Systems Thinking, Big Data, and Data Protection Law

Using Ackoff's Interactive Planning to Respond to Emergent Policy Challenges

Henry Pearce[*]

Abstract

This article examines the emergence of big data and how it poses a number of significant novel challenges to the smooth operation of some of the European data protection framework's fundamental tenets. Building on previous research in the area, the article argues that recent proposals for reform in this area, as well as proposals based on conventional approaches to policy making and regulatory design more generally, will likely be ill-equipped to deal with some of big data's most severe emergent difficulties. Instead, it is argued that novel, and possibly unorthodox, approaches to regulation and policy design premised on systems thinking methodologies may represent attractive and alternative ways forward. As a means of testing this general hypothesis, the article considers Interactive Planning, a systems thinking methodology popularized by the organizational theorist Russel Ackoff, as a particular embryonic example of one such methodological approach, and, using the challenges posed by big data to the principle of purpose limitation as a case study, explores whether its usage may be beneficial in the development of data protection law and policy in the big data environment.

Keywords: big data, data protection, data minimization, systems thinking, interactive planning.

A Introduction

This article represents an attempt to build on a previously published article in which it was suggested that rather than relying on conventional policy-making logic to respond to some of most serious regulatory challenges linked to the emergence of big data, European regulators might better be served by pursuing novel and possibly unorthodox approaches to designing and implementing data protection law and policy in the emerging big data environment.[1] Here, it was argued that approaches to policy making premised on systems thinking methodologies

* University of Hertfordshire, Lecturer in law, e-mail: h.pearce@herts.ac.uk.

1 H. Pearce, 'A Systems Approach to Data Protection Law and Policy in a World of Big Data?', *Computer and Telecommunications Law Review*, Vol. 22, No. 4, 2016.

doi: 10.5553/EJLR/138723702016018004004

may represent a particularly promising avenue of enquiry. This article examines this notion in greater detail and, with a particular focus on the practice known as 'Interactive Planning', further makes the case for systems thinking, and its associated methodologies, to be incorporated into the shaping of data protection law and policy in the big data regulatory environment. Accordingly, though throughout the course of the article certain specific emergent regulatory problems linked to big data are considered, with one in particular being considered in some detail, it is not necessarily the article's primary intention to advance definitive solutions to any of the challenges mentioned. While some embryonic ideas for future regulatory frameworks are suggested, the primary objective of the article is to highlight the potential of Interactive Planning, as one prominent example of a systems thinking methodology, as a methodological means of improving the design and implementation of data protection law, policy, and regulation in the big data environment.

The article's structure is as follows. First, the emergence of the phenomenon of big data is considered, and it is explained how, despite its considerable potential for good, it raises a number of notable regulatory concerns, some of which give rise to serious, possibly fundamental, reservations about the continued suitability of some of the European data protection framework's core tenets. Second, the concept of systems thinking is introduced, and it is explained why, in the big data environment, the use of systems thinking methodologies may be suitable for deployment as a means of responding to emergent policy challenges. Following this, as a case in point, the systems thinking methodology of Interactive Planning, popularized by the American organizational theorist Russell Ackoff, is outlined as an example of one such methodology that may be of particular use in this context. The final substantive section of the article then, using the challenges faced by the principle of data minimization – central tenet of the European data protection framework, as a case study, deploys the Interactive Planning methodology outlined in the previous section as a means of demonstrating the potential of systems thinking as a tool for designing regulation in the big data environment. Some suggestions as to how the policy and regulatory challenges identified might be responded to are then advanced.

B Big Data

Generally speaking, big data can be considered a loosely defined term, which is broadly used to describe data sets that are so large and complex that they have become awkward to work with using standard statistical software, or data which are too large to be stored, managed, or analysed in a single organization.[2] Essentially, the existence of such data sets is made possible by the unprecedented, and exponentially increasing, amount of data produced and put into circulation in the

2 C. Snijders, U. Matzat & U.D. Reips, '"Big Data": Big Gaps of Knowledge in the Field of Internet Science', *International Journal of Internet Science*, Vol. 7, No. 1, 2012, p. 1.

