Data Policies: Regulatory Approaches for Data-driven Platforms in the UK and EU

Report as part of the project:

Policy Frameworks for Digital Platforms: Moving from Openness to Inclusion

Arne Hintz & Jess Brand

Data Justice Lab

Cardiff University

Executive Summary

This report summarizes the findings of the research project *Data Policies: Regulatory Approaches for Data-driven Platforms in the UK and EU* which members of the Data Justice Lab at Cardiff University, UK, conducted throughout the year 2018. It analyses the implications of current policy reform for data collection, analysis and sharing via platforms. In particular, it interrogates emerging regulatory frameworks that shape, constrain or advance citizens' control over data that concerns them and that affects their lives. The report focuses on the EU General Data Protection Regulation (GDPR), the UK Investigatory Powers (IP) Act and the UK Digital Economy (DE) Act but addresses broader regulatory trends that emerge from these. It draws from extensive document analysis and interviews with different stakeholders in the UK.

The report explores components, trends and prospects of a citizen-centric perspective on data policy. The GDPR, particularly, offers policy mechanisms, such as the right to data portability, access to data, the right to explanation of data-based decisions, the strengthening of consent requirements, restrictions to profiling and limitations to algorithmic decision-making, that enhance people's influence over data and protect them from data harms. Embedded in an emerging set of norms and discourses on data ethics and citizen control, such policies point to a growing recognition of the need for regulating datafication and incorporating citizen concerns.

However this trend clashes with the expansion and normalisation of data collection through laws such as the IP Act and DE Act. A stronger focus in policy debate on the uses of data – through data protection rules and data ethics norms – may move attention away from the risks of data collection and exacerbate harms and uncertainties connected to the persistent monitoring of citizens. Moreover, the centrality of rules that seek to empower the 'informed user' (e.g., the right to access and move data and to demand explanations, and the continued reliance on 'consent' mechanisms) disregards unequal power relations between citizens and platforms and overstates the existing capabilities of citizens in negotiating datafied environments. Further, the focus on 'personal data' in current data protection rules leaves out the variety of inferred and derived data that are getting ever more significant for the ways in which people are assessed, rated and categorized.

The report proposes a combination of different strategies and components to advance citizen-centric policy frameworks, including: the improvement of citizen rights and capabilities in exerting control over data; the protection of citizens from data harms, including limitations to data collection and data sharing; normative frameworks that focus on citizen control and civic rights; and the development of new policy concepts, such as models of collective as well as decentralised control over data.

Introduction

The datafication of social life has led to a profound transformation in how society is ordered, decisions are made, and citizens are governed. The emerging capacities in analysing 'big data' have generated vast new opportunities 'to extract new insights or create new forms of value' (Mayer-Schönberger & Cukier, 2013, p. 8). Datafication has become a defining feature of contemporary societies and political-economic systems which have been termed, amongst others, *datafied society* (e.g., Hintz et al., 2018) and *surveillance capitalism* (Zuboff, 2019). Data collection and analysis has allowed commercial and state institutions to predict and change human behaviour; to sort, categorize and assess citizens; and thus to significantly affect the roles of citizens and the protection and understanding of civic rights. The rules and norms that regulate the collection and use of data are therefore crucial cornerstones of emerging societal formations.

This has become a particularly prominent concern with the proliferation of social media platforms, cloud services and the so-called 'sharing economy' whose core business model is the collection, analysis and monetization of user data. Platforms are a 'data mine' (Andrejevic, 2012) where personal data is systematically extracted, processed, and combined with additional datasets in order to create detailed profiles of people that are valuable to the business sector. Consumers, but also citizens, are increasingly profiled, categorized and assessed according to this data (Lyon, 2015). The increasingly fundamental role of platforms for contemporary society has been conceptualized as *platform society* (van Dijk et al., 2018) and *platform capitalism* (Srnicek, 2016), yet platform businesses have largely operated in a policy vacuum, and many of their activities and social, political and economic consequences remain unregulated. This points to a particular need for policy development (Belli & Zingales, 2017).

This report reviews how data collection and analysis on platforms are regulated, and it investigates current trends and developments. In particular, it interrogates emerging regulatory frameworks that shape, constrain or advance citizens' control over data that concerns them and that affects their lives. In doing so, it serves to advance scholarly and public debate about these developments, and to point to areas of potential intervention to address citizen needs and concerns. The report, thus, both outlines and critically reviews components for citizen-centric data policies in the context of contemporary regulatory debate and policy reform.

We focus on a particular national and regional jurisdiction – the United Kingdom (UK), in the context of the European Union (EU) – to offer a perspective on an advanced economy where platforms play a significant role in social and economic life, yet where considerable debate has occurred on data collection and use. The UK has played a prominent role in recent controversies, from the Snowden revelations to the Cambridge Analytica / Facebook scandal and to newer debates on data ethics and data justice (Greenwald, 2015; Greenfield, 2018) while the EU has offered innovative approaches to data protection legislation. The UK/EU focus also provides insights into the contradictory aspects of policy development, with some laws enhancing, and others, restricting data collection, and into the interplay between national and regional policy. Despite the recent decision by the British government to leave the EU, the UK remains bound by European law and will do so for the foreseeable future.

The report summarizes the results of empirical research that was conducted throughout the year 2018 by members of the Data Justice Lab, a research unit situated in the School of Journalism, Media and Culture at Cardiff University, UK. The Data Justice Lab is dedicated to the study and practice of datafication from a social justice perspective, highlighting the politics of data processes from a range of different angles. The research included an analysis of policy documents, stakeholder statements and commentary regarding recent policy reform; semi-structured interviews with members of different stakeholder groups (government, business, civil society); and a multi-stakeholder fact-finding workshop.

The report begins by situating the topic in relevant academic and historical contexts, focusing on a critical assessment of the datafication of life and of the regulatory architecture. It then describes the empirical research that was conducted over the year of 2018 and the methods that were used to collect relevant data. The findings of the research address necessary components of a citizen-centric policy environment and evaluate their implementation in recent instances of policy reform.

Background & Context

Data

Datafication has transformed both the private and the public sector. In the private sector, data analysis has enabled new business models based on the processing of data about people, which has been hailed as a 'new industrial revolution' (Hellerstein, 2008). Data brokers and credit agencies develop increasingly sophisticated ways of rating, ranking and categorising consumers by aggregating different datasets (Dixon & Gellman, 2014). In many cases, these combine specific consumer data with a wider range of social and contextual data, aiming at the prediction of consumption patterns based on a variety of social, cultural, health and other information (McCann et al, 2018).

Government departments and state agencies in many countries now apply data analytics to inform policy and decision-making. Public services are increasingly allocated based on data analytics about claimants, leading to automated welfare eligibility systems and the use of predictive risk models in, for example, child protective services and the health sector (Eubanks, 2018). In education, data scores support personalized learning and individualized instruction of students, and assess teacher performance (Warrell, 2015; O'Neill, 2016). In criminal justice systems, risk assessment tools are used to produce 'risk scores' on defendants to estimate their likelihood of re-offending and thus determine sentencing (Angwin et al., 2016). In border control, data-driven profiling based on a cross-set of aggregated data is increasingly used for 'vetting' the 'threat' of migrants and refugees to society (Metcalfe & Dencik, 2019; Tucker, 2016). In the UK, data analytics have been used in, among others, predictive policing, criminal justice, housing, and child welfare (Dencik et al, 2018; Big Brother Watch, 2018; McIntyre & Pegg, 2018).

In the aftermath of the Snowden revelations on mass surveillance by intelligence and security agencies, including the British Government Communications Headquarters (GCHQ), the use of personal data for intelligence gathering and crime prevention has led to particular controversies regarding the interaction of state power and citizen rights (Greenwald, 2015; Lyon, 2015). As the revelations demonstrated, a large portion of the data used for surveillance purposes is generated through social media and other platforms and internet companies. Overall, the various ways in which data is now collected and analysed mean that citizens are increasingly, and in great detail, monitored, categorized, scored and assessed, and then treated according to an analysis of data that is gathered about them (Hintz et al., 2018).

The emerging academic field of *critical data studies* has interrogated both the premises and implications of the use of 'big data' and associated algorithmic processes. Scholars have questioned the supposedly value-neutral, impartial and objective character of data that, and have instead pointed out that data is always constructed based on the goals, interests and cultures of institutions and individuals (Kitchin, 2014). This also means that the representation of 'reality' by data and, more specifically, the relationship between people and the data that is collected about them is not self-evident (van Dijck, 2014). Data analytics may provide a reduced lens on society (Berry, 2011) and shape the reality they measure by focusing on specific objects, methods of knowing, and understandings of social life (boyd & Crawford,

2012; Cheney-Lippold, 2017). Rather than representing society, data may construct it – as Kitchin (2017, p. 25) notes, data 'are engines not cameras.'

