on provision of readily accessible portals, machine readability, sufficient structure to enable comparability and aggregation for statistical purposes, interoperability with other systems to enable analytics.

There are tensions between the social and economic utility of access to personal data on one hand and the concerns of data security and personal data protection and privacy (each discussed above) on the other. Such tensions shift with populations' sensibilities to such issues, but also with technological means of exploiting data while reducing the risks. For instance, anonymization techniques may make it feasible technologically to use location and driving data for intelligent transport systems, or medical data for research without compromising privacy.

B. Concentrated data, competition, data portability and access

1. Data concentration and competition

There is increasingly widespread concern that competition is insufficient to deliver public goods from digital services, particularly in a platform economy. The availability of data has led to the rapid emergence of important new digital platforms. These include digital matchmakers, transaction platforms and multi-sided markets. Examples arise in vertical markets such as retail sales (Amazon), transport (Uber and Lyft), hospitality (Airbnb), payments (M-Pesa) and digital credit (M-Shwari), among many others. In data-centric areas, they include search (Google, Baidu) and social networking (Facebook family of companies, WeChat).

Many digital platforms generate externalities through network effects, benefitting consumers as they expand. Data reduces some information asymmetries by allowing firms to know their consumers better, and to assess their needs and risk better, and so to tailor products, services and prices. Yet data may increase information asymmetry between firms able to extract systemic information from large datasets and consumers for whom such data is mere noise, reducing consumers' bargaining power. It may also increase information asymmetry between

successful platforms and firms without access to such scale of data, algorithms and processing power. The same network effects may lead more to competition *for* market ('winner take most') than competition *in* the market.

Moreover, a firm may aggregate data from customer activities on the platform, generate inferences that together with the input data are excluded from rivals, and consolidate its market lead through ever better-tailored matchmaking services (e.g., improved design, recommendations, search results, and targeted advertising). Where this is used to foreclose rivals, competition concerns may arise about market failure in data-driven sectors.

Competition interventions have been aligned with traditional privacy, which has long emphasised control over one's personal data as a fundamental tenet (Westin 1962). This has been applied in law, for example, Europe's GDPR, codes of conduct and norms like the Fair Information Practices (FIPs), and in commercial offers such as where Google and Apple allow the user to control location tracking. This confluence of competition and privacy laws is being tested in Germany, where in 2019 the German *Bundeskartellamt* ordered Facebook to de-link data from the Facebook, Instagram and browsing history data on the basis that the aggregation of the data without effective consumer consent amounted to an abuse of dominance that violated the competition laws.

After recent years of academic investigation (e.g., Stucke & Grunes, 2016; Khan 2016 & 2019), regulatory policy makers are today vigorously re-examining their competition laws and enforcement practices as they relate to big data and artificial intelligence (CMA 2019; ACCC 2018; Bundeskartellamt 2019). Much of the concern about market power arises from aggregation of data through vertical and horizontal consolidation, and leverage of market power from one market to another. While data is in theory non-rivalrous, competition has often involved excluding data from rivals. Fear of creative destruction (Schumpeter 1942) may result in inefficiency from hoarding data. In some sectors, this has been identified for specific intervention, such as by imposing open banking requirements (as in the UK, for example).

These are not just first world problems, but are already noticeable in markets reliant on customer data and featuring network effects, such as digital financial services.

2. Data portability and access

Data portability has been proposed as a solution for competition problems arising from concentrated data (APC 2017; Crémer 2019; Furman 2019; GDPR). This may be more feasible in the case of some vertical sectors, such as open banking (CMA 2017; EU PSD2 2018). It can reduce switching costs by enabling the consumer to make relevant data available to an alternative service provider that depends on such data for example to assess the customer's creditworthiness.

However, data is often unstructured, or structured in different ways in different organizations. Its relevance depends on the purpose to which it is put, and what other data it is linked to. There may be benefits to applying portability measures sector-by-sector, as the Australian Productivity Commission recommended when introducing the 'consumer right to data.' The Australian Government has begun with banking and telecommunications services before moving on to others. The UK Open Banking Order of the Competition and Markets Authority (CMA 2017) and the EU's 2nd Payment Services Directive (PSD2) depend on developing sector specific Application Programming Interfaces (APIs), data structures and security architectures to enable data portability (Open Banking Order 2017).

Portability is a potentially intrusive and heavy remedy, and needs to be deployed where the benefits are likely to exceed the costs, both financial and administrative. In some cases, data portability may not remedy the perceived harm, for instance if switching costs are less driven by concentration of data but more by network effects.