world today.[3] As noted by boyd[4] and Crawford, however, big data is in many ways a poor term. In their words, there is little doubt that the quantities of data now available in the world are often huge, but that is not the defining feature of this new data ecosystem.[5] Big data, in fact, is less about data that are big, and more about increased capacities to search, aggregate, and cross-reference large data sets.[6] Working with data sets of such extraordinary size and scale allows for those in possession of sophisticated analytical tools to identify patterns, and make inferences and predictions that would not have previously been possible when working with data sets of a smaller size. Quite often this will entail data that have been collected for a specific purpose being repurposed to serve an entirely different end. The types of scenarios that would fall within these categories can range drastically, from data derived from vast international science projects, such as the Large Hadron Collider of the European Organization for Nuclear Research, to the wealth of data collated by online companies such as Facebook and Google.[7]

While big data's latent value is seemingly huge, and its uses are likely to lead to a significant number of benefits, both social and economic in nature, it has also been noted that certain big data analytics operations may also be capable of causing harmful and undesirable consequences for individuals, and thus they require regulation. In particular, automated-algorithmic profiling, a key constituent part of many big data analytics operations, has been identified as having the potential to lead to discriminatory practices and the diminution of individual autonomy.[8] Significantly, in this regard, it is important to note that many big data analytics operations will involve the processing of personal data of individuals, and so will

3 For instance, *The Economist* reported in its 2013 Outlook that the quantity of global digital data expanded from 130 exabytes in 2005, to 1,227 in 2010, and will have increased to 7,910 by 2015 – to highlight the enormous quantity of this amount of data in lay terms, the 1,227 exabytes of data, if stored on DVDs, would require a fleet of 16 million Boeing 747 aircraft in order to transport it globally. The Economist, 'Welcome to the Yotta World', available at: <www.economist.com/node/2153792> (last accessed September 2016).

4 In accordance with her wishes, this paper refers to danah boyd using lower case letters only.

5 Some of the data encompassed by big data, for instance trending Twitter posts, will not be as large as earlier data sets that we have not considered big data, such as national censuses.

6 D. Boyd & K. Crawford, 'Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon', *Information, Communication and Society*, Vol. 15, No. 5, 2012, pp. 662-679.

7 J. Fishleigh, 'A Non-Technical Journey into the World of Big Data: An Introduction', *Legal Information Management*, Vol. 14, No. 2, 2014, pp. 149-151.

8 *See* M. Hildebrandt, 'Profiling and the Rule of Law', *Identity in the Information Society*, Vol. 1, No. 1, 2008, pp. 55-70; L. Magnani, 'Abducing Personal Data, Destroying Privacy: Diagnosing Profiles Through Artificial Mediators', *in* M. Hildebrandt & J. de Vries (Eds.), *Privacy, Due Process and the Computational Turn: The Philosophy of Law Meets the Philosophy of Technology*, Oxon, Routledge, 2013, pp. 63-86.

European Journal of Law Reform 2016 (18) 4
doi: 10.5553/EJLR/138723702016018004004

fall under the jurisdiction of the European data protection framework.[9] Troublingly, however, the emergence of big data has presented significant, perhaps insurmountable, challenges for the smooth operation of this framework. In particular, there are concerns that, as traditionally envisaged, the concepts of personal data, informed consent, and, as will be considered in much greater detail below, the principle of data minimization, all central tenets of the data protection framework, are no longer fit for purpose. Accordingly, European regulators and policy makers are now being put to task to find fresh policy options, which will allow for a number of fundamental tenets and key underlying principles of data protection law to be modernized and brought into alignment with the practical realities of contemporary data-handling practices.

Given the apparent novelty of big data's challenges, it has been suggested that big data's emergence represented a perfect opportunity for a fundamental rethink of contemporary data protection law and policy, with some observers proposing that the time was ripe to attempt to devise novel approaches to regulation premised on 'brave new thinking.'[10] However, the response of European lawmakers and other regulatory bodies has been more measured. This is particular well-evidenced by the General Data Protection Regulation, which was supposedly drafted with the intention of bringing the European data protection framework into alignment with the practical realities of the twenty-first century.[11] In particular, in an apparent attempt to tackle the challenges posed by big data's emergence noted above, the Regulation contains revised definitions and conceptions of some of the data protection framework's fundamental pillars, such as the con-

9 At the time of writing, the European Union's main legislative instrument in the data protection framework is Directive 95/46/EC, commonly known as the Data Protection Directive, which, in Art. 2 (a) defines personal data as: "…any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or one or more factors specific to his physical, physiological, mental, economic, cultural or social identity". The General Data Protection Regulation, which will replace the Directive and apply in all EU Member States from May 2018, provides an updated version of the Directive's earlier definition: "…any information relating to an identified or identifiable natural person 'data subject'; an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, online identifier or to one or more factors specific to the physiological, genetic, mental, economic, cultural or social identity of that person". An agreed text of the General Data Protection Regulation was adopted by the European Parliament in April 2016, bringing four years of negotiations on the overhaul of European data protection rules to a close. *See* European Parliament, 'Data Protection Reform – Parliament Approves New Rules Fit for the Digital Era', 2016, available at: <www.europarl.europa.eu/news/en/news-room/20160407IPR21776/Data-protection-reform-Parliament-approves-new-rules-fit-for-the-digital-era> (last accessed September 2016).