Further, critics have highlighted the risks and implications of increased monitoring and surveillance of populations through data (Van Dijck, 2014; Lyon, 2015) and have analysed a wider range of harms, such as discrimination, that may be the result of using past patterns to predict future behaviour and occurrences (Gangadharan et al., 2015; Redden & Brand, 2018). They have raised concerns regarding the 'operative logic of preemption' (Massumi, 2015) inherent in data-based governance that challenges practices and understandings of the democratic process (Andrejevic, 2017) and focuses on managing the consequences, rather than seeking to understand underlying causes, of social ills.

The black boxed nature of big data processes, i.e. the lack of transparency about how and according to what criteria data about people is analysed, poses a significant problem for populations that are assessed by them and whose services are affected by them (Pasquale, 2015). Research has consistently highlighted a lack of public knowledge of data processes and demonstrated a public unease regarding pervasive data collection and analysis, yet coupled with a feeling of disempowerment due to a lack of understanding of the workings and consequences of datafication. This dynamic has been described as 'digital resignation' (Draper & Turow, 2017) and 'surveillance realism' (Dencik & Cable, 2017) and has raised significant questions regarding the agency of, supposedly, active and informed 'digital citizens' (Hintz et al., 2018).

Policy

This is significant as the idea of the 'informed user' has been at the heart of regulatory frameworks for datafication and, particularly, data extraction by platforms, where tentative interpretations of user consent have formed the core of what have largely been self-regulatory regimes. Platforms and apps are required to seek acceptance from users for the ways in which these companies track their browsing habits and use their data. For example, the EU Directive on Privacy and Electronic Communications from 2002 (and amended in 2009) required 'explicit consent' from those who visit websites for the installation of 'cookies' that may identify, track and profile them. However, this model of user consent has, in practice, required users to agree to the comprehensive collection of their data if they wish to partake in digital life through the most widely used platforms and services. The model places the burden of privacy protection on the individual and 'merely legitimises the extraction of personal data from unwitting data subjects' (Edwards & Veale, 2017, p. 49).

In addition to such self- and co-regulatory mechanisms, the regulatory framework for data collection and analysis is affected by, on the one side, data protection legislation and, on the other, rules that allow the state and other actors to collect and share data, for example for security purposes. In the UK, the Data Protection Act from 2018 controls access to, and use of, personal data, provides limitations for data collection and sharing, and gives citizens the right to, e.g., access their data and object to some of its uses. The EU General Data Protection Regulation (GDPR) from 2018 offers a comprehensive set of protections of citizens' personal data, particularly with regard to internet platforms and cloud computing, as it limits the use and sharing of personal data by companies inside the EU as well as the export of data outside the EU, and it addresses new challenges that have emerged with datafication. For example, it strengthens consent rules; makes new forms of automated and algorithmic decision-making

more transparent; assigns citizens a right to explanation and to challenge outcomes of algorithmic decisions; requires impact assessments for potentially harmful data uses; and mandates data protection by design. Many elements of the GDPR have been controversial (e.g., Edwards & Veale, 2017; Wachter, et al., 2017) but as a broad regulatory framework, it fills some of the gaps in the regulation of the data economy and offers new directions for providing citizens with some control over their personal data.

Yet that control is also affected by laws that regulate state surveillance and interception of communication (and thus of data traffic). The UK Regulation of Investigatory Powers Act (RIPA) from 2000, amended by the Data Retention and Investigatory Powers Act 2014, allowed a Secretary of State to authorize the interception not only of the communications of a specific individual but also of wide-ranging and vaguely defined types of traffic in bulk. Similar powers have been included in other relevant laws, such as the Telecommunications Act 1984 and the Wireless Telegraphy Act 2006. The UK Investigatory (IP) Powers Act 2016 provides comprehensive legislation to combine these previously fragmented rules for state-based data collection and analysis under one law and addresses a wide range of surveillance practices. While it opens up many of the traditionally secret surveillance measures to public scrutiny and oversight, it largely confirms, legalizes and expands existing surveillance practices. It allows, for example, the bulk interception of data that is generated, not least, on internet platforms; requires the collection of 'internet connection records' (i.e., people's web browsing habits) and enables a wide range of state authorities to view these without judicial approval; and allows security agencies to hack into people's computers and mobile phones (Hintz & Brown, 2017).

Data collection, sharing and analysis is affected, moreover, by the UK Digital Economy (DE) Act from 2017 which updates regulations on electronic communications infrastructure and services, as well as criminal justice issues such as copyright infringement. In particular, it facilitates data sharing between government departments, and it requires age verification by, and filters for, websites that provide adult content. While less prominent and less publicly controversial than the IP Act, the DE Act received strong criticism from digital rights groups because of its potential privacy implications (Open Rights Group, 2016a).

As these laws and regulations demonstrate, the policy environment is an interplay of national, regional and international rules. It includes, for example, the European Convention for the Protection of Human Rights and Fundamental Freedoms (ECHR), which was incorporated into UK law in the Human Rights Act 1998. Article 8 of the Convention guarantees everyone's 'right to respect for his private and family life, his home and his correspondence' (Council of Europe, 1950). Regional courts, such as the European Court of Human Rights (ECHR), can hear complaints and advise on the lawfulness of government action. Directives adopted by the European Commission are implemented by all member states and thus have far-reaching consequences for national law. For instance, the Data Retention Directive from 2006 required telecommunications services to retain communications data – such as, who communicates on the internet with whom, at what time, and from what IP address – for up to two years. It was revoked in 2014 by the Court of Justice of the European Union but was effectively continued by the UK government at the national level when it adopted the Data Retention and Investigatory Powers (DRIP) Act. Following a legal challenge, this Act was ruled unlawful by the European Court of Justice in 2016 (Hintz & Brown, 2017).

In addition to platform self-regulation and national and regional law, normative frameworks play an important role in guiding policy development and, potentially, policy reform. These may include international and UN declarations (such as the Universal Declaration of Human Rights from 1948 or the Declaration of the World Summit on the Information Society from 2003) and national policy statements, such as the UK Digital Charter from 2018. The Charter recognises that "personal data should be respected and used appropriately".[1] Concerns regarding the collection and analysis of personal data have also informed the creation of the UK Centre for Data Ethics and Innovation in 2018 whose task is to develop norms and guidelines on data use. Civil society organisations and campaign groups have played a particular role in advancing normative frameworks and raising public pressure. Digital rights and privacy organizations – such as Privacy International, Open Rights Group and Big Brother Watch – have long advocated for citizen rights in digital and datafied environments, but more recently, organizations such as Amnesty International have joined these efforts in recognition of the increasing role that datafication plays for a wider range of social justice issues. In some instances, they have been joined by internet companies which, although being primary gatherers of personal data, have often argued for restrictions to data collection by state agencies, not least due to concerns about the implications of governmental data collection for user trust in their services (Wizner, 2017).

As these examples show, the policy environment is dynamic and subject to ongoing change, and it is affected by different stakeholders and different policy levels. Yet the regulatory frameworks for datafication are often contradictory and unclear. While parts of them leave the citizen in a vulnerable position, some openings are emerging that point to possibilities for enhanced protection of, and control by, citizens.

Citizen-centric Policies

At its core, the need for citizen-centric regulatory frameworks refers to the question of where control lies regarding the data that is collected from and about citizens. Does it rest with the citizen, or with platforms and other internet companies, or with government institutions? In the platform economy, as noted above, it is collected and analysed by large platforms and other internet businesses (Facebook, Google, Alibaba, etc.) and accessed by governments, with little knowledge of citizens when, where and for what purpose collection, access and analysis take place. A citizen-oriented policy environment would thus enhance citizen control. This, then, leads to the question how to implement control: Should citizens have ownership over 'their' data, allowing them to trade and sell it or to protect it financially (e.g., through premium models of platform membership), or should data be seen as a right and a part of the self that cannot (or should not) be traded? While models of data ownership are increasingly part of the policy debate, the approach taken in this project (and this report) is to consider policies in which control over data remains with the citizen and cannot be handed over to the platform economy for financial gain. If the use and manipulation of data can have serious implications for questions of social justice and democracy (see above), merely adding a layer of potential income for the citizen does not address the wider challenges of the datafied society.