Facebook's market power in online behavioural advertising is based on profiling from large amounts of personal data. However, it appears to be the network effect – whereby consumers need to be on Facebook or Instagram to communicate with friends and family on the social network – that creates the market failure more than the data it holds.

Similarly, online ecommerce platforms Amazon.com and South Africa's Takealot.com may secure their market power through cross-side network effect of bringing together more buyers with more sellers and extraordinary innovation in delivery logistics than holding large amounts of data. The data may enable them to strengthen market power further by better predicting customer interest, or even anticipating the market responses of downstream competitors. However, portability of the data may not be a particularly effective remedy compared with prohibitions on platforms acting with conflict of interest.

On the supply side, a number of proposals involve providing rivals with access to data of incumbents to neutralise their data advantage and allow competition directly on the basis of algorithms (Argenton & Prüfer 2012[3]; Furman 2019). This might rely on the essential facility doctrine of competition law – that it's an essential input to a competitor's service. It will be important to examine whether this should be used to enable competition *in* the market, or to encourage innovation and competition *for* new markets (Graef 2016).

The remedy of access to data faces some of the same challenges as data portability, and it will be important to examine again whether it is best introduced on a sector-by-sector basis relying on open standards developed and agreed amongst industry members.

South Africa will need to be ready to deal with these sorts of issues as platform economy grows, and is already confronting them in the areas of healthcare, financial services and ecommerce.

C. Cross-border data flows – data sovereignty, ownership and localization

As mentioned above, one of the features of data is the increasingly low cost of moving it, and this mobility creates huge efficiency and innovation opportunities. Data from disparate geographical locations and population groups can be aggregated in optimal locations, enabling firms to manage customer identities, deliver services and carry out analytics on a regional or global scale. Linked data centres can be located in jurisdictions with adequate human resources, cheap power and high-speed connectivity, enabling a wide range of cloud services and caching of content for distribution.

Access to such services across borders offers huge opportunities, particularly for export to countries whose domestic tech industries may take time to develop such services. Cross-border trade in goods in the physical economy depends on cross-border flows of information to communicate demand and ability to supply, and to manage logistics and process of transport and delivery.

[3] Argenton & Prüfer proposed access to search query data to allow competition only on basis of algorithms.

The cross-border dimension presents questions for policy makers and regulators. Legal powers can be enforced only by institutions within their borders, through bilateral or multilateral cooperation (e.g., mutual recognition), or by treaty arrangements with foreign national or international bodies. Such limits are leading to an increased emphasis on data sovereignty, data ownership and data localization.

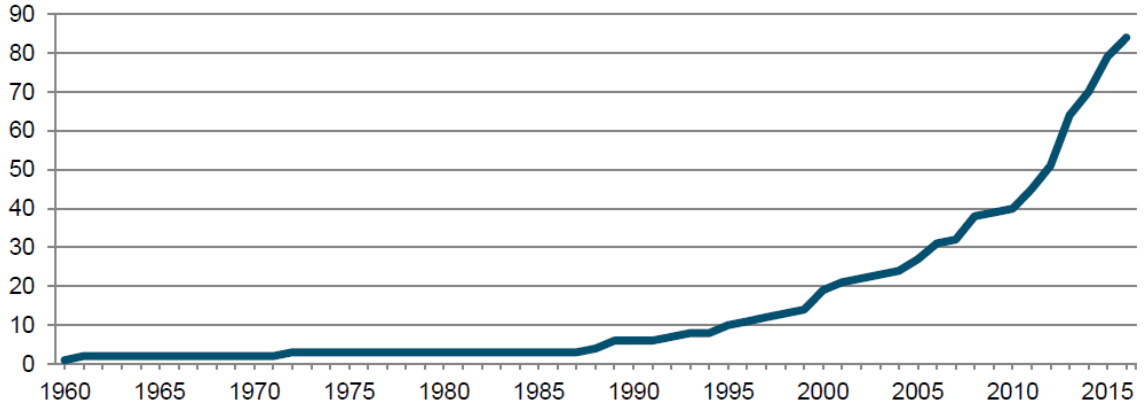
Data sovereignty seeks to apply laws and governance structures to data within the jurisdiction in which it is collected or held, even where the data is held abroad. For example, the GDPR asserts extraterritorial jurisdiction, applying to any data controller or processor whose data subjects are EU citizens, regardless of whether the data is held or processed within the EU. The US Department of Justice sought to require Microsoft to divulge data held on Hotmail accounts in Ireland that could be used to enforce against illegal trade in drugs.