10 P. Lee, 'A Brave New World Demands Brave New Thinking', *IAPP*, 2013, available at: <https://iapp.org/news/a/a-brave-new-world-demands-brave-new-thinking> (last accessed September 2016).

11 An agreed text of the General Data Protection Regulation was adopted by the European Parliament in April 2016, bringing four years of negotiations on the overhaul of European data protection rules to a close. *See* European Parliament, 2016, *supra* note 9.

cept of personal data itself,[12] informed consent[13] and the principle of data minimization.[14] There is reason to suspect that for one reason or another this legislative redrafting may be unlikely to achieve its desired objectives, with a variety of concerns now being voiced about their rationale and practical workability. An in-depth consideration of these issues is beyond the scope of this article. What is extremely significant in respect of the central objective of this article, however, is the way in which the Regulation's reformulation of many of the key tenets of the data protection framework appears to be an initiative born out of what might be termed conventional policy-making logic. It has been suggested, for instance, that they appear to be approaches to reform that have been designed as a means of addressing symptoms of the problems posed by big data rather than the underlying causes.[15] As a result, in even a best case scenario, they may fail to make any successful long-term impact, even if some short-term accomplishments are achieved. This is evidently a far cry from the brave new thinking that was clamoured for at the point of big data's initial emergence. What is also significant is that it is increasingly possible to argue that not only will the proposals for data protection reform contained within the General Data Protection Regulation likely fail to provide adequate answers to big data's most serious emergent challenges, but that conventional or linear approaches to regulation and policy of any form are not well-suited to problems of a chronically intricate nature in any circumstances, including those in the field of data protection.[16] In simple terms, in an age of mystifyingly complex data-driven business models and globalization, the ability of European lawmakers to bring about desired data protection policy objectives, and to bring the subject matter of data protection law, the use of personal data itself, under regulatory control through the use of traditional 'command and control', rights-based, rule-heavy, legalistic approaches to regulation, has perhaps reached its limit. One key general criticism of contemporary policy development, for instance, and one that appears to be particularly pertinent in the big data context, is that public policy responses in these forms are habitually too linear for the complexity of the issues for which they are devised. Many contemporary problems faced by policy makers might in fact be described as so-called 'wicked' problems. These are problems that are highly resistant to solutions, and challenge traditional governance structures and the capacity of traditional approaches to policy making and regulation.[17] Big data appears to be a problem with these characteristics. The obvious implication of this assertion being, that if European lawmakers wish to truly engage and grapple with the full extent of big data's emergent regulatory challenges, and develop responses to regulatory issues in the 'brave new thinking' mould, they must adopt new, non-conventional, methodo-

12 General Data Protection Regulation, Art. 4(1).
13 *Ibid*., Art. 4(11).
14 *Ibid*., Art. 5(c).
15 *Ibid*.
16 *Ibid*.
17 B. Head & J. Alford, 'Wicked Problems: Implications for Public Policy and Management', *Administration & Society*, Vol. 47, No. 6, 2015, pp. 711-739; J. Canty-Waldron, 'Using Systems Thinking to Create More Impactful Social Policy', *Journal of Future Studies*, Vol. 19, No. 2, 2014, pp. 61-62.

European Journal of Law Reform 2016 (18) 4
doi: 10.5553/EJLR/138723702016018004004

logical approaches to policy making. One particular school of methodological approaches that may be of use, but one that has hitherto been unexplored in the literature on big data and data protection, is that of systems thinking.