1 https://www.gov.uk/government/publications/digital-charter/digital-charter

Based on the examples raised in the previous section, we can conclude that a policy environment that thus connects questions of control with civic rights and social justice would address different aspects that intersect in the regulation of datafication. First of all, even if we reject ownership models, empowering the citizen in handling and controlling data that concerns them and is collected about them is crucial. Following the GDPR, this may include the right to challenge algorithmic decisions and to access and withdraw personal data from the platforms that have collected it. Secondly, it would require legal limitations to data collection, sharing and use, and the protection of citizens' privacy and datafied selves. Examples, again from the GDPR, may include restrictions to the collection of sensitive data and to the sharing of data between commercial and public entities.

Thirdly, the recognition that datafication now affects all areas of life and society, including the very core of social justice and democracy, would require a perspective that moves beyond 'digital rights' (such as online privacy) to include the various ways in which data can discriminate and can transform state-corporate-citizen relations. Data regulation may thus need to intersect with, and be developed in the context of, rules that affect non-data aspects, such as labour law, election regulations, anti-discrimination law, etc. Fourthly, a citizen-centric policy environment would require both legal and normative frameworks – with citizen-centric norms guiding policy, and laws and regulations as rigorous implementations of laws. Finally, it would have to be responsive to citizen voices and interventions.

In the following, we will address some of these aspects in more detail and explore their implementation and future prospects. We will focus, in particular, on issues that have been discussed and implemented as part of recent regulatory reform initiatives in the UK and the EU and thus on the legal and normative dimensions of citizen-centric policies.

Methods

This project explored current trends in the regulation of data collection and analysis by and through platforms. It focused, in particular, on a set of recent laws and regulations in the UK and the EU – incl. the UK Investigatory Powers Act, the UK Digital Economy Act, and the EU the General Data Protection Regulation – but it expanded its perspective to wider regulatory developments. We were interested, broadly, in the intersection between the expansion of data collection, sharing and analysis in the context of datafication, on the one hand, and emerging calls for the protection of citizens against data-related harms and for citizens' control over data that concerns them, on the other. In that context, specific questions addressed the agendas that inform and underpin policy change, the responsiveness of policy processes to public concerns, regulatory responses to exclusions, inequalities and discriminations in the platform economy, gaps and inconsistencies in the current policy framework, and new or alternative approaches towards regulating the use of citizen data.

These questions were investigated through a combination of desk research and expert interviews. As a first step we conducted a review of both academic and public literature on the policies that formed the core of the analysis – IP Act, DE Act and GDPR. We reviewed journal articles, blog entries and other literature that emerged in the immediate aftermath of or, in the case of GDPR, before the adoption of the new policies. This provided a perspective on key debates that emerged around the new laws and regulations during their development and adoption phase. This review was conducted between January and March 2018.

Secondly, we conducted semi-structured interviews with members of different stakeholder communities, with the goal of exploring different perspectives on the specific laws and regulations as well as on key themes that emerged, and to investigate broader regulatory trends. The interviews took place between August and October 2018, and each lasted between 30 minutes and 1 hour. The interviewees are listed in the table below. All are policy officers or policy directors in their organisations and concerned with questions closely related to those addressed in this research project; 5 interviewees were male and 2 female; and all are based in the UK.

Interviewee #	Stakeholder group	Organisation
Interviewee 1	Government	UK Department for Digital, Culture, Media and Sport (DCMS)
Interviewee 2	Government	UK Department for Business, Energy and Industrial Strategy (BEIS)
Interviewee 3	Business	Internet Service Providers Association (ISPA)
Interviewee 4	Business	techUK
Interviewee 5	Civil Society	Open Rights Group (ORG)
Interviewee 6	Civil Society	doteveryone
Interviewee 7	Civil Society	Privacy International (PI)

Finally, the research was informed by meetings and workshops that brought together different stakeholders and affected groups. In particular, we held a fact-finding workshop as part of the conference *Data Justice* at Cardiff University, 21-22 May 2018, to review current policy frameworks, identify gaps and shortcomings, and explore proposals for policy reform.

Findings

This chapter combines insights from the document analysis and the interviews. It starts by addressing elements of recent policy reform that may empower the 'user' in exerting control over their personal data and then discusses questions of data protection, data collection, the political-economic environment, data localisation, and normative frameworks.

Empowering the Citizen

The most immediate condition for empowerment and control is transparency and knowledge. Articles 13-15 of the GDPR offer a significant step in this direction by setting out **information and access rights** for data subjects and requiring them to be informed of instances of automated decision-making and profiling, in addition to 'meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject' (Edwards and Veale 2017). Together these articles provide, as some have claimed, a **right to explanation** as well as a means to regulate and challenge the power of algorithmic decision-making (ibid).

However the effectiveness of this right (and other GDPR provisions) has been contested by academic observers. Wachter et al (2017) characterize it as a much narrower right to be informed which amounts to an explanation of the functionality of automated decision-making systems but not specific automated decisions. The Article 29 Working Party confirmed that GDPR provides a 'more general form of oversight', as opposed to 'a right to an explanation of a particular decision' (A29WP 2017, Edwards and Veale, 2017). In practice 'meaningful information' refers to input information provided by the data subject; relevant information provided by others (such as credit history); and relevant public information used in the decision (such as insolvency records) (A29WP, 2017, p. 14). The actual inner workings of algorithms are not covered, including the training dataset used, which has been seen as an overly restrictive and prescriptive approach to algorithmic transparency (Edwards and Veale, 2017).

Kaltheuner and Bietti (2017, p. 15), similarly, criticise that this 'becomes the right to a general explanation, rather than a right that would allow individuals to obtain an explanation for a particular individual decision that affects them'. Veale and Edwards (2017, pp. 66-67) note that relying on the right to explanation risks creating a new 'transparency fallacy' similar to the illusion of online consent as 'individual data subjects are not empowered to make use of the kind of algorithmic explanations they are likely to be offered […] individuals are mostly too time-poor, resource-poor and lacking in the necessary expertise to meaningfully make use of these individual rights.' An explanation alone, they argue is likely not meaningful enough to confer much autonomy 'on even the most empowered data subject' (ibid: 67).

The right to access one's data leads to a key GDPR provision that aims directly at challenging platform power – the right to **data portability**. The regulation requires platforms to allow users to move their data across services and thus encourages competition between digital services. The goal is to facilitate switching from one service provider to another, while advancing 'user choice, user control and consumer empowerment' (A29WP, 2016). Commentators have argued that this will prevent vendor lock-in, particularly prevalent on social media platforms: '[I]f the

switching costs are high, providers will be able to create a high degree of lock-in. For providers that rely heavily on data provided by users, restricting data portability is a way to tie users to their services' (Graef et al, 2014). By enabling users to transfer their data easily from one system to another, competition will be enhanced as market access for new services is facilitated and anti-competitive network effects are alleviated (Vanberg and Unver, 2017). The GDPR requires for personal data to be downloadable in a 'structured, commonly used and machine-readable format', i.e. the data must be standardised and re-usable. To that end, the European Data Protection Board (which replaced A29WP under the GDPR) recommended that industry stakeholders and trade associations work together to create 'a common set of interoperable standards and formats'.

The problems of data portability were highlighted, to varying degrees, in our stakeholder interviews. Government interviewees praised data portability as a significant step towards citizens' control over data and highlighted it as policy priority. The technical implementation, though, requires improvement, as a BEIS official told us:

'The right for portability is there, it's fully workable in terms of being able to actually download your data in a meaningful way or in a format that is then re-useable in ways to encourage consumers to very easily be able to switch providers. [However] portability is not in a place where, whilst you can have a download of all of your data in a reasonably intuitive useful format, you cannot then easily give it to other platforms.'

The business representative from techUK agreed that while consumers would benefit in the long term, 'on a technical level it's a challenge.' However he raised more significant concerns: 'I think over time, users will realise that this isn't a carte blanche right that they can move things from one area to another incredibly easily. It's a very technical thing, and with portability you need interoperability as well.' The civil society representative from ORG reiterated the intersection of technical and regulatory issues as a fundamental problem and noted, 'portability creates an obligation for companies to make the data available but it's not very clear whether it creates an obligation on companies to accept the data.' Further, 'portability relies on common formats and data structures but I think you are going to need government [...] to force companies to sit down around the table in a particular sector and come up with a common format and common processes to do it, and I think that's a problem. Just putting an Article on portability by itself is not going to be enough.' Moreover, implementing portability might incur disproportionate costs as 'what is technically feasible for one data controller might not be technically feasible for another' (Vanberg and Unver, 2017, p. 4). Vanberg and Unver suggest the requirement for data portability will place a higher burden on SMEs than on larger companies.