Data ownership is also sometimes used as a metaphor to emphasise control over our data, and in some cases such as India's recently proposed ecommerce policy which refers to 'taking back control' over data, to emphasise that data is a national resource and should be protected as such by preventing it from being transferred out of the country (India 2019). Requirements to hold data on servers within the country are more commonly and better referred to as 'data localisation,' and where the rules aim to preserve value within the economy, it might be referred to as 'data nationalism.'

Some consider the extraction of data – particularly what Zuboff has called 'behavioural surplus' (which takes data beyond what is needed to develop and provide the service) – by the big tech companies to be reminiscent of a pattern of natural resource extraction in Africa under and since colonisation (Zuboff 2019). Recent criticism of Nigeria's online market Jumia as being more European than African shows how fired up this debate can get (Pilling, 2019).

Data localization concerns legal requirements to keep data within the geographic boundaries of the jurisdiction in which its subjects are located, or to ensure that certain protections are assured if data is transferred abroad. Concerns about surveillance by foreign governments of data passing over systems to which they have access, as well as weak or absent data protection regimes in many countries, led to a sharp increase in such laws (see Figure 5). Indeed, any country should have a view on how it should control transfer of data that could have national security implications (Scassa, 2018).

Figure 5 Number of data localization measures globally



Source: USITC 2017, ECIPE Digital Trade Estimates database

Data sovereignty and localization can sometimes be seen as an extension of the policies that underly domestic personal data protection and privacy (discussed above), sometimes protecting types of information that are viewed as particularly sensitive, such as health data (e.g., Australia), telecommunications metadata (e.g., Germany), payment systems data (e.g., India), and in some cases even all personal data (e.g., Canada and Russia).

They may also represent a form of digital mercantilism, whereby countries hope to force high-tech activity to be carried out within their borders and to promote local processing operations.

Some level of data localisation may be necessary to ensure that data that is particularly sensitive for national security or personal reasons is protected by the domestic laws, but it will be vital to examine whether excessive requirements to keep data within the country may undermine the efficiency and innovation opportunities of big data, the cross-border provision of cloud services, customer relationship management and regional and global value chains.

Data localization can be counterproductive.

It can impede the development of cross-border commercial activity and public sector initiatives at regional levels. Cross-border electronic commerce and mutual recognition of digital identification may be held back. The inability to aggregate data in data centres in advantageous locations to serve multiple countries may result in increased costs, and lost efficiencies and lost opportunities to provide services.

Numerous studies seeking to put value on the cost of data localization suggest it is significant (US International Trade Commission 2014; Leviathan Security Group 2015; Center for International Governance Innovation 2016; European Center for International Political Economy 2014). For example, one study suggests that data localization would have increased the cost of cloud computing services by over 50% in the case of Brazil if it had gone ahead with proposed restrictions measures (O'Connor 2015). Data localization may become a tool for protectionism if it effectively prevents foreign providers from offering services in a country, and so is today a key component of trade policy.

Countries such as the US, EU, Australia and Canada are increasingly using e-commerce and digital trade chapters of free trade agreement negotiations to roll back data localization rather than effectively working through the World Trade Organization (WTO). The US has included clauses on free flow of data in the Comprehensive and Progressive Agreement for Trans-Pacific Partnership (CPTPP) and the United States-Mexico-Canada Agreement (USMCA), making free flow of data across borders a default. These agreements commit signatories to establish online privacy protection, ban data localization requirements and prevent countries from requiring source code as a condition of doing business. The EU agreements permit free flow of data but on the basis of adequate data protection. China restricts flow of data across its borders as part of the country's wider controls on data.

However, trade negotiations are conducted at bilateral and plurilateral levels, and are not producing universal or interoperable rules governing the use of data. Challenges inherent in the nature of data relate to determining what is an import or an export, when data is subject to domestic law, and when "trade" as such is even occurring (Aaronson 2018). The difference of approach among the US, EU and China, the leading jurisdictions, results in a patchwork approach posing inconsistent options for countries negotiating trade with them. FTAs have different definitions, different scope and objectives, define barriers to trade differently and have different requirements for domestic legal frameworks (Monteiro & Teh, 2017).

