Downstream Consent: A Better Legal Framework for Big Data

Andrew Nicholas Cormack*

1. Introduction

'Big data' is characterised by computer scientists as presenting the challenge of four V's – volume, variety, velocity, veracity.[1] For lawyers the challenge is that it involves data-driven, rather than hypothesis-driven, investigation. Whereas traditional research proposes a hypothesis and then collects evidence to confirm or refute it, big data approaches often start with collection of evidence and then use correlations to identify candidate hypotheses for further investigation. Although the term 'big data' is often applied to any investigation using large datasets, here the focus is specifically on this data-driven aspect. Both research ethics and data protection law apply their principal controls at the hypothesis stage – through ethical review and user consent respectively. This paper considers how data protection law can deal with an approach where data are collected and processed long before it is possible to obtain valid, informed, consent.

The paper first reviews how data protection law handles research on personal data – using either informed consent or the research exemption – and why these approaches are unsatisfactory for data-driven analysis. An alternative model is proposed – dividing big data processing into pattern-finding and pattern-matching stages – which provides a better correspondence between what the law requires and how investigations are performed in practice. This leads to a framework using provisions of the current Data Protection Directive – though not those normally invoked for big data – which can draw on existing guidance on purpose limitation, legitimate interests and consent to provide better information for individuals, richer guidance for those wishing to use data-driven techniques responsibly, and more opportunities for regulators to ensure appropriate behaviour. Though based in the 1995 Directive, the framework also reflects a more recent trend, notably in the new General Data Protection Regulation, that seeks to make data controllers more accountable for their decisions why and how to process personal data. Conducting data-driven analysis within such a framework should reduce fears that it may be used in unethical or harmful ways, increase confidence in the technique, and allow its benefits to be delivered to individuals, organisations and society. Finally the limits of the model are examined, concluding that unfettered use of data-driven techniques may in fact be socially, not just legally, unacceptable. If so, the proposed framework can help organisations and society to identify the boundaries of acceptable big data use.

2. The Legal Challenge of Big Data

In traditional research, the objective of the study is the first thing to be identified. Where the research will involve human subjects, ethics codes such as the Nuremberg Code[2] and Menlo

* Chief Regulatory Advisor, Jisc Technologies

[1] According to IBM, 'The Four V's of Big Data' <http://www.ibmbigdatahub.com/infographic/four-vs-big-data> accessed 9 Aug 2016, though other sources differ on the fourth V.

[2] Originally from *Trials of War Criminals Before the Nuremberg Military Tribunals Under Control Council Law No. 10* Vol. 2, Nuremberg, October 1946 - April 1949. (Washington, DC: US Government Printing Office, 1949). pp 181-182. Available at <https://history.nih.gov/research/downloads/nuremberg.pdf> accessed 20 May 2016.

Report[3] insist that the objective and methods be assessed and approved before any data gathering takes place. This assessment must consider the consequences of all possible outcomes, whether the researchers' hypothesis is supported or contradicted. Under the UK Data Protection Act two different approaches can be adopted. For some studies, especially those involving health data, the consent of each research subject will be sought for data collection and processing, providing the same information on methods and consequences as the ethical review to ensure that consent is fully informed. For studies using data that have already been obtained for other purposes, consent may not be required provided there are safeguards to ensure the additional processing is neither likely to cause substantial damage or distress to any data subject, nor be 'used to support measures or decisions with respect to particular individuals'.[4]

When big data involves processing or collecting data in order to identify possible hypotheses, neither of these approaches can be used. At the time of data collection, the possible results are unknown: Strandburg considers it a 'legal fantasy' that individuals can assess the benefits and costs, as is required if they are to give valid consent.[5] Barocas and Nissenbaum note that analysis that is not constrained by a prior hypothesis 'may discover correlations neither sought nor anticipated', making it impossible to perform the harm assessment required by the Data Protection Act's research exemption.[6] And in many big data applications – from drug side-effects to student support to personalised advertising – there will be a desire to use the outcome to benefit individual research subjects. Whereas in the past the duration of research meant it could only affect future cohorts or generations, big data may offer results fast enough to support the same individuals as it studies.

Not only do traditional 'consent' and 'research' approaches provide little practical control to individual subjects of big data processing, they also fail to guide processors on how to conduct data-driven research and they give regulators few tools to manage it. When using consent, researchers are required to tell subjects what they plan to do with their data. Once the individual has granted permission, the fair processing principle requires that researchers do what should have been expected from that notice, while the security principle requires that data be kept safe. But none of these provide any guidance on what the researcher should ask the experimental subject: consent takes the view that if the subject agrees then the request must have been acceptable. Enthusiastic investigators might well be tempted to see what they can get away with. Likewise consent only lets regulators intervene if the privacy notice was unclear or inaccurate, not if the proposed processing was inappropriate, either in individual or societal terms. Conversely the research exemption, by prohibiting activities whose outcome cannot be foreseen, effectively prohibits data-driven approaches entirely.