Further, data portability points to a core problem of the GDPR and data protection more generally: What **types of data** are addressed? The wording of 'data provided' is sufficiently vague that it can be interpreted 'restrictively' or 'extensively' (De Hert et al, 2017). In the former interpretation portability applies only to personal data that the controller has received from the subject (such as profile information or questions answered during account registration) while in the latter portability extends to data observed by the controller (for example through GPS, cookies, preferences, analysis of browsing data, etc). WP29's guidance states that portability be interpreted broadly but formulates clear limits. It includes 'personal data that relate to the data subject activity or result from the observation of an *individual's*

behaviour but not subsequent analysis of that behaviour. By contrast, any personal data which has been generated by the data controller as part of the data processing, e.g., by a personalisation or recommendation process, by user categorisation or profiling is data which is derived or inferred from the personal data provided by the data subject, and is not covered by the right to data portability.' This interpretation is unlikely to disrupt platform power as it places control over **inferred and derived data** firmly in the hands of platforms and other internet businesses. Particularly intrusive data activities, such as profiling, often utilize inferences from observed data such as Facebook clicks and likes but, according to this perspective, do not belong to the data subject but to the system that generates them (Edwards & Veale, 2017, p. 67). Ultimately, it may be up to the courts to determine which data portability applies to and whether the right is interpreted restrictively or extensively (De Hert, 2017).

The GDPR, further, expands and refines a classic notion of user empowerment: **consent** rules. It defines consent as an ongoing and actively managed choice, rather than a one-off compliance box to tick, and requires consent to be actively obtained, giving users the option to withhold or withdraw at any moment. The tightening of consent rules by GDPR makes it more difficult for platforms that have a direct relationship with users (e.g., Facebook and Google) to use the personal data they hold for advertising purposes without user permission. While they can process personal data necessary to provide services that their users request, using data for any other purpose requires additional user permission. In practice, this means that users will have to opt-in to tracking (Ryan, 2017). A company such as Facebook now has to explicitly name all third-party data sources. As the ICO notes in their guidance on consent: 'name your organisation and any third parties who will be relying on consent – even precisely defined categories of third-party organisations will not be acceptable under the GDPR.' Third party data processing will therefore face disruption as users now have to consent to the sharing of their data by the platform with any third party. Insufficient consent models already led to a court ruling in Belgium pre-GDPR where Facebook's model of consent was found to be invalid because users are not 'sufficiently informed' about being tracked as they browse the web through the platform's social plug-in which allowed sensitive data such as health-related, sexual and political preferences to be gauged (Ryan, 2018a). This has implications for any platform which shares this kind of data with companies during real-time bid requests for advertising purposes as the new GDPR consent rules collide head on with this practice (Ryan, 2018b).

Explicit consent is even harder to obtain for platforms with no direct relationship with the data subject, such as companies working in online behavioural advertising that track users across the net. As consent must be sought for each specific purpose, each data broker or adtech vendor is now required to obtain consent for each profile that is bought and sold. Some commentators in the adtech sector are predicting that the GDPR will severely weaken the third party data market, while others are calling for advertisers to stop relying on personal data and instead monetize 'non-personal data' (Ryan, 2018a). Indeed, an unnamed big tech executive has claimed that 'personal data is quickly becoming a toxic asset' and that 'surreptitiously gathered personal data [is] the radon gas of business and a silent killer' (quoted in Rainie & Anderson, 2017). Third party data processors may have to rely on first party platforms (like Google and Facebook) for consent.

The construct of **legitimate interest**, however, offers an alternative legal basis for processing data and thus a possible way for platforms to bypass the GDPR's consent rules. Recital 47 of

the GDPR states: '[t]he processing of personal data for direct marketing purposes may be regarded as carried out for a legitimate interest.' The Article 29 Working Party, on the other hand, indicated that this would require adequate user controls and safeguards. In its recent guidance on legitimate interest the ICO (2018) introduced a balancing test whereby 'the individual's interests, rights and freedoms' must be weighed up against the interests of both the data processor and third parties. The guidance says that 'in particular, if [the data subject] would not reasonably expect you to use data in that way, or it would cause them unwarranted harm, their interests are likely to override yours. However, your interests do not always have to align with the individual's interests. If there is a conflict, your interests can still prevail as long as there is a clear justification for the impact on the individual.'

Even though consent is an established legal construct, most stakeholder interviewees regarded it as problematic and ambiguous, with differing opinions as to what it should look like or whether the mechanism should be abandoned altogether. Consent meant different things to different interviewees and the criteria for setting the benchmark was a point of contention. One government official said that 'meaningful consent' was the ideal, while another focused on 'informed consent'. A business representative agreed that informed consent is desirable but added 'valid consent'. It was not always clear what these different versions meant. On this note the doteveryone representative highlighted that informed consent is context dependent:

'We have a view that if you're asked to consent to terms and conditions at the beginning of using a platform or a service, it's not really consent because actually the monitoring of data for each individual case from then on will change and this is depending on context.'

The notion of consent as information emerged as a point of divergent opinions. While government officials accept the need to improve consent mechanisms as users are struggling to understand what they consent to, they maintain the value of, and need for, the principle of consent itself. The DCMS official said 'we might want to make it easier to look at terms and conditions, or make it easier for consumers to understand terms and conditions, and we need to focus on understanding what they give consent to, rather than just having consent as the solution for everything.' For the BEIS official consent is about 'communicating information' for consumers to 'absorb.' To ensure that users understand terms and conditions, privacy notices should be redesigned and presentation formats changed. Personal Information Management Systems might help enhance user control. The civil society view, on the other hand, was that consent should not primarily be about information but should entail a real choice with the viable option to opt out. The ORG representative noted that current models of consent, 'are based on choices that are not real […]. Consent should be totally free, it should be a real choice. If there is no choice, the choice is either you use this service or you don't.' The PI representative concurred:

'the way that especially advertisement driven companies have implemented consent is a joke. It's not freely given and unambiguous [...] it has to be as easy to opt-in as it has to be to opt-out but that's proving not the case. It's rigged against you. It's very easy to say click yes, agree and not very easy to actually say no, I don't want this and this is something we want to challenge. So currently no, it's not a game-changer in the way I see it being implemented.'

The techUK representative agreed with some of these concerns: 'If it's not genuine consent and someone can't really prevent themselves from consenting to it, then you shouldn't be using consent because that's misleading the consumer at the end of the day.' The doteveryone

representative situated the consent principle in the broader debate regarding the societal implications of data, calling it a 'red herring': 'as an individual, you will consent to give your data over because the individual benefits you receive far outweigh any potential risks really to you on a personal level' and yet the most challenging issues around datafication are felt on a societal level. 'It's very hard for people to internalise the societal impacts into their individual decisions.'

The critique of consent as a regulatory principle led some civil society interviewees to fundamentally reject consent as inadequate. The ORG representative noted: 'I think the state of the art understanding of privacy nowadays is that consent is definitely not the way to go. What you need to do is to create engineered systems that don't collect as much data.' The PI representative added: 'We can rethink consent as much as we like but we're not going to get it.' Rather than being a key component in the arsenal of user empowerment, the doteveryone representative suggested that consent is currently a burden for individual users and thus disempowering. Referring to a recent survey of public attitudes, he noted:

'43% of our respondents say that they say yes to terms and conditions because tech companies will do what they want anyway. So there's a feeling of total lack of empowerment there and almost resignation to tech companies. 89% of people say that they just say yes to terms and conditions without reading them.'

Overall, interviewees differed in their assessment of user empowerment both as a goal and in its actual implementation, with government representatives being most supportive and civil society members most sceptical. Most agreed that the GDPR's potential for empowering individuals and assigning new data subject rights has not yet been fulfilled. The ISPA representative implied that while theoretically the GDPR puts people more in control of their data 'whether it does in practice, I think it remains to be seen.' A government official noted that she had expected more data portability and subject access requests and that 'people are not making as much use of their rights as we could have predicted.' The other government official added: 'I think the right to be informed how your data is being used has probably had more immediate and obvious impacts, although I wouldn't say it's being used to the full extent that policy makers might have envisioned or has been as impactful as [they] might have hoped.'

Citizen Protection

If the strategy of user empowerment has limitations, it needs to be embedded in policy that protects citizens from negative implications of datafication and formulates stricter rules to underpin citizen control of data. In the GDPR, this includes the principle of **purpose limitation** – Article 5 (1) – which means that personal data must only be 'collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes.' Further, the regulation sets **limitations to automated decision-making**: Article 22 prohibits any 'decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.' While this rule cannot quite capture current power shifts in data-based decision-making in the public sector where humans remain part of the process but with diminishing influence over the results of data processing (see Dencik et al, 2018), it nevertheless poses an important safeguard in the context of emerging forms of algorithmic processing.