A key goal must be to enable the use of data for productive economic goals, and this depends on aggregating data with other data outside the country where there is processing capability. As a result, it is important not to be seduced by data nationalism arguments that claim that the country's data should be kept in the country and used only for the country's benefit. A restrictive approach risks the country never exploiting the opportunity that data presents.

South Africa's POPI[4] restricts transfer of data outside the country except where transferred to a recipient that is subject to a law, binding corporate rules or binding agreement that provides substantially similar protections to those provided by POPI. However, transfers may occur if the individual consents or the transfer is necessary in connection with a contract, or is for the benefit of the individual. The overriding impact of consent weakens these protections.

These provisions are reasonably aligned with jurisdictions like the EU. Development of standard binding corporate rules and standard contractual clauses for use by organizations seeking to transfer data in and out of South Africa would be a helpful development. South Africa has not imposed a requirement like the EU and many other countries that require an authority to confirm in advance which countries have laws providing data protection adequate to qualify for transfers.

Some studies suggest that strict restrictions on cross-border data flows negatively and significantly impact trade in data-intense services, and that this negative impact is stronger for countries with better developed digital networks (Ferracane, 2019). This is potentially significant for South Africa, which may have an opportunity for export of digital services and data processing, where the rest of the continent badly lags behind. On the same continent, Africa reportedly has less than 1% of the World's total available data processing centre capacity while it contains 17% of the population (Munshi, 2020). But there is no reason to restrict the vision to Africa if South Africa can offer lower costs to Europe and beyond, while promising a robust, professional, secure data environment.

Firms like Teraco Data Environments can provide data centre capacity to other countries relying on submarine cable fibre to transport data into and out of the data centres. Amazon Web Services' plan to open a cluster of data centers in the Cape Town area promises a powerful boost to South Africa's and indeed the continent's processing capability. The capacity of the submarine cables that circle Africa should place South Africa well to compete with other countries like Nigeria, which are also rapidly building up data centre capacity.

Overall, South Africa appears likely to benefit from a relatively liberal regime for cross-border transfers, one where its restrictive policies do not provoke reciprocal data localization rules that would result in barriers to South African services, yet one in which it can tout its credibility as a trusted location for secure data processing and storage. This would be supported by greater multilateral efforts to find common ground among countries on data regulation and common rules for trade in e-commerce and digital services.

VI. Inferences, policy inputs and algorithmic accountability

Lack of data has historically been a major barrier to access to services, such as financial services, for many lower income population members. Data about a person's borrowing history, scale and sources of income, ownership of assets, or family and broader social network, has not been readily available. Identification data about people has not been

[4] Section 72.

sufficient to verify that they are who they say they are. As a result, traditional lenders and insurers have not developed attractive business cases for broadening lending, insurance and other risk-based financial services to such populations.

Today, a vast number of attributes are widely available rather than those merely submitted by an individual upon presenting himself or herself for a transaction or other interaction. A person's 'data exhaust' from interactions with and transactions over electronic networks includes extensive data points that may be collected, analysed and shared, and used to identify and enable inferences about us, whether using our own self-assertions or through assessments of third-party algorithms.

Governments and businesses can use such data to build a detailed personal profile of an individual and their behaviour (preferences, activities and movements) which may be used for commercial offers, State and private surveillance. They may be used to identify an individual and to determine their eligibility for a service or product. For example, inferences drawn by algorithms using machine learning techniques over large data sets are reducing information asymmetries between financial service providers and consumers. The former can assess risk and offer financial services not previously offered, increasing financial inclusion.

Financial services are only one of numerous areas in which artificial intelligence is extracting value from data through making inferences about people and their preferences and habits, often with revealing levels of accuracy (Stephens-Davidowitz, 2017). Use cases offering substantial public benefit include healthcare provision, medical research, transport, education, advertising, policing and the justice system.

However, use of algorithms to make decisions based on these datasets presents a new set of risks.

Big datasets drawn from structured and unstructured data gathered from multiple direct and indirect sources over time risk being inaccurate or out of date. Whether used as training data for algorithms that generate automated decisions about people or as inputs about the subjects of such decisions, inaccuracies may lead to erroneous inferences and decisions.

Algorithms trained on data from past experience may reflect and perpetuate the biases embedded in historical treatment of ethnic, religious or gender groups. Even where special or protected categories of data about a person (such as ethnicity, religion or gender) are specifically avoided, such data may often be inferred from non-special data. Religion might be inferred from a name, a medical condition from purchase history at a pharmacy, or ethnic group from a postcode.