The same big data techniques can support uses of data that are highly beneficial to individuals and society:

---

[3] 'The Menlo Report: Ethical Principles Guiding Information and Communiation Technology Research' (August 3rd 2012) <https://www.dhs.gov/sites/default/files/publications/CSD-MenloPrinciplesCORE-20120803.pdf> accessed 9 Aug 2016.

[4] Data Protection Act 1998, s.33.

[5] Katherine J Strandburg, 'Monitoring, Datafication and Consent: Legal Approaches to Privacy in the Big Data Context' in Julia Lane et al (eds), *Privacy, Big Data and the Public Good* (Cambridge University Press 2014) 8.

[6] Solon Barocas and Helen Nissenbaum, 'Big Data's End Run Around Anonymity and Consent' in Julia Lane et al (eds), *Privacy, Big Data and the Public Good* (Cambridge University Press 2014) 60.

> scientists can use new forms of data to do research that improves the lives of human beings; federal, state and local governments can use data to improve services and reduce taxpayer costs; and public organisations can use information to advocate for public causes, for example[7];

as well as those uses that, accidentally or intentionally

> perpetuate existing prejudices and stereotypes … aggravate the problems of social exclusion and stratification … increase the economic imbalance between large corporations and consumers … result in other significant adverse impacts on individuals[8].

Clearly a more nuanced approach is needed to help data users, data subjects and regulators distinguish the beneficial uses from the harmful ones.

3. An Alternative Legal Framework

Article 7 of the Data Protection Directive offers six different justifications for processing personal data: consent, contract, legal obligation, vital interests, public interest and legitimate interests of the controller.[9] Rather than treating 'big data' as a single process requiring the same legal treatment throughout the stages of data collection, data processing and individual intervention, separating these three processes and applying the appropriate justification to each provides a much better match between law and practice. The resulting framework can then use existing analysis of these justifications – including by the Article 29 Working Party and the European Data Protection Supervisor – to provide much more guidance on when and how data-driven techniques ought to be used.

Most big data applications involve extracting new insights from 'data exhaust': the information we already create in our daily interactions with organisations and systems.[10] For example a history of credit card use can provide valuable indicators of fraudulent use. The sequence of web pages visited by an IP address may allow a site owner to identify where their navigation cues are unhelpful or, more worryingly, the price that an individual is likely to be prepared to pay for particular goods or services.[11] In each case the collection of the information is already justified and subject to law and guidance: whether as necessary for a contract, a legal duty, or the legitimate interest of the processor to identify and investigate misuse of their systems. Provided the information collected is, indeed, necessary for those purposes (explained by the Article 29 Working Party as meaning no 'less invasive means are available to serve the same end'[12]), its collection is justified. This 'necessity' test provides helpful guidance for processors, regulators and users: information that is not necessary for the declared purposes should not be collected.

---

[7] Julia Lane et al (eds), *Privacy, Big Data and the Public Good* (Cambridge University Press 2014) xi.

[8] Article 29 Data Protection Working Party, *Opinion 03/2013 on purpose limitation* (00569/13/EN WP 203) 45.

[9] Data Protection Directive, Article 7.

[10] European Data Protection Supervisor, *Opinion 7/2015: Meeting the Challenges of Big Data* 10.

[11] Kevin Peachey, 'How technology opens the door to personalised pricing' (BBC News, 21 November 2012) <http://www.bbc.co.uk/news/business-20396091> accessed 9 Aug 2016.

[12] Article 29 Data Protection Working Party, *Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC* (844/14/EN WP 217) 29.

Data-driven techniques are commonly used to re-process existing information, for example to identify possible improvements in the service or other patterns. This is hard to justify as strictly 'necessary' for the original contract or legal obligation. However the Working Party note that it may be a 'legitimate interest' of the organisation.[13] The Data Protection Directive[14] and the Working Party's Opinion[15] on this justification again provide useful guidance. Where personal data are processed for legitimate interests, there must be a clearly stated purpose, the processing must be necessary for that purpose, the impact and risk for the individuals whose data are processed must be minimised, and any remaining impact or risk must be justified by a balancing test against the claimed interest. Interests, even though legitimate, cannot justify processing that involves an inappropriate risk to the individuals whose data are processed. The requirements to identify a purpose, ensure its legitimacy and demonstrate the necessity of processing to its achievement provide guidance to responsible big data users on what they should apply the technique to. The requirements to minimise harm and ensure the benefits are justified guide them on how, and whether, to conduct their investigations. All five requirements provide regulators with means to guide, and if necessary enforce, appropriate behaviour. Within these constraints it should be possible to examine data, identify patterns and consider how these might be used to help individuals, the organisation, or society.