The GDPR's specific rules against the **processing of sensitive personal data** contribute a further dimension to these restrictions by limiting the ability to profile data subjects. Article 9 states that 'Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited.' Data brokers collect and trade this kind of sensitive personal data to partners like credit scoring and insurance companies and rely on it to target select adverts for users, personalise services and products, and categorise users based on their online activity. These profiles may also include income, age, gender, sexual orientation, religion and political leaning. As the Spanish court in a case against Facebook noted in September 2017, '[Facebook] data on ideology, sex, religious beliefs, personal preferences or browsing activity are collected directly, through interaction with their services or from third party pages without clearly informing the user about how and for what purpose will use those data' (Cabanas et al, 2018, p. 4). However this collection (rather than the processing) remains legal under Article 9. Moreover (and as noted above), it is uncertain what happens when sensitive data has been inferred or derived from a user's activity on that platform or generated by its proprietary algorithms. Inferences can gain unobservable (and potentially sensitive) data about a person from pieces of observable (but 'non-personal') data. The risk scoring in the criminal justice system in the US, where a wide range of data points has led to racial bias, may serve as a prominent example (Angwin et al., 2016). This issue will only get more important as data analytics is increasingly at the centre of governance. But as noted before, sensitive data inferred from 'non-sensitive' data may not be covered by GDPR rules.

As this example shows, there are significant concerns regarding the GDPR's unclear formulations and its potentially limited impact. A common thread in our civil society and industry interviews was the fact that the GDPR's impact still needs to be tested in the courts and without this case law it is just 'too early to say'. A Policy Manager from the industry association techUK noted: 'Time will tell. We haven't had any cases under GDPR yet. There's a lot of case law to be written on GDPR.' A representative from Privacy International said 'the law's ambiguous sometimes. We still need court cases that will set precedents.' An Open Rights Group interviewee told us that good cases will be essential to implementing the GDPR because without them it will be 'business as usual'.

As an area of necessary further protection, interviewees highlighted the practice of **profiling**, specifically in the context of **targeted advertising**. Both government and civil society representatives agreed on the current regulatory shortfalls and attributed societal harm of targeted advertising and profiling – especially in the wake of the Cambridge Analytica/Facebook scandal. The doteveryone representative highlighted that extra scrutiny is needed, particularly in the area of political advertising, and called for a limit to micro targeting:

'The problem is that 'you can't see how an advert has reached you. So it would be good to highlight the provenance of an advert and exactly how they're being targeted and based on what demographics. Things that are available to the Electoral Commission offline, [...] we think should apply online as well.'

The DCMS official explained that her department is currently looking into the advertising model of platforms and she is particularly concerned about the lack of public knowledge as to

how Google and Facebook 'finance themselves.' Yet some civil society interviewees, while agreeing with the general critique, raised concerns about the (both public and policy) focus on social media adverts and regarded this as too narrow for the scope of the problem. The ORG representative noted that regulatory attempts in the wake of election meddling and Cambridge Analytica scandals are so far 'very focused on narrow superficial things like Facebook adverts', with less attention given to 'data collection outside of Facebook platforms that doesn't get controlled. Or the online advertising systems that, in many cases, now are run by Google and Facebook but it's not just Google and Facebook, it's lots of media marketing companies.' Similarly, the PI representative disliked the focus on 'platforms' as such, preferring to think of the online advertising 'ecosystem' and business models behind the platforms. She explained that 'platforms only exist because of the advertisement system behind them' and this 'relies on the exploitation of data, of not just selling peoples' data but of selling their attention on the basis of collecting vast amounts of data.' For her this means that the interests of platforms and those of its users will always be diametrically opposed.

Despite the GDPR's attention to profiling and sensitive data, there are concerns among civil society on whether ongoing advertising practices on social media might be effectively addressed by the European regulation. Its, in the words of the ORG interviewee, 'limited' provisions for profiling may not be 'serious enough' to address the problem sufficiently. The BEIS official implied more regulation may be needed regarding profiling and attributed the current data 'backlash' to the consequences of online personalisation, especially price discrimination and altered search results. He drew a line between acceptable and unacceptable forms of personalisation (giving the example of the insurance market using data to issue quotes as acceptable).

Data Collection

While there have been some advances in protecting citizens' data rights and enhancing their control over data, legal provisions (and requirements) for data collection and sharing have increased, too. A policy environment has emerged that, on the one hand, recognizes citizens' needs for controlling (some) data and enhancing (some) citizen rights while, on the other hand, expanding the collection of citizens' data via platforms and opening it up to government access. The IP Act includes, for example, the **mandatory retention of communications data** which requires the generation and collection of 'relevant communications data' by 'telecommunications operators' for up to 12 months. Most civil society and technology organizations have opposed mandatory data retention on the grounds that it is excessive and violates the right to privacy, as well as security risks such as breach, theft, misuse, and abuse of the data. As with Internet Connection Records (ICR, see below) data retention may well require platforms to produce data outside of their current business practices. In their written evidence[2] to the UK Parliament Facebook, Google, Microsoft, Twitter and Yahoo expressed concern that the IP Act's data retention and ICR provisions would require them to reconfigure their networks or services in order to generate data.

[2] http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/draft-investigatory-powers-bill-committee/draft-investigatory-powers-bill/written/26367.html

Further, the IP Act introduced the requirement for internet service providers to capture **Internet Connection Records** (ICRs). Similar to the data retention rules, ICRs force platforms to produce large volumes of new datasets. ICRs are an artificial construct with no concrete definition in the Act, and they are not a term recognized by the computing industry. It remains unclear whether ICRs can be matched to real categories of data processed by internet companies (ISPA, 2016). They are typically understood as a user's 'browsing history'. ICRs are concerned with using communications data to enable law enforcement to find out the identification of a device and the identifications of services accessed. However, many platforms require constant connection to push content to their users, meaning that if an ICR showed an individual to be using a particular service, the value of this insight as evidence may be limited and it may fail to show what was communicated or to whom (Coats and Komisarczuk, 2016). ICRs and data retention employ, specifically, the use of platforms, but data collection practices enabled by the IP Act extend far beyond to the **bulk interception** of communications data, the **hacking** into devices and networks, and a range of other measures that make citizens' data vulnerable to collection, interception and analysis.

Some platforms are forced to monitor and censor both their users and content. The **age verification** rules of the DE Act require platforms to enforce age verification both on their own sites and other third parties – known as ancillary service providers – and to block infringing sites. The Electronic Frontier Foundation stated that 'the possible impact of the law extends beyond video hosting websites, but also extends to payment services providers, hosting providers, and advertisers on those websites, whether they are based in the United Kingdom or overseas' (Malcolm, 2016). This may mean, for example, that the British Board of Film Classification (the government's appointed age-verification regulator) requires the withdrawal of services such as advertising and payment services, and internet service providers and mobile network operators to block access to non-compliant services. Further, websites will be compelled to create databases of users' viewing habits along with their personal data – including credit card details – to ensure verification (Open Rights Group, 2016b). There are concerns that data protection is not a strong enough safeguard here because consent is 'forced' in this context (ibid).

As noted above, the DE Act also expands rules for **data sharing** between government departments and therefore contravenes the efforts made in the GDPR to limit the widespread distribution of data gathered through platforms. The former chair of the Government's privacy and consumer advisory group (PCAG) has noted that Part 5 of the DE Act directly contradicts the GDPR and also the government's Technology Code of Practice by transferring control of personal data away from the citizen over to government (Fishenden, 2016). For example, this section allows for government departments to share personal data with private companies including debt collectors in order to 'improve public service delivery' – without citizens' knowledge – which raises questions about the GDPR's focus on consent-based data processing. While the use of data by the government and commercial actors is regulated separately, the increasing use of commercial data aggregation tools by the public sector blurs this line and points to a potential conflict between different legislations (see Dencik et al., 2018).

The theme of data collection demonstrates some of the most significant differences in stakeholder perspectives. Government interviewees highlighted the economic benefit of data collection and pointed to the 'trade-offs' between benefits and risks, as well as the need to 'balance' innovation with data privacy concerns. Consequently, advancing user control over

data was encouraged but situated within a need to advance the digital economy overall. Rather than limiting collection and sharing, government interviewees tended to advocate closer regulation of data use and analysis, together with robust mechanisms for informed and meaningful consent and frameworks for enhancing data ethics. In other words, sparsely restricted collection of data should be balanced by rules constraining its use. Civil society interviewees disagreed. They maintained that the first step of data analytics and of any potential privacy violations remains the collection of personal data. They pointed to GDPR provisions that demand data privacy by default and the minimisation of data collection. They also pointed to academic research on the chilling effect of data collection, as well as court decisions on how 'the existence of data in the first place was already having an effect on people. If people knew that the data was being collected, they already changed their behaviour' (ORG interviewee).