In addition to these risks, at the same time as they facilitate access to financial services, new technologies pose challenges to privacy and autonomy. The installation of starter interrupter devices (SIDs) on cars for subprime vehicle lending strengthens lenders' repossession rights in case of default. However, SIDs also give lenders unprecedented amounts of information about borrowers' location and activities (Elvy, 2020). Consumer benefits of access to services and innovative goods need to be considered alongside concerns raised about what some term surveillance capitalism (Zuboff, 2019) and power over consumers' daily use of household equipment in the internet of things (IoT) era.

Initiatives are underway in many countries from several angles to address such problems. These include legislating for a right to opt out of automated decisions made entirely through

data processing, to receive an explanation for automated decisions, or the right to appeal to a human, as in the GDPR and some countries' national data protection legislation (e.g., Brazil). They also include efforts to develop standards and ethical frameworks for the use of artificial intelligence (IEEE).

These efforts will need to be applied beyond algorithms to their application in the physical world as self-driving cars, robots and other machines increasingly interface with humans. Data security will be vital to avoid malicious and accidental harm, greater amounts of personal data will be collected and need protection, and automated decisions are taken with ethical stakes beyond the trolley problem.

The use of algorithms may also include potential competition concerns, whether by facilitating the implementing, monitoring, and policing of cartels, or reducing competition through industry-wide adoption of predictable reactions to changing market conditions (Stucke & Ezrachi, 2016).

South Africa's POPI Act[5] provides that a person may only be subject to automated processing under certain conditions if it "results in a decision with legal consequences" or affects him or her "to a substantial degree" based solely on automated processing of information for profiling purposes. Conditions that make such processing acceptable include where it is carried out for the purpose of a contract or there is a law or code of conduct in place that provides information to the individual and allows him or her to make representations.

At this time, these provisions are reasonably aligned with other jurisdictions' efforts to introduce some controls on use of AI, machine learning, profiling and automated decision-making, but they are generally likely to be very inadequate over time. They amount to a classic notice model without even really relying on consent rather than building in substantive regulation of how these technologies are used in a manner that mitigates their risks. Further efforts will be needed to determine how much further law and regulation should stretch into setting norms rather than leaving such steps to industry bodies, standards and ethics.

Various ideas are being developed and could be considered for introduction in South Africa's legal and regulatory framework:

A. Transparency

Governance depends on accountability, and accountability often requires an explanation for the basis and method of the decision (Doshi-Velez et al., 2017). The EU and some other jurisdictions have established a consumer right to an explanation where a solely automated decision has legal or other significant effects (e.g., a declined loan application or reduction in a credit limit).

Brazil's Data Protection Act 2018 provides the consumer with the right to request a review of automated processing decisions designed to define his profile or evaluate aspects of his personality, and the right to request clear and relevant information on the criteria and procedures used for the automated decision. South Africa's POPI Act does require providers to provide sufficient information about the underlying logic of the automated processing to enable him or her to make representations about the decision.

[5] Section 71.

However, this still faces a significant challenge of explaining “opaque” machine learning models that appear to be “black boxes” makes it difficult to produce useful explanations (Pasquale, 2015). And the more accurate a system is intended to be, the more detailed its levels and so the more difficult to explain it will be, i.e., there may be a trade-off between explainability and accuracy. Additionally, machine learning models that produce automated decisions may be the subject of trade secrets and software copyright resulting from investment in a competitive market.

Use of counterfactuals have been suggested by some in order to provide the information that a consumer really wants, which is how to get a better decision from the system (Wachter, Mittelstadt & Russel, 2018). Other approaches would require documentation of how the machine learning model was chosen, providing a legal and technical analysis to support this, including identifying the trade-offs made between explainability and accuracy.

B. Algorithmic bias

One of the reasons data protection laws impose stricter conditions on processing of sensitive data – relating to race, religion, ethnicity among other things – is to reduce risk of discriminatory outcomes resulting from using such data. However, as explained above, sensitive data can be inferred from non-sensitive data.

There is widespread concern that automated decision making may systematize and conceal these and other forms of discrimination. A country with South Africa’s diversity and inequalities among the population will need to develop means of preventing unfair discrimination generated in machine learning systems. This will involve working out how existing laws should apply to such technologies, and potentially regulating to mitigate the risk of discriminatory outcomes. This might be through a strict liability for disparate impact of decisions, or requiring organizations to introduce technical measures that reduce discrimination.