Legitimate interests cannot, however, be used to justify any activity where the intention is to personalise a service or otherwise affect individual users, since this would contradict the requirement that the impact on individuals be minimised. Once the organisation has identified patterns in data that enable it to identify and design such an intervention, however, it should also have sufficient information to seek valid consent from those individuals who may be affected by it. Whereas at the time the data were collected the results of data-driven analysis and their consequences could not be foreseen or explained to individuals, now they can. Consent can now be fully informed. Offering a choice between personalised and generic versions of the service should increase the likelihood that consent to personalisation is freely given.[16]

This approach divides what is commonly described as a monolithic 'big data' approach into two stages: analysis, which aims to find patterns in data, and intervention, which uses those patterns to identify and affect relevant individuals. Analysis, conducted under conditions that ensure impact on data subjects is minimised, can produce results indicating that people matching a particular pattern might benefit from a specified action. Intervention, with the aim of maximising impact, identifies the individuals that match the pattern and invites them to consent to the action. When discussing safeguards, the Article 29 Working Party recognise these as distinct types of processing:

> In order to identify what safeguards are necessary, it may be helpful to make a distinction between two different scenarios. In the first one, the organisations

[13] *Ibid*, 25.

[14] Data Protection Directive, Article 7(f).

[15] Article 29 Data Protection Working Party, *Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC* (844/14/EN WP 217).

[16] Compare the European Data Protection Supervisor's discussion of a choice between free and paid services in *Opinion 5/2016 on the Review of the ePrivacy Directive* 16.

> processing the data want to detect trends and correlations in the information. In the second one, the organisations are interested in individuals.[17]

For some big data applications, the 'trends and correlations' resulting from the analysis stage will be all that is needed, for example to suggest how an overall service might be improved. Here no intervention stage is required: the results of the analysis stage can be implemented to general, rather than specific, benefit. In other cases, the organisation will want to use information from the correlations to change the way they treat particular individuals. Here the two types of processing identified by the Working Party will occur in succession.

Something like this separation seems to have been foreseen by the Working Party in their 2011 Opinion on Consent. As an alternative to obtaining consent at the time of data collection, they suggest:

> consent can also be requested "downstream", when the purpose of the processing changes. In this case the information to be provided will have to focus on what is needed in the specific context, in relation to the purpose.[18]

This appears to recognise the opportunity to provide more, and more specific, information by postponing the request for consent until the time a specific intervention is proposed.

The European Data Protection Supervisor also notes the value of separating the processes of 'detect[ing] trends and correlations' and 'directly applying any insights … to the individuals concerned'. By ensuring that individuals can specifically authorise individual interventions:

> 'functional separation' [by way of technical and organisational measures] may potentially play a role in reducing the impact on the rights of individuals, while at the same time allowing organisations to take advantage of secondary uses of data.[19]

In terms of transparency, 'downstream consent' also supports Barocas and Nissenbaum's proposal that requests for consent be used to highlight a departure from 'norms, standards and expectations'.[20] In this formulation the norm, standard and expectation is that processing of personal data will be conducted in ways that minimise the impact on the individual, as required by the legitimate interests justification used for the analysis stage. By seeking consent for a particular personalised intervention the data controller highlights both the proposed change to an impact-maximising intention and, by defining the specific scope of the consent, the boundary outside which processing will continue to be subject to the impact-minimising norm. Consent is thereby being sought at the time when most information about the proposed intervention is available but least impact has been caused by it. By granting consent the individual agrees both to the – presumably mutually-beneficial – impact-maximising intervention and to its boundaries. This suggests that downstream consent could have wider benefits than in the Working Party's example: agreeing the change from impact-minimising to impact-maximising is important even when the purpose of processing does not change.

---

[17] Article 29 Data Protection Working Party, *Opinion 03/2013 on purpose limitation* (00569/13/EN WP 203) 46.

[18] Article 29 Data Protection Working Party, *Opinion 15/2011 on the definition of consent* (01197/11/EN WP 187) 19.

[19] European Data Protection Supervisor, *Opinion 7/2015: Meeting the Challenges of Big Data* 15.

[20] Solon Barocas and Helen Nissenbaum, 'Big Data's End Run Around Anonymity and Consent' in Julia Lane et al (eds), *Privacy, Big Data and the Public Good* (Cambridge University Press 2014) 65.