Political-economic Context

As we could see already in the section on consent, data collection and analysis by platforms are often connected to their economic context. Civil society representatives, in particular, pointed to the business models sustaining platforms as underlying concern for citizens' data protection and control. Challenging these business models was a strategic priority for the representatives from Open Rights Group and Privacy International. Without such broader measures, interviewees demonstrated a lack of trust in the effectiveness of several mechanisms addressed by the GDPR, such as consent, and in the prospects for GDPR compliance and enforcement – the latter in direct contrast to the industry interviewees.

In the context of the economic obstacles to regulation, interviewees shared a concern with the limits of competition law (or anti-trust regulation) in regulating digital markets. All interviewees were explicit in noting the difficulty of defining and measuring this market as traditional market boundaries do not apply and sectors are blurred, with companies like Google/Alphabet straddling multiple sectors.

Civil society interviewees shared a strong scepticism about the GDPR's impact on platforms overall. The ORG representative distinguished between 'front end changes' and 'back end continuity', noting that 'GDPR is not having a fundamental effect on privacy at the platform level, but bringing some… I wouldn't say cosmetic but it's bringing some changes to the surface of the data practices but not changing the fundamentals.' He added that dominant platforms would be able to remedy any potential GDPR limitations on, e.g., third party tracking and data sharing by bringing some of these activities in-house, thereby leading to further concentration:

'So before, Cambridge Analytica built a psychometric profile of lots of Facebook users and now it seems that Facebook themselves, they are building political profiles and all sorts of detailed measures of their users.'

Civil society members thus felt that the GDPR's impact is limited by the wealth and power of dominant platforms, rendering compliance less of a challenge than for SMEs, while enforcement becomes a more significant problem for regulators. For example, a representative from doteveryone said 'bigger international platforms, who have the capacity and resources and legal teams to deal with the GDPR have been able to adjust and adapt quite well.' This was reiterated by the PI official who claimed that although the GDPR shifts the onus from data

subject to data controller 'for big platforms that [have] always had compliance departments, that's not a game-changer.' She added that enforcing the GDPR is a significant challenge with many companies still 'blatantly not complying [...] because traditionally data protection is not a very strong area of enforcement and even if fees are considerably higher under GDPR, you can still make the decision to simply not comply. Even if then you have to pay a fine of 4% of global turnover, if you do the math over ten years, this can still be a profitable decision to make.'

Data localisation

Calls for the localised storage and processing of data – and thus for taking it out of the realm of (mostly) US-based platform businesses – have emerged in many countries, and certain types of data are required to be localised in jurisdictions as different as Australia, Nigeria and Russia (Browman, 2017). The GDPR does not explicitly advocate for data localisation, but its strict requirements for transferring personal data to non-EU countries may certainly encourage it. However all of our interviewees rejected this approach, although for different reasons and, in some cases, with differing understandings of what data localisation means.

The techUK representative was most outspoken on the issue: 'The global trend of data localisation is incredibly concerning and I'll be perfectly clear, data localisation does not lead to good data protection. There's no argument that is correct anyway that says you need data localisation in order to protect data. If anything, it makes data less secure rather than more secure.' He added: 'The freeing up of data flows across the world would improve the data economy and the end result for users considerably.' This, in his view, also has cybersecurity implications: 'You need the data to flow round the world as you track it so that you can keep it secure from various threats across the world. If you require a piece of data to be held in one territory, you can't operate on that model.'

UK government representatives agreed, in principle. As the DCMS official told us: 'We've always advocated against legislation which would basically provide unjustified data localisation measures.' While the business perspective on the issue and, to some extent, the government perspective may be unsurprising, critique of data localisation also came from civil society. The PI representative said: 'In many cases, data localisation is a pretext for increased, direct access to data that otherwise wouldn't be possible, but the idea of localising data is not compatible with the way that data flows and platforms work.' The ORG representative added: 'Mandating data localisation by the Russians or Chinese, that model of data localisation, is seen as problematic by everyone and I would say civil society as well.' There is thus strong concern that data localisation would merely enhance further data access by national governments, rather than enhance citizen rights and alleviate unequal power relations in both the digital economy and digital geopolitics.

This view was grounded, however, in an understanding of data localisation as 'a regulatory issue around jurisdiction of data' (as the doteveryone representative noted). An alternative suggestion was raised by the ORG representative who drew a distinction between localisation as a form of data nationalism and localisation as a form of decentralisation. As he noted, 'the problem is decentralising the internet, that is something that would be good and […] should involve more decentralisation of data.' Despite scepticism towards national solutions for

localising data, there is thus a strong interest amongst civil society in decentralising monopolised (economic and geopolitical) power over data, recognising that approaches based on, for example, communities and municipalities can play an important role in enhancing citizens' data control.

Norms: Data Ethics

The normative context of laws, regulations and (business and government) practices can be an important part of the policy environment. Data ethics has become a particularly popular framework for data use in both public institutions and the private sector. Frameworks for data ethics have been developed by a range of actors, from scholars[3] to the UK government[4], and the new 'Centre for Data Ethics and Innovation'[5] may institutionalise data ethics in a governmental context. The Centre has drawn significant attention in the UK policy debate and most of our interviewees regarded it as a positive step in the policy environment and as a place where questions of data protection, privacy and innovation could be discussed in an integrated way. As the doteveryone representative noted, the Centre may serve to balance competing interests because 'there's routinely situations where peoples' rights and ethical considerations are in conflict. So you need to weigh those up.' The DCMS official explained governmental enthusiasm for the Centre 'because we see that there needs to be this balance struck between innovation and ethical use of individuals' data, and that sometimes you might still want to collect the data but then limit the use or ensure that actually perhaps it's then used in ethical ways for good outcomes.'

The PI interviewee acknowledged the debate on ethics as a way to move beyond data protection and towards a more comprehensive understanding of data, 'which I think is very good because data protection has many flaws' and pointed out that the UK is playing a leading role in these debates. Ethics, she told us, is part of the UK's drive to have public trust in technology. However, civil society interviewees were cautious about the implication of a turn towards ethics in replacing robust law and legitimising both the use and collection of data. The latter was visible in the above-mentioned government quote about collecting data for ethical use. Ethics norms might thus facilitate further and more expansive data collection. As for the potential tension between data ethics norms and the creation of hard law, the PI representative pointed to 'corporate attempts to use discussions about ethics and data governance to undermine existing laws.' She expressed skepticism about the currency of data ethics as a potential distraction from the core problem of protecting citizens through adequate laws and regulations.

Between enthusiasm for, and criticism of, data ethics approaches, the interviewee from doteveryone advocated for the new Centre to take a 'stewardship role for how the public would like data to be used and technologies more broadly' and that 'there needs to be an ongoing in-depth public conversation about the values that we want our digital technologies to embody.' Such a conversation about underlying norms and policy goals would be a significant

3 https://www.oii.ox.ac.uk/news/releases/what-is-data-ethics/

4 https://www.gov.uk/government/publications/data-ethics-framework/data-ethics-framework

5 https://www.gov.uk/government/groups/centre-for-data-ethics-and-innovation-cdei

component of a citizen-centric regulatory environment. Whether ‘data ethics’ lends itself to this task, however, remains contested.

Towards Citizens' Control Over Data?

As we noted at the beginning of this report, the context for this study has been marked by the rise of a lightly regulated platform sector that has generated value from extracting and analysing user data, and by pervasive data collection by state agencies. The digital citizen, as we noted, is increasingly monitored, categorized, scored and assessed by both commercial and governmental actors, with little influence over, or even understanding of, how this is done. This questions the idea of the active digital citizen in fundamental ways.

However, we can observe a different trend emerging, one that advances the **active control by citizens over data** that characterises them and that is collected about them. In part, this trend is fuelled by data-related scandals, such as the Cambridge Analytica case, and by a public discourse that puts pressure on both the public and business sector to address data-related harms. As one civil society interviewee noted, 'people are starting to wake up to it and there's been a lot of movement in the UK.' He calls this 'regulation by outrage' and acknowledges it as a driver for policy debate. Policy scholars speak of a 'policy window' (Kingdon, 1984) that opens up and allows for new debates and policy concepts to take hold in response to external events or changes in the political equilibrium.

Yet the momentum for citizen control emerged, not least, as part of persistent critiques of platform power and data extraction over extended periods of time, exacerbated by the Snowden revelations and expressed through policy initiatives such as the multi-year effort to develop the GDPR. As our interviews demonstrate, the need for data autonomy has reached the UK policy debate, too, and there is some political will to address citizen concerns with regards to datafication, including the data extraction practices of platforms. The UK Digital Charter and Data Protection Act, both from 2018, refer explicitly to citizen control over data.