C. Rights and mechanisms to contest decisions

Although laws do not typically provide a general right to contest the accuracy of decisions, data protection laws are increasingly entitling consumers to contest decisions made on the basis solely of automated processing. Automated decision-making introduces risks in important areas from financial services (credit, insurance and risky financial products) to prison sentencing. There is increasingly widespread consensus that people should have recourse to human appeal where an AI system may not appreciate extenuating circumstances (IEEE, 2019). Increasingly, data protection laws, including the GDPR[6], provide the right to obtain human intervention, express one’s views and contest the decision. South Africa’s POPI Act only confers a right to make representations in the case of automated decision-making.

D. Ethics for AI

The line between application of law with the force of the State and the development of ethics and standards in this area is a challenging one to set. In many cases, no approach is sufficiently satisfactory to create financial and other liability, and in these cases, it may be particularly important to develop ethics that will be integrated into the full product lifecycle, from development to sale.

[6] GDPR, Article 22(3).

Numerous organizations have developed ethical frameworks, including the Association for Computing Machinery (ACM)[7] and the Institute of Electrical and Electronic Engineers (IEEE),[8] as well as Partnership on AI,[9] Software & Information Industry Association (SIIA),[10] and companies such as Google[11] and Microsoft.[12] These are accompanied by work by organizations such as Fairness, Accountability, and Transparency in Machine Learning (FAT/ML),[13] Privacy International,[14] the Future of Life Institute,[15] Center for Democracy & Technology (CDT),[16] and the Leadership Conference.[17]

In some countries, authorities have developed principles to apply to AI, such as the Monetary Authority of Singapore's *Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector*.

Common threads of such principles include providing for internal authorities within organizations to approve decisions, recording design choices, accountability for models, proactive awareness raising including to board level, channels for inquiries, appeals and reviews.

In important areas of the economy, including data-driven sectors where South Africa stands to gain from exporting services, the engagement with these sorts of principles by the authorities and industry bodies, as well as with transparency, algorithmic bias and mechanisms for contesting automated decisions, is likely to bolster trust and reputation. This may lead to growth in digital services, and demand from abroad.

---

[7] When computers decide: European Recommendations on Machine-Learned Automated Decision Making, Informatics Europe & EUACM 2018. http://www.informatics-europe.org/news/435-ethics_adm.html.

[8] IEEE, Ethically Aligned Design: A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems version 2 (2018), https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf..

[9] *Tenets*, Partnership on AI, *available at* https://www.partnershiponai.org/tenets/.

[10] *Ethical Principles for Artificial Intelligence and Data Analytics*, Software & Information Industry Association (Sep. 15, 2017), *available at* http://www.siia.net/Portals/0/pdf/Policy/Ethical%20Principles%20for%20Artificial%20Intelligence%20and%20Data%20Analytics%20SIIA%20Issue%20Brief.pdf?ver=2017-11-06-160346-990.

[11] *AI at Google: our principles*, Google (June 7, 2018), *available at* https://www.blog.google/technology/ai/ai-principles/. Also, Google, PERSPECTIVES ON ISSUES IN AI GOVERNANCE, *available at* https://ai.google/static/documents/perspectives-on-issues-in-ai-governance.pdf.

[12] *Our approach*, Microsoft, *available at* https://www.microsoft.com/en-us/ai/our-approach-to-ai.

[13] *Principles for Accountable Algorithms and a Social Impact Statement for Algorithms*, Fairness, Accountability, and Transparency in Machine Learning, *available at* https://www.fatml.org/resources/principles-for-accountable-algorithms.

[14] Privacy International, PRIVACY AND FREEDOM OF EXPRESSION IN THE AGE OF ARTIFICIAL INTELLIGENCE (2018), *available at* https://privacyinternational.org/sites/default/files/2018-04/Privacy%20and%20Freedom%20of%20Expression%20%20In%20the%20Age%20of%20Artificial%20Intelligence.pdf

[15] *Asilomar AI Principles*, Future of Life Institute (2017), *available at* https://futureoflife.org/ai-principles/.

[16] *Digital Decisions*, Center for Democracy & Technology, *available at* https://cdt.org/issue/privacy-data/digital-decisions/.

[17] *Civil Rights Principles for the Era of Big Data*, The Leadership Conference (Feb. 27, 2014), *available at* https://civilrights.org/civil-rights-principles-era-big-data/.