If a government wished to make some mandatory, rather than consensual, intervention based on information obtained from data-driven analysis then it could either legislate, making the subsequent data processing 'necessary for compliance with a legal obligation',[21] or else demonstrate that the intervention and processing were 'necessary for the performance of a task carried out in the public interest'.[22] Both of these justifications permit impact-maximising actions, so could take the place of consent as the basis for the intervention stage of the big data framework. However before either legislation or a public interest activity is initiated there should be a clear understanding of the likely impact on both individuals and society. This understanding should be obtained using an impact-minimising analysis stage subject to the additional controls of the legitimate interests justification. However the new General Data Protection Regulation may create a paradox as Article 6(1) states that legitimate interests 'shall not apply to processing carried out by public authorities in the performance of their tasks'. One alternative would be to use the public interest justification for the analysis stage but limit this to a specific purpose and apply additional controls to ensure impact minimisation and balancing of interests. Legislating before the analysis stage, to make that stage necessary for compliance with a legal obligation, would seem to risk a law that is either unacceptably wide, or that may not cover a desirable outcome once it is identified.

Finally, whether in the public or private sector, if the analysis stage identifies a possible purpose that was not envisaged and stated at the time of data collection then the purpose limitation rule must be applied to determine whether the analysis may continue. Some new purposes may be found to be compatible with those originally declared, but continuing processing for an incompatible purpose will be unlawful.[23] In this case the organisation may either seek the individual consent of existing data subjects to process their data for the new purpose, or else offer the new purpose to new data subjects when their data are first collected. In either case the agreement of data subjects only authorises the analysis stage: specific consent will still be required for any subsequent interventions.

This framework considers a typical big data application as a series of activities, each with a different objective and subject to appropriate legal rules. Data collection is done for some other primary purpose and justified by the necessity of obtaining data for that purpose. Legitimate secondary purposes, for which data-driven techniques will be applied, are stated at this time. Analysis, to identify patterns in the data, must be done in ways that minimise the risk to individuals and subject to a balance of interests test; it should be done under rules applying to the legitimate interests justification, whether formally that justification or, in the public sector, public interest is actually used. Intervention, to apply the insights gained from a pattern to a particular individual, should occur only after obtaining the individual's free, informed consent. Public sector interventions may alternatively be justified by legislation or public interest.

4. Benefits of the Framework

Separating the pattern-finding and pattern-matching stages of the big data approach, and applying the appropriate legal justifications to each, provides benefits for individuals, organisations using big data techniques, and regulators. All three should obtain greater confidence that these techniques can be used safely, and that processing that creates unacceptable risks can be detected and terminated quickly.

[21] Data Protection Directive, Article 7(c).

[22] *Ibid*, Article 7(e).

[23] Article 29 Data Protection Working Party, *Opinion 03/2013 on purpose limitation* (00569/13/EN WP 203) 21.

The framework should also satisfy the requirements of the new General Data Protection Regulation, in particular helping data controllers comply with its stricter rules on consent. By making controllers more accountable for their use of big data techniques it conforms to recent trends in regulatory thinking.

4.1 Benefits for Individuals

For individuals, the approach provides greater transparency and control over use of their personal data. At the time when data are collected, organisations will be required to declare both the primary purpose for which the data are necessary and any secondary purposes for which they claim a legitimate interest. According to the Data Protection Directive such purposes must be 'specified, explicit and legitimate':[24] sufficiently detailed, according to the Article 29 Working Party, to 'delimit the scope of the processing operation' and allow necessary privacy safeguards to be identified.[25] Vague, general statements of purpose, insufficient to identify the risks involved and safeguards required, should not be permitted.

Whatever secondary purposes are claimed, the use of legitimate interests rather than consent requires that analysis be conducted in ways that minimise the risk of impacting individuals, and subject to a continuing balancing test of interests. Individuals have a legal right to object 'on compelling legitimate grounds relating to [their] particular situation'; if the objection is 'justified' then the organisation must cease processing those data.[26] The European Data Protection Supervisor suggests that organisations might go further and offer an 'unconditional opt-out.'[27] Everyone whose personal information is processed should be reassured by the rule that any data-driven pattern-finding will be done in ways that minimise any likely impact, and then only if that remaining impact can be objectively justified.

Since the aim of data-driven analysis is to reveal unexpected patterns, the risks and balance of interests may also change unexpectedly during the course of analysis. For example algorithms may identify patterns that relate to small groups of individuals, increasing the risk of accidental or deliberate re-identification. Organisations performing analysis must ensure that if such changes mean that the interests of individuals do now override the legitimate interest of the organisation, then those analyses should not continue. Similarly if a pattern suggests an insight that is not compatible with the purposes stated when data were collected, that pattern and insight may not be acted upon unless the individual agrees.