An important component of a policy environment focused on citizen control over data is the **empowerment of the citizen** to take informed decisions in a datafied environment. The GDPR enhances citizens' active role in many different ways, including the right of access to personal data and to data portability**,** a right to explanation, and stronger consent requirements. However, as our interviews showed and many other observers have confirmed, user empowerment has severe limitations without robust rules that **restrict the exploitation of people's data**. The GDPR has proposed useful elements through, for example, the principle of purpose limitation, restrictions to profiling and to automated decision-making.

Yet while these provisions push the regulatory framework in the direction of increased citizen control and protection, our research also points to a number of concerns that place significant constrains on this emerging trend. Most prominently, the **unrestrained collection of data** by both commercial and state actors, and the expansion of **data sharing** across different institutions and agencies counter-acts some of the advances of data protection legislation. While the GDPR limits data sharing between platforms, the IP Act and DE Act require platforms to make more of their data available to public authorities. As our interviews have shown, there is no appetite among policymakers for restrictions to the collection of data in the name of citizen control. The debate on ethical data use and the protections by the GDPR may actually turn attention away from questions (and risks) of data collection. As civil society

participants in our research noted, data collection inhibits citizen control and therefore requires careful implementation and robust restrictions.

The specific types of data that are regulated require critical interrogation. The GDPR, according to most of its interpretations, only applies to personal data concerning an individual that he or she has 'provided to' a data controller and thus excludes combinations of this data with data from other sources or providers, as well as information inferred or derived from this data. **Inferred and derived data** has become more valuable than the rather limited 'personal data' that the data subject knowingly provides, yet it is typically owned by the data analytics company that processed and thus 'created' it. New rules are required to address the collection and use of inferred data.

Data localisation proved to be an unpopular approach among all interviewees, at least in its implementation as a national strategy that may assign greater control over data to national governments. Localisation policies at municipality level – as attempted, for example, in Barcelona – may offer avenues for enhancing citizen control over data without necessarily empowering national governments. In that respect, decentralisation was highlighted by one interviewee as a guiding principle that is, so far, underexplored.

This points us to the conceptual context of data policy which may require significant attention. While the role of active citizens in the datafied society is important, the dominant **construct of the informed user** as guideline for policy is not always helpful. Rights to, for example, access to personal data require detailed knowledge and significant engagement by the citizen. The model of consent may lead to a 'consent fallacy' and an illusion of control if, as all our interviewees agreed, its implementation does not reflect informed and meaningful decisions by the user. Civil society representatives, in particular, shared a fundamental critique of the consent principle, which was partly, and interestingly, supported by representatives of the private sector. Instead of allowing the sparsely informed user a limited choice to hand over their data, they argued for both technical and regulatory restrictions to data collection. While a proactive digital citizen making informed choices about their data-related activities would be an ideal scenario, it remains an unrealistic one for the time being.

A focus on the informed user and thus on individual responsibility, moreover, has problems addressing societal harm of datafication. This points to the **limits of individual approaches** to regulating data. Data typically denotes a relation to others, and the individual's place within a broader collective, from which they can either be distinguished or to which they belong. This is the case even for 'personal' data, but much more so for the wider range of inferred and derived data. Data, in that sense, is only ever valuable in relation to others, which is particularly apparent in case of the categorizations, rankings and 'risk scores' produced by data analytics companies and used, among others, by public services. Such evaluations compare citizens and consumers with each other and allocate resources based on the results. Further, data always affects others. A data subject's online communication, 'friending', and service use in the platform economy inevitably connects them with others and affects the data inferred about those others – e.g., the characteristics of their social network, and potentially their personal credit score, their 'risk' according to police or social services, etc. Data, then, is never entirely 'individual'. Yet the key concepts in use continue to focus on personal data and individual privacy.

Our interviews demonstrated that there is limited appetite yet for developing regulatory concepts that move beyond a focus on the individual. Government interviewees' understanding of the main future policy challenges focused on the need to create and support empowered individuals. Civil society interviewees hinted at alternative, collective understandings of data but expressed a need to focus their everyday advocacy work on advancing, improving and implementing the laws that exist.

Normative development is therefore crucial to underpin and inform efforts to advance citizen-centric policy frameworks. Citizens' rights and people's control over data form a growing discourse but continue to compete with other understandings of policy needs and aims. As the debate around the IP Act demonstrated, dominant sectors of the state have successfully established 'security' (or rather, a specific understanding of national security) as a prominent benchmark. Business interests (supported, as we have seen, by government) have used 'innovation' as the frame for guiding data policy and arguing for reduced restrictions to data uses. The protection of citizens and the enhancement of their control over data still need to assert their place (and be 'balanced') against goals of national security and economic innovation. There is no primary benchmark and thus no clear goal for policy (yet) but different objectives, as well as the concepts on which they are based, factor into policy decisions.

Data ethics have emerged as a strong normative approach to guide datafication and can offer important advances in the responsible treatment of data. However, as some of our interviewees argued, both the concept and practice of data ethics have significant limitations. Without being accompanied by a robust regulatory framework, they risk transforming the protection of citizen rights into a self-guided act by public and private sector entities that is either voluntary or negotiated between those stakeholders. Legislative and regulatory rules for, and restrictions of, the collection and analysis of citizens' data are therefore essential, if the goal is to advance citizen control over data. Data ethics frameworks and institutions can, as our interviewees pointed out, advance interactions between different stakeholders and agendas and help develop consensus towards common policy goals. They cannot, however, replace legislation.

Innovative ways of approaching the **collective**, rather than individual, **dimension of data** have emerged, for example, through the concept of indigenous data sovereignty. Based on the need to both preserve and develop their cultural heritage, traditional knowledge and traditional cultural expressions, indigenous communities have formulated programmes for the right to maintain, control, protect and develop their intellectual property over these and, more broadly, over data that is collected about them (Kukutai & Taylor, 2016). Yet such concepts remain underdeveloped in a European policy context.

Overall, as this analysis shows, the **discourse** of increased citizen control and empowerment is growing and is gaining traction in policy debate. There is an emerging understanding that the policy environment for the data-related activities of platforms (and, by extension, for data collection and analysis more broadly) is insufficient. However the actual **implementation** of citizen control, so far, is subject to significant limitations based on narrow definitions of such control and an expansion, rather than reduction, of data collection and sharing. Citizens are gaining new capabilities due to the GDPR but are also subject to increased monitoring, and the data they have access to and power over remains a limited section of the wider range of data that is now used in the private and public sector. We may be a long way off actual citizen control over data, but we are witnessing openings and new avenues towards that goal.

Bibliography

A29WP. (2016). *Guidelines on the right to data portability. Adopted 13 December 2016*. Available at http://ec.europa.eu/information_society/newsroom/image/document/2016-51/wp242_en_40852.pdf

Andrejevic, M. (2012) Exploitation in the data mine. In C. Fuchs, K. Boersma, A. Albrechtslund and M. Sandoval (Eds.), *Internet and Surveillance: The Challenges of Web 2.0 and Social Media*. pp. 71–88. Abingdon: Routledge.

Andrejevic, M. (2017) To pre-empt a thief. *International Journal of Communication*, 11, 879–96.

Angwin, J., Larson, J., Mattu, S. & Kirchner, L. (2016) Machine Bias. *Pro Publica*, 23rd May 2016. Available at: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Belli, L. & Zingales, N. (2017). *Platform Regulations*. Rio de Janeiro: FGV Direito Rio.

Berry, D. (2011) The computational turn: Thinking about the digital humanities. *Culture Machine*, 12, 1–22. Available at: http://www.culturemachine.net/index.php/cm/article/viewDownloadInterstitial/440/470.

Big Brother Watch (2018) A closer look at Experian big data and artificial intelligence in Durham Police. Blog post, 6th April 2018. Available at: https://bigbrotherwatch.org.uk/2018/04/a-closer-look-at-experian-big-data-and-artificial-intelligence-in-durham-police/

boyd, d. & Crawford, K. (2012) Critical questions for Big Data. *Information, Communication & Society*, 15(5), 662–79.

Browman, C. (2017) Data Localization Laws: an Emerging Global Trend. *Jurist: Legal News and Research*. Available at https://www.jurist.org/commentary/2017/01/Courtney-Bowman-data-localization/

Cabanas, J.G., Cuevas, A., and Cuevas, R. (2018) *Facebook Use of Sensitive Data for Advertising in Europe*. Available at https://arxiv.org/abs/1802.05030

Cheney-Lippold, J. (2017) *We Are Data*. New York: New York University Press.