VII. Conclusions

The growth of data generated, amassed, used, reused, stored and transferred is increasing at high rates. It is set to increase with more people coming online, advent of 5G communications and proliferation of IoT-enabled devices. Data governance is not only important to protect the population but is strategically central to economic success and social cohesion in the future – near and far.

South Africa has some strong measures in place, including a relatively well developed data protection law in the POPI Act, even if it is not yet properly in force. But there are numerous steps that it should be considering taking in order to capture the opportunity presented by data. The protections are prerequisites to building the trust necessary to assure the growth potential, and such trust will only enhance the South African brand as the country pursues the realistic opportunity to be a regional and even international hub for data-centric services.

Various data protection measures could be considered to bolster the POPI Act, including embracing privacy by design and default (rather than allowing privacy to be consented away), required use of data protection impact assessments, adoption of substantive privacy requirements based on the reasonable expectations of consumers and requirement that data be used for a legitimate purpose that benefits the individual concerned. Novel mechanisms can be tried, such as legal frameworks for information fiduciary responsibility, encouragement of data intermediaries, and personal data management service providers.

Steps to open up data for public policy making and other public goods should also be pursued. This concerns not only public data but also proprietary and mixed data held by private companies, can also be taken but with great care. Data trusts are one mechanism that might be used to ensure confidence necessary to bridge distrust between public and private sector stakeholders, as well as to provide secure and trusted means for cross-border data sharing.

Addressing the risk of foreclosure of data-driven markets through accumulation of market power needs to be approached with caution to ensure that remedies will address the means by which such market power is protected. Heavy remedies such as data portability need to be deployed strategically and likely on a sector-by-sector basis.

Openness to cross-border data flows is also likely to benefit a country that can offer data processing capability in well-connected data centres to an underserved continent, and to which it can export innovative digital services as well as internationally. South Africa appears unlikely to benefit from subjecting cross-border data flows to data localization restrictions except for strategically important cases and where national security issues arise.

Citations

Aaronson, S.A. (2018). *Data Is Different, Why the World Needs a New Approach to Governing Cross-border Data Flows*, CIGI Papers No. 197 – November 2018

Acquisti, A., Taylor, C. and Wagman, L. “The Economics of Privacy”, *Journal of Economic Literature* 2016, 54(2), 442–492

ACCC (2019), *Digital Platforms Inquiry Report*, available at https://www.accc.gov.au/focus-areas/inquiries-ongoing/digital-platforms-inquiry/final-report-executive-summary

Balkin, J. & Zittrain, J. (2016), "A Grand Bargain to Make Tech Companies Trustworthy," *The Atlantic*, 3 October 2020, at https://www.theatlantic.com/technology/archive/2016/10/information-fiduciary/502346/

Bambauer, D. "Privacy Versus Security," *Journal of Criminal Law & Criminology*, Vol. 103 (3) (2013), p. 669, at http://scholarlycommons.law.northwestern.edu/cgi/viewcontent.cgi?article=7454&context=jclc

Benkler, Y. (2007). *The Wealth of Networks: How Social Production Transforms Markets and Freedom*

Buchanan, J.M. (2003). What is Public Choice Theory. Vol. XLIII No. 5 *Economic Education Bulletin*, May 2003

Bundeskartellamt (2019), *Bundeskartellamt prohibits Facebook from combining user data from different sources*

Cavoukian, A. *Privacy by Design: The Seven Foundational Principles* (Information and Privacy Commissioner of Ontario, 2011), https://www.ipc.on.ca/wp-content/uploads/Resources/7foundationalprinciples.pdf

Center for International Governance Innovation (2016). *Estimating the Economic Impact of Data Regultaions*

CMA (2019), *Unlocking Digital Competition*, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/785547/unlocking_digital_competition_furman_review_web.pdf

Custers, B. (2016). "Click Here to Consent Forever: Expiry Dates for Informed Consent," *Big Data & Society*, January–June 2016: 1–6. http://journals.sagepub.com/doi/10.1177/2053951715624935.