Under the framework, organisations must conduct their data-driven analyses in ways that minimise the impact on individuals. This norm can only change when the individual agrees to waive it, in favour of a specific intervention that is intended to maximise impact on the individual in some mutually-beneficial way. This request for consent gives the individual a clear signal when the intention of processing changes. Since agreeing to the change will give the organisation permission to seek to maximise impact, individuals will be rightly suspicious if the consequences of this change are not clearly explained, or if the scope of the requested waiver is unreasonably vague or widely drawn.

Furthermore, as the analysis will by now have identified patterns and proposed interventions, organisations can be expected to provide much more detailed information on their proposed

[24] Data Protection Directive, Article 6(1)(b).

[25] Article 29 Data Protection Working Party, *Opinion 03/2013 on purpose limitation* (00569/13/EN WP 203) 12.

[26] Data Protection Directive, Article 14(a).

[27] European Data Protection Supervisor, *Opinion 7/2015: Meeting the Challenges of Big Data* 11.

action and its possible consequences than would have been possible at the time of data collection. Requests for consent must be specific to the proposed intervention: any other processing of personal data that is not covered by either the consent or the original primary service purpose should continue to follow an impact-minimisation approach.

Finally, separating the delivery of the primary service from consent to individual secondary interventions makes it harder for organisations to compel consent as a condition of service. Individuals should not feel under pressure to give consent unwillingly – to an intervention they do not want, or one that has not been sufficiently explained – for fear of losing the service they originally signed up to.

4.2 Benefits for Organisations

For organisations wishing to use big data techniques, combining the requirements of the necessity, legitimate interests and consent justifications provides guidance on responsible conduct throughout processing.

To be collected and available for data-driven analysis, information must be necessary for the purpose of the original interaction between the organisation and the individual. Organisations must specify the secondary purposes to which data-driven analysis may be directed, ruling out completely unconstrained searches for unexpected correlations. However the Article 29 Working Party recognise both the 'detection of trends and correlations'[28] and the 'concept of an overall purpose, under whose umbrella a number of separate processing operations take place'.[29] The UK Information Commissioner suggests that 'a privacy notice can allow for development in the way you use personal data, whilst still providing individuals with enough detail for them to understand what you will do with their information'.[30] The need to declare purposes in advance need not be a barrier to big data techniques, so long as the data controller has some clear objective in mind. Indeed it appears that a valid statement of purpose can be wider than a valid consent notice: the latter must give the individual 'an appreciation and understanding of the facts and implications of an action'[31] (as discussed above, likely to be impossible before data-driven analysis has commenced) whereas a purpose statement need only be sufficient to delimit the scope of the operation and identify appropriate safeguards.[32]

For the analysis stage the requirement that the interests served be legitimate can inform organisations which analyses should be performed; the requirement of necessity and the balancing test provide rules on how they should be conducted. The Article 29 Working Party's Opinion on Legitimate Interests[33] provides guidance on these questions and on the procedural and technical safeguards that should be applied. The ongoing requirement to minimise impact should help to identify, for example, circumstances when a pattern relates to too small a number of individuals to be investigated without their consent; the requirement to

---

[28] Article 29 Data Protection Working Party, *Opinion 03/2013 on purpose limitation* (00569/13/EN WP 203) 35.

[29] *Ibid*, 16.

[30] Information Commissioner, 'Privacy Notices, Transparency and Control' <https://ico.org.uk/about-the-ico/privacy-notices-transparency-and-control/fairness/> accessed 9 Aug 2016.

[31] Article 29 Data Protection Working Party, *Opinion 15/2011 on the definition of consent* (01197/11/EN WP 187) 19.

[32] Article 29 Data Protection Working Party, *Opinion 03/2013 on purpose limitation* (00569/13/EN WP 203) 12.

[33] Article 29 Data Protection Working Party, *Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC* (844/14/EN WP 217).

balance interests can identify discriminatory or otherwise harmful correlations and help organisations avoid activities that will damage trust in how they handle personal data.

Using legitimate interests, rather than consent, for the analysis stage may also have practical benefits for the accuracy of results. When statistical analysis relies on information provided by consenting users there is a risk of bias due to self-selection – that the users who consent will not be representative of the overall population. Processing a more complete dataset under impact-minimising rules can therefore provide both more accurate results and better safeguards for individuals. Where analyses will be used to inform public policy, rather than individual interventions, open debate may be facilitated by explicitly minimising the risk of harm and excluding the possibility of intervention.[34] Indeed it seems likely that many people would be willing to allow their data to be used for public benefit if they were guaranteed a separate choice whether or not to accept any individual impact.