C Systems Thinking

In simple terms, systems thinking is a methodological approach to seeing, and talking about, reality that helps us better understand and work with systems around us in our everyday environment, to influence and improve the quality of our lives. Systems themselves can broadly be defined as sets of interacting parts – people, cells, or molecules, to give a few examples – interconnected in such a way that they produce their own behaviour over time. A system may be buffeted, constricted, triggered, or driven by outside forces, but the system's response to these sources is characteristic of itself.[18] Systems thinking approaches to problem solving are intended to be of benefit in situations where we are confronted by complex problems that involve a variety of different actors and have no obvious solutions. By taking a systems approach to problems of this character, we allow ourselves to see the whole picture of the issue at hand, and raise our thinking to the level at which solutions to multifaceted issues are forthcoming. This is something that competing conceptual frameworks do not always provide.[19]

However, despite what can be considered true systems thinking now being more than half a century old,[20] it appears that little has filtered down to everyday thinking or indeed policy design and implementation. This is perhaps unfortunate, as research in various fields has suggested that systems methodologies may have untapped potential in respect of deciphering the complexity inherent in many prominent contemporary regulatory challenges worldwide.[21] The obvious question to be asked in the immediate context, therefore, is whether the big data policy area is one such area where they may be suitable for deployment. As has been argued elsewhere, for a number of compelling reasons it appears, however, that this is a question that can be answered in the affirmative.[22] In order to examine this notion in greater detail, the next logical step is to identify precisely which system thinking methodology, or methodologies, may be best suited for deployment in this context. While methodologies like Strategic Assumption Surface Testing (SAST)[23] and Soft Systems Methodology (SSM)[24] were both considered as embryonic possibilities for analysis, it is, given its forward-looking nature, the methodology of Interactive Planning that was chosen as the focus of this article.

18 D. Meadows, *Thinking in Systems: A Primer*, London, Earthscan, 2009, p. 2.
19 J. Cordoba-Pachon, *Systems Practice in the Information Society*, Abingdon, Routledge, 2010, p. 30.
20 *See*, for instance L. Von Bertalanffy, 'The History and Status of General Systems Theory', *The Academy of Management Journal*, Vol. 15, No. 4, 1972, pp. 407-426.
21 J. Stewart & R. Ayres, 'Systems Theory and Policy Practice: An Exploration', *Policy Sciences*, Vol. 34, No. 1, 2001, pp. 79-94.
22 For a detailed exposition of these reasons, *see* Pearce, 2016.
23 *See* R. Flood & M. Jackson, *Creative Problem Solving*, Chichester, Wiley, 1991, p. 119.
24 *Ibid*.

D Interactive Planning

Interactive planning, developed and popularized by the American organizational theorist Russell Ackoff, is a systems thinking methodology that differs significantly from two more commonly used types of planning: reactive and pre-active, the use of which is far more common.[25] Reactive Planning tends to take the form of tactically oriented, bottom-up planning, the main objective of which is to identify existing deficiencies, and devising initiatives to eliminate or diminish them on a one-by-one basis. However, in respect of problems that are of a deeply complex nature, in the majority of situations it will be ineffective for three notable reasons. First, reactive planning is dedicated to the removal of deficiencies. However, when dealing with problems of a complex and nonlinear nature, the removal or elimination of one deficiency will not automatically result in an overall improvement in situation. In fact, in such scenarios it is not uncommon for the removal of individual deficiencies to give rise to unintended and undesirable consequences, leading to an overall worsening of the status quo. Second, given that Reactive Planning tends to focus on the removal of what one does not want rather than preserving or developing what one does want. One who focuses on the past while attempting to move into the future would have no control over where one is going.[26] Third, and perhaps most significantly, the way in which reactive planning tends to approach complex issues and challenges within a system in an isolated manner is too simplistic, as the overall state and behaviour of the system will often be more than the sum of its constituent parts and problems.[27] Pre-active planning initiatives, on the other hand, tend to consist of strategically oriented, top-down planning initiatives that are made up of two primary components: prediction and preparation. Initiatives of this sort are generally based on the assumption that, although the future is inherently unpredictable, with good forecasting the system's future can be controlled. Pre-active planning initiatives, therefore, concern themselves with planning *for* the future, not planning *the* future itself. The limitation in this methodological approach is that the future that manifests in reality is often drastically different from the one that is planned for, meaning that in practice very few pre-active plans are carried out to completion.[28] Significantly, policy-making initiatives premised on conventional logic tend to embody these types of approaches.[29]

Conversely, Interactive Planning is directed at creating the future, rather than predicting it or responding to it. It is premised on the underlying belief that

25 R. Ackoff, 'A Brief Guide to Interactive Planning and Idealized Design', 2001, available at: <www.ida.liu.se/~steho87/und/htdd01/AckoffGuidetoIdealizedRedesign.pdf>; R. Ackoff, 'Systems, Messes and Interactive Planning', *in* E. Trist, F. Emery & H. Murray (Eds.), *The Social Engagement of Social Science, Volume 3: A Tavistock Anthology – The Socio-Ecological Perspective*, Philadelphia, University of Pennsylvania, 1997, pp. 417-438; M. Jackson & P. Keys, 'Towards a System of Systems Methodologies', *Journal of the Operational Research Society*, Vol. 35, No. 6, 1984, p. 480.