Coats, D. and Komisarczuk, P. (2016) The Investigatory Powers Act 2016 and Internet Connection Records: Some surprising truths? *Royal Holloway University of London, ISG MSc Information Security thesis series*. Available at https://www.royalholloway.ac.uk/isg/documents/pdf/technicalreports/2017/danielcoatsisg.pdf

Council of Europe. (1950) *European Convention of Human Rights*. Available at: https://www.echr.coe.int/Documents/Convention_ENG.pdf

De Hert, P., Papakonstantinou, V., Malgieri, G., Beslay, L., & Sanchez, I. (2017) The right to data portability in the GDPR: Towards user-centric interoperability of digital services. *Computer Law & Security Review* 34 (2), 193-203.

Dencik,L. & Cable, J. (2017) The Advent of Surveillance Realism: Public Opinion and Activist Responses to the Snowden Leaks. *International Journal of Communication*, 11(2017), 763-781.

Dencik, L., Hintz, A., Redden, J., & Warne, H. (2018) Data Scores as Governance: Investigating Uses of Citizen Scoring in Public Services. Project Report. Cardiff: Data Justice Lab.

Dixon, P. & Gellman, R. (2014) The Scoring of America: How Secret Consumer Scores Threaten Your Privacy and Your Future. Report for the World Privacy Forum.

Draper, N. and Turow, J. (2017) Toward a sociology of digital resignation. Paper presented at *Data Power*, 23 June 2017, Ottawa.

Edwards, L. & Veale, M. (2017) Slave to the algorithm? Why a 'right to an explanation' is probably not the remedy you are looking for. *Duke Law and Technology Review*, 16(1), 18–84.

Eubanks, V. (2018) *Automating inequality: How high-tech tools profile, police, and punish the poor*. New York: St Martin's Press.

Fishenden, J. (2016) Sharing of citizen data. *Ntouk blog*. Available at https://ntouk.wordpress.com/2016/10/18/the-digital-economy-bill-part-5-data-sharing/

Gangadharan, S.P., Eubanks, V. & Barocas, S. (eds.) (2015) Data and Discrimination: Collected Essays. Open Technology Institute, New America. Available at: http://newamerica.org/downloads/OTI-Data-an-Discrimination-FINAL-small.pdf

Graef, I., Valcke, P., & Verschakelen, J. (2013) Putting the right to data portability into a competition law perspective. *Law: The Journal of the Higher School of Economics*, Annual Review.

Greenfield, P. (2018) The Cambridge Analytica Files. *The Guardian*. Available at https://www.theguardian.com/news/2018/mar/26/the-cambridge-analytica-files-the-story-so-far

Greenwald, G. (2014) *No Place to Hide: Edward Snowden, the NSA and the surveillance state*. London: Hamish Hamilton.

Hellerstein, J. (2008) The Commoditization of Massive Data Analysis. *Radar*, 19 November 2008. Available at: http://strata.oreilly.com/2008/11/the-commoditization-of-massive.html

Hintz, A. & Brown, I. (2017) Enabling digital citizenship? The reshaping of surveillance policy after Snowden. *International Journal of Communication*, 11, 782–801.

Hintz, A., Dencik, L. & Wahl-Jorgensen, K. (2018) *Digital Citizenship in a Datafied Society*. Cambridge: Polity Press.

ICO (2018) Lawful basis for processing Legitimate Interests. Available at https://ico.org.uk/media/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/legitimate-interests-1-0.pdf

ISPA (2016) Internet industry outlines major concerns ahead of the 2nd Reading of the Investigatory Powers Bill. Available at https://www.ispa.org.uk/internet-industry-outlines-major-concerns-ahead-of-the-2nd-reading-of-the-investigatory-powers-bill/

Kaltheuner, F. and Bietti, E. (2017) Data is power: Towards additional guidance on profiling and automated decision-making in the GDPR. *Journal of Information Rights, Policy and Practice, 2(2).* Available at https://jirpp.winchesteruniversitypress.org/articles/abstract/10.21039/irpandp.v2i2.45/

Kingdon, J.W. (1984) *Agendas, alternatives, and public policy*. Boston, MA: Little Brown.

Kitchin, R. (2014) *The data revolution*. London: Sage.

Kitchin, R. (2017) Thinking critically about and researching algorithms. *Information, Communication & Society*, 20(1), 14–29.

Kukutai, T. and Taylor, J., eds. (2016) *Indigenous Data Sovereignty: Toward an Agenda*. Acton: ANU Press.

Lyon, D. (2015) *Surveillance after Snowden*. Cambridge: Polity.

Massumi, B. (2015) *Ontopower: War, Powers, and the State of Perception*. Durham, NC: Duke University Press.

Mayer-Schönberger, V. & Cukier, K. (2013) *Big Data: A Revolution That Will Transform How We Live, Work and Think*. New York: John Murray.

Malcolm, J. (2016) New Censorship and Copyright Restrictions in UK Digital Economy Bill. *Electronic Frontier Foundation*, 2016, July 8. Available at: https://www.eff.org/deeplinks/2016/07/new-censorship-and-copyright-restrictions-uk-digital-economy-bill

McCann, D., Hall, M. & Warin, R. (2018) Controlled by calculations?: Power and accountability in the digital economy. New Economics Foundation, report published 29th June 2018. Available at: https://neweconomics.org/2018/06/controlled-by-calculations

McIntyre, N. & Pegg, D. (2018) Councils use 377,000 people's data in efforts to predict child abuse. *The Guardian*, 16th September 2018. Available at: https://www.theguardian.com/society/2018/sep/16/councils-use-377000-peoples-data-in-efforts-to-predict-child-abuse

Metcalfe, P. & Dencik, L. (2019) The politics of big borders: Data (in)justice and the governance of refugees. First Monday, 24(4). Available at: https://firstmonday.org/ojs/index.php/fm/article/view/9934/7749

O'Neil, C. (2016) *Weapons of math destruction: How big data increases inequality and threatens democracy*. Penguin.

Open Rights Group (2016a). *Digital Economy Bill: Briefing to the House of Commons on Second Reading*. Available at: https://www.openrightsgroup.org/ourwork/reports/digital-economy-bill-briefing-to-the-house-of-commons-on-second-reading

Open Rights Group (2016b). *A database of the UK's porn habits. What could possibly go wrong?* Available at: https://www.openrightsgroup.org/blog/2016/a-database-of-the-uks-porn-habits-what-could-possibly-go-wrong

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

Rainie, L., & Anderson, J. (2017) The Fate of Online Trust in the Next Decade. *Pew Research Center*, August 10, 2017. Retrieved from: http://www.pewinternet.org/2017/08/10/the-fate-of-online-trust-in-the-next-decade/

Redden, J. & Brand, J. (2018) Data Harm Record. Available at: https://datajusticelab.org/data-harm-record/

Ryan, J. (2017) How the GDPR will disrupt Google and Facebook. *Pagefair*, August 30, 2017. Available at: https://pagefair.com/blog/2017/gdpr_risk_to_the_duopoly/

Ryan, J. (2018a) GDPR consent design: How granular must adtech opt-ins be? *Pagefair*, January 8, 2018. Available at: https://pagefair.com/blog/2018/granular-gdpr-consent/

Ryan, J. (2018b) Facebook and adtech face a turbulent time in Europe's courts: the Brussels case. *Pagefair*, May 3, 2018. Available at: https://pagefair.com/blog/2018/facebook-brussels-case/

Srnicek, N. (2016) *Platform Capitalism*. Cambridge: Polity.

Tucker, P. (2016) Refugee or Terrorist? IBM thinks its software has the answer. *Defense One*. Available at: http://www.defenseone.com/technology/2016/01/refugee-or-terrorist-ibm-thinks-its-software-has-answer/125484/

Van Dijck, J. (2014) Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & Society*, 12(2), 197–208.

Van Dijk, J., Poell, T., & de Waal, M. (2018) *The Platform Society*. Oxford: Oxford University Press.

Vanberg, A.D., & Unver, M.B. (2017) The right to data portability in the GDPR and EU competition law: odd couple or dynamic duo? *European Journal of Law and Technology*, 8(1).

Veale, M. & Edwards, L. (2017) Clarity, surprises, and further questions in the Article 29 Working Party draft guidance on automated decision-making and profiling. *Computer Law & Security Review*, 34(2), 398-404.

Wachter, S., Mittelstadt, B. & Floridi, L. (2017) Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. *International Data Privacy Law*, 7(2), 76–99.

Warrell, H. (2015) Students under surveillance. *Financial Times*. Available at: https://www.ft.com/content/634624c6-312b-11e5-91ac-a5e17d9b4cff

Wizner, B. (2017). What changed after Snowden? A US perspective. *International Journal of Communication*, 11, 897–901.

Zuboff, S. (2019) *The Age of Surveillance Capitalism*. London: Profile Books.