Where individual intervention is proposed, postponing the request for consent lets organisations provide clear and precise information about their proposed action and its consequences for the individual. With growing concern about 'opaque privacy policies, which encourage people to tick a box and sign away their rights',[35] organisations that are specific about the consents they seek are likely to be more trusted. Unlike consent at time of collection, being precise in this consent request does not constrain all future investigations: that boundary continues to be provided by the original statement of purpose.

Indeed declaring additional purposes as part of a service offering might give organisations an opportunity to distinguish different types of services, or to compete on the basis of the secondary purposes they offer. For example the Working Party discuss a hypothetical storecard that offers customers a choice between generic and personalised discounts;[36] on-line services from apps to magazines already let users choose between paying for their subscription or receiving adverts (a secondary purpose, increasingly based on big data techniques). With stores now offering to share the financial proceeds of a transaction with selected charities, in future we might even see competition based on socially beneficial uses of informational proceeds. Such markets in purposes might themselves create new business opportunities serving consumers and groups that prefer either more or less information re-use than average. They could also, as discussed below, provide a mechanism for determining the acceptable limits of big data techniques.

4.3 Benefits for Regulators

Any regime seeking to regulate secondary processing of data must be enforced. Once individuals have provided information to an organisation for a primary purpose (whether to receive a service, comply with a legal duty, or otherwise) there is no practical measure they can take to control whether and how the organisation might use that data for secondary purposes. Their only remedy against unwanted processing is through the law. Neither consent nor the framework proposed here can control secondary uses of big data unless data controllers are regulated. However, whereas regulation of consent concentrates on the information given to individuals, the framework presented here offers considerably more

[34] Katherine J Strandburg, 'Monitoring, Datafication and Consent: Legal Approaches to Privacy in the Big Data Context' in Julia Lane et al (eds), *Privacy, Big Data and the Public Good* (Cambridge University Press 2014) 33.

[35] European Data Protection Supervisor, *Leading by Example: the EDPS Strategy 2015-2019* 11.

[36] Article 29 Data Protection Working Party, *Opinion 03/2013 on purpose limitation* (00569/13/EN WP 203) 61.

scope for regulators to develop, and if necessary enforce, rules on how personal data are actually processed.

Enforcement against data controllers claiming the consent justification has concentrated almost entirely on privacy notices, ensuring these provide 'clear, unambiguous and comprehensive information regarding the data processing'.[37] This is because the consent justification gives regulators little scope to question the acts of processing themselves, so long as they were described accurately and thus satisfy the fairness requirement that processing be what the data subject 'would reasonably expect'.[38] There seem to have been very few instances of regulators questioning if even clearly-described processing might still be unfair or whether free consent was actually possible in the circumstances.[39] In contrast the framework proposed here for big data techniques lets regulators assess whether the purpose of processing is legitimate, whether that purpose can be achieved in a less invasive way, whether appropriate safeguards have been adopted and whether the remaining risk to individuals is justified by the benefit to other interests.[40] Each of these questions provides opportunities for regulators and big data users to work together to develop guidance on responsible practice.

4.4 Future Data Protection Law

Early data protection law, including the 1995 Data Protection Directive, concentrated on enabling individuals to control the use of their own personal data.[41] More recent discussion has concerned those who direct the processing of personal data, with Cate and Mayer-Schönberger reporting a broad view that 'new approaches must shift responsibility away from data subjects towards data users, and towards a focus on accountability for responsible data stewardship'.[42] In big data in particular, the European Data Protection Supervisor sees a need to make data controllers more responsible for why and how they process personal data.[43]

The General Data Protection Regulation, which will come into force in 2018, reflects this trend by making the limits of consent explicit. In particular consent will not be valid if the individual 'has no genuine or free choice' (Recital 42) or there is 'a clear imbalance between the data subject and the controller' (Recital 43). In many big data applications – for example those involving governments, dominant service providers or observed data – these tests may well raise doubts whether legally valid consent to a secondary purpose can be obtained at all.

---

[37] Article 29 Data Protection Working Party, *Appendix to letter to Google 23rd September 2014* (Ref. Ares(2014)3113072 – 23/09/2014) <http://ec.europa.eu/justice/data-protection/article-29/documentation/other-document/files/2014/20140923_letter_on_google_privacy_policy_appendix.pdf> accessed 9 Aug 2016.

[38] Information Commissioner, 'Processing personal data fairly and lawfully (Principle 1)' <https://ico.org.uk/for-organisations/guide-to-data-protection/principle-1-fair-and-lawful/#fair-processing> accessed 9 Aug 2016.