Doshi-Velez, F. and others, "Accountability of AI Under the Law: The Role of Explanation" [2017] arXiv preprint arXiv:1711.01134

Edwards, J. (2019). "Click to Consent? Not Good Enough Anymore." Privacy Commissioner blog post, 2 September. https://privacy.org.nz/blog/click-to-consentnot-good-enough-anymore/

Elvy, S-A. (2020) Big Data in Consumer Lending Transactions, *The Internet of Things: The Future of Commercial Law & Privacy.* Cambridge University Press

European Center for International Political Economy (2014), *The Costs of Data Localization: Friendly Fire on Economic Recovery*

Ferracane M. and der Marel, E. (2019), *Do Data Policy Restrictions Inhibit Trade in Services?* European Centre for International Political Economy (ECIPE)

Herrera, S. and Haggin, P., New Apple Sign-In Option Could Keep More Personal Data Away From Facebook, Google, *Wall Street Journal*, 6 June 2019, https://www.wsj.com/articles/new-apple-sign-in-option-could-keep-more-personal-data-away-from-facebook-google-11559839438

Institute of Electrical and Electronic Engineers (IEEE) (2019). *Ethically Aligned Design*, https://ethicsinaction.ieee.org/

Jones, C. and Tonetti, C. (2018) *Non-rivalry and the Economics of Data*

Kerry, C. & Morris, J.B. (1999). *Why data ownership is the wrong approach to protecting privacy*, Brookings Institute, 26 June 2019

Khan, L.M. Amazon's Antitrust Paradox, *126 Yale L.J.* (2016).

Khan, L.M. The Separation of Platforms and Commerce. *119 Columbia L.R.* (2019)

Koponen, J. (2014) *We need algorithmic angels*, TechCrunch, https://techcrunch.com/2015/04/18/we-need-algorithmic-angels/

Lessig, L. (1999). *Code and Others Laws of Cyberspace*. Basic Books. See also Lessig, L. (2006) *Code v2.* Basic Books

Leviathan Security Group (2015), *The Costs of Cutting Access to Global Cloud Services*

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

Mayer-Schonberger, V. and Ramge, T. (2018) *Reinventing Capitalism in the Age of Big Data*. Hachette UK

Monteiro, J. & Teh, R. (2017) *Provisions on Electronic Commerce in Regional Trade Agreements*, WTO Working Paper ERSD-2017-11

Munshi, N., Africa's cloud computing boom creates data centre gold rush, Financial Times, 2 March 2020

National Institute of Standards and Technology (NIST) (April 30, 2019). *NIST Privacy Framework, An Enterprise Risk Management Tool* (discussion draft). Available at https://www.nist.gov/sites/default/files/documents/2019/04/30/nist-privacy-framework-discussion-draft.pdf

Nissenbaum, H. (2010). *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford University Press

Nubian Rights et al. v Attorney General of Kenya [add full citation]

O'Connor, B. (2015), *Quantifying the Cost of Forced Localization*, Leviathan Security Group, June 2015

Orcutt, M. (2017). "Personal AI Privacy Watchdog Could Help You Regain Control of Your Data," *MIT Technology Review, 11 May 2017, https://www.technologyreview.com/s/607830/personal-ai-privacy-watchdog-could-help-you-regain-control-of-your-data/, and the related Privacy Assistant mobile app, https://play.google.com/store/apps/details?id=edu.cmu.mcom.ppa&hl=en.*

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

Puttaswamy v Attorney General of India [add full citation]

Robinson v Attorney General of Jamaica [add full citation]

Romer, Paul M. (1990). "Endogenous Technological Change." Journal of Political Economy 98: S71–S102

Scassa, T. (2018). "Considerations for Canada's National Data Strategy." In *Data Governance in the Digital Age.* Waterloo, ON: CIGI. www.cigionline.org/sites/default/files/documents/Data%20Series%20Special%20Reportweb.pdf

Schumpeter, J. (1942), *Capitalism, Socialism and Democracy*

Sen, Amartya (1999), *Development as Freedom*

Stephens-Davidowitz, S. (2017) *Everybody Lies: Big Data, New Data, and What the Internet Can Tell Us About Who We Really Are*, Harper Collins

Stucke, M.E. and Ezrahi, A. (2016) *Virtual Competition: The Promise and Perils of the Algorithm-Driven Economy*, Harvard University Press

Stucke, M.E. and Grunes, A.P. (2016) *Big Data and Competition Policy*. Oxford University Press

US International Trade Commission (2014), *The Impact of Foreign Digital Trade Barriers on the U.S. Economy*

Wachter, S., Mittelstadt, B. & Russell, C. "Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR," *Harvard Journal of Law & Technology*, 2018. https://arxiv.org/abs/1711.00399

Westin, A. (1967) *Privacy and Freedom*. Scribner

Zuboff, S. (2019) *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. PublicAffairs