[39] E.g. Judith Rauhofer, 'Of Men and Mice: Should the EU Data Protection Authorities' Reaction to Google's new Privacy Policy Raise Concern for the Future of the Purpose Limitation Principle' (2015) 1(1) European Data Protection Law Review 15.

[40] Data Protection Directive, Article 7(f).

[41] Office of the Privacy Commissioner of Canada, 'Consent and Privacy' <https://www.priv.gc.ca/information/research-recherche/2016/consent_201605_e.pdf> accessed 9 Aug 2016 4.

[42] Fred Cate and Viktor Meyer-Schönberger, 'Tomorrow's Privacy: Notice and consent in a world of Big Data' (2013) 3(2) International Data Privacy Law 69.

[43] European Data Protection Supervisor, *Opinion 7/2015: Meeting the Challenges of Big Data* 15.

The new requirements for withdrawing of consent to be as easy as granting it, and for withdrawal to be put into effect at any time (Article 7(3)) may also be challenging. By making the consent justification harder to use, the Regulation encourages the use of alternative justifications that place more responsibility on the data controller.

Although most of the other justifications are unaltered between the Directive and Regulation, a change in the rules for opt-outs hints at increased support for the use of legitimate interests. Under Article 14(a) of the Directive, processing must cease 'where there is a justified objection'. Article 21(1) of the Regulation instead considers whether the controller can 'demonstrate[] compelling legitimate grounds' for continuing processing after receiving an objection. This appears to recognise that strong safeguards may be a valid alternative to stopping important processing entirely.

The 'downstream consent' framework presented here already reflects this change in emphasis, requiring data controllers to behave responsibly when using data-driven techniques to identify patterns. Consent is only used for individual interventions, where the Regulation's new conditions are most likely to be fulfilled. Data collection and processing are conducted under other justifications, principally legitimate interests which places the strongest obligations and responsibility on the data controller. With the Regulation taking 'utmost account' of whether consent was claimed as a condition of service,[44] the framework's separation of service provision from any subsequent consent process for intervention supports good practice.

A consent regime expects individuals, by granting or withholding consent, to control what processing takes place: the data controller's duty is merely to describe their proposal accurately and conduct any processing securely. By contrast, legitimate interests requires organisations to analyse their proposed activities, assess the risks and benefits, and only offer them to individuals if the balancing test is satisfied. 'The user agreed to it' can no longer be an excuse for irresponsible conduct.

5. Exploring the Limits of Big Data

The framework described here does not cover all the possible applications for which big data techniques have been proposed. In particular it presumes that existing data are being re-processed, and for some clearly defined purpose. This section considers situations where those presumptions are not met.

Big data techniques are commonly used to extract new information from data that the organisation has already collected for some other purpose. For example this covers most of the potential applications in city management identified by Goerge.[45] If an organisation wished to collect new information solely for data-driven analysis it could either make its own observations or else invite individuals to provide information voluntarily. Under data protection law the latter would be based on individuals' informed consent; if observations constitute personal data they would require some legally-recognised permission or duty. Either approach is likely to require the purpose of collection and processing to be defined in advance, probably more precisely than is required when notifying a secondary purpose of processing existing data. Observation of personal data might sometimes constitute a legitimate interest of the collector, but the Information Commissioner's Code of Practice on

[44] General Data Protection Regulation, Recital 43 and Article 7(4).

[45] Robert M Goerge, 'Data for the Public Good: Challenges and Barriers in the Context of Cities'in Julia Lane et al (eds), *Privacy, Big Data and the Public Good* (Cambridge University Press 2014) 153.

WiFi Location Analytics warns that this, too, requires a clear purpose, strong safeguards and a detailed analysis of the impact on privacy.[46] Once data have been collected under these more restrictive conditions, the same analysis and intervention framework could apply to subsequent processing.

The framework requires that clear purposes be identified and declared before big data collection or analysis may begin. This has been seen as placing a limit on the potential of big data to discover unanticipated correlations in unexpected fields.[47] However, this limit appears to be an inherent social one, rather than an artificial legal constraint. According to Nissenbaum,[48] all information flows take place within particular social contexts. Each context involves a set of more or less formal context-relative informational norms linking context, actors, information types and transmission principles.[49] Contexts such as doctor-patient, manager-employee or friend-friend set very different expectations as to what information will be transferred, in which directions, and what purposes it will be used for.[50] Any transfer or use that is unexpected within the current context is likely to breach the 'contextual integrity' and be perceived as a *prima facie* violation of privacy. This need not exclude all possibility of change: Nissenbaum suggests that some such breaches may nonetheless be accepted as morally legitimate, so long as they advance the core values of the context.[51] However, this still requires the new processing to relate to the originally understood purposes.

The importance of establishing and respecting a shared understanding of context is shown by NHS England's care.data programme. Here public opinion turned against a proposal to re-use individual patients' records for wider research, resulting in the programme being suspended.[52] One witness to the House of Commons' Science and Technology Committee suggested this was because 'care.data had been framed as helping the NHS as an institution rather than helping individual patients'.[53] The Committee later contrasted this with 'the success of a scheme similar to care.data in Scotland',[54] concluding that 'patients and healthcare professionals are not against the sharing of patient records"'[55] but 'support for data usage is highly dependent upon the context within which the data is collected.'[56]

---

[46] Information Commissioner's Office, 'WiFi Location Analytics' (16/02/16) <https://ico.org.uk/media/for-organisations/documents/1560691/wi-fi-location-analytics-guidance.pdf> accessed 9 Aug 2016.

[47] Article 29 Data Protection Working Party, *Statement of the WP29 on the impact of the development of big data on the protection of individuals with regard to the processing of their personal data in the EU* (16 September 2014) (14/EN WP 221) 2.

[48] Helen Nissenbaum, *Privacy in Context* (Stanford Law Books 2010).

[49] *Ibid*, 140.

[50] *Ibid*, 146.

[51] *Ibid*, 164.

[52] Nick Triggle, 'Care.data: How did it go so wrong?' (*BBC News*, 19 Feb 2014) <http://www.bbc.co.uk/news/health-26259101> accessed 9 Aug 2016.

[53] Professor Liesbet van Zoonen, Loughborough University, to House of Commons Science and Technology Committee, *Responsible Use of Data* (Fourth Report of Session 2014-15 HC 245) 15.

[54] House of Commons Science and Technology Committee, *The Big Data Dilemma* (Fourth Report of Session 2015-16 HC 468) 25.

[55] *Ibid*, 25.

[56] *Ibid*, 23.

Where big data techniques are used within a clear, agreed context, the contextual integrity framework can be used to assess whether or not they are likely to be perceived as acceptable, or at least legitimate, giving big data users some measure of the risk of reputational or operational damage resulting from their activities. However those wishing to make unrestricted use of big data techniques, without choosing a purpose in advance, have a much greater problem. Without any defined context they have no norms against which to assess the compatibility of their proposed information flows. Worse, Nissenbaum suggests that users of new technologies such as social media services are likely to select the existing context that most closely matches their own use of the service, and base their expectations on the norms of that context.[57] Where big data techniques are used on information gathered in a wide variety of different social contexts, the risk of users perceiving an integrity violation of one or more of those contexts seems extremely high.

The market in purposes suggested above may help to explore this new territory. If individuals can choose among various additional purposes when they sign up to a service, this may help the provider identify clusters of purposes that are perceived as compatible. Where a purpose appears acceptable (or unacceptable) to a minority, this might form the basis of a separate option, such as the personalised discount storecard discussed by the Article 29 Working Party. Such an approach should at least reduce the risk of a service suffering a catastrophic loss of confidence like that affecting care.data. Meyer-Schönberger and Cukier have compared big data techniques to a 'treasure hunt';[58] offering data subjects a choice of new purposes may help to identify the extent of the island within which digging is socially acceptable.

6. Conclusion

With a growing consensus that prior consent is an inadequate solution to the challenges of big data techniques, a solution drawing on a wider range of data protection tools is required. The Article 29 Working Party's concept of 'downstream consent' appears more fruitful, incorporating additional transparency and control through the rules of purpose limitation and legitimate interests. Splitting big data into distinct analysis/pattern-finding and intervention/pattern-matching stages allows appropriate controls, guidance and regulatory enforcement to be applied to each.

Big data raises social, as well as legal, concerns. The House of Commons' Science and Technology Committee warn:

> Given the scale and pace of data gathering and sharing, however, distrust and concerns about privacy and security is often well founded and must be resolved by industry and Government if the full value of big data is to be realised.[59]

With one of the richest and longest-established data protection frameworks, Europe has an opportunity to lead the development and adoption of responsible big data practices that can build public confidence and support. Individuals, businesses, governments and society all stand to benefit from such an achievement.

---

[57] Helen Nissenbaum, *Privacy in Context* (Stanford Law Books 2010) 223.

[58] Viktor Mayer-Schönberger and Kenneth Cukier, *Big Data: A Revolution That Will Transform How We Live, Work and Think* (John Murray 2013) 15.

[59] House of Commons Science and Technology Committee, *The Big Data Dilemma* (Fourth Report of Session 2015-16 HC 468) 4.