26 *Ibid*.

27 *See* Ackoff, 2001; Jackson & Keys, 1984, p. 480.

28 *Ibid*.

29 Pearce, 2016.

European Journal of Law Reform 2016 (18) 4
doi: 10.5553/EJLR/138723702016018004004

a system's future depends at least as much on what it does between the present and the future as on what is done to it. It is a methodological approach to planning, therefore, that consists of the design of a desirable present and the selection or invention of ways of approximating it as closely as possible. It creates the future by continuously attempting to close the gap between where it is at the present, and its desired state. In essence, therefore, the main benefit that may be attributed to the use of Interactive Planning is that it can actively facilitate the exploration of better and more desirable futures and can help to develop future states of being that are actively sought-after, as opposed to growing one that is merely adequate, or even one that is suboptimal.[30] As such, if European regulators were to make use of Interactive Planning, they would theoretically be well-placed to devise regulatory interventions that *dissolve* emergent problems, rather than *resolve* them and, as a result, it should be possible for them to work towards a system of data protection law and policy that is acceptable to all relevant stakeholders.

Interactive Planning consists of two primary components: the idealization stage and the realization stage, each with their own individual sub-components. In the idealization stage, the planner is tasked with designing an idealized future for the system that is being planned for from scratch. All constraints other than technological feasibility, such as financial or political restrictions, are discounted. The prospect of future technological innovations can be taken into account, but these must be restricted to what is reasonably believed to be possible. The planner's design, therefore, is an explicit formulation of their conception of the system they would create if they were to create or design it exactly as they desired.[31] Once an idealized vision has been established and agreed upon by consensus, the next step is to begin planning how that vision can be achieved: Interactive Planning's realization stage. Both the idealization and realization phases are divisible into six interrelated phases, namely: formulating the mess, ends planning, means planning, resource planning, design of implementation, and design of controls. These six phases of interactive planning may not necessarily be carried out in this sequence, but they are usually initiated in this order. Because they are strongly interdependent, they usually take place simultaneously and interactively.

I Formulating the Mess

Every system is faced with a set of interacting threats and opportunities. This is known as a mess. The aim of this stage of planning is to determine how the system would eventually be destroyed, or at least severely damaged, if it were to continue behaving as it is currently. Identifying the presence of the system's Achilles heels provides a focus for the planning that will later be followed by identifying precisely what undesirable futures must be guarded against and avoided if at all possible. Accordingly, the formulation of the mess stage of the Interactive Planning process will require the planner to prepare the following: a detailed consideration of how the system in question operates at present; an analysis of the char-

30 Cordoba-Pachon, 2010, p. 30.
31 Ackoff, 1997, pp. 417-438.

acteristics and properties of the system that obstruct its progress; a projection of the system's future assuming that no changes are made to existing plans for its future development and the future environment that can be expected; and a description of how and why the system would likely be destroyed or severely damaged if the above-mentioned projections manifested in practice.[32]

II Ends Planning

Having formulated the mess, Interactive Planning's ends planning stage involves the development of the idealized design of the system's future, as mentioned earlier. The planner is asked to determine what the system in question would be like in the present, if it could take whatever form was desired. In other words, the ends planning stage involves identifying the gaps between the idealized design and the state of the system as envisaged during the formulation of the mess. Once this has been done, the remainder of the planning process can be dedicated to an analysis of how these gaps can be removed or reduced, both collectively and interactively.[33]

III Means Planning

The means planning stage involves making determinations in respect of what ought to be done to remove or reduce the above-mentioned gaps or barriers, which exist between the 'mess' that has been formulated and the desired goals identified during the ends planning stage. It is here where the planner should select or invent the courses of action, practices, projects, programmes and policies that are to be implemented in the pursuit of the system's idealized redesign.[34]

IV Resource Planning

The resource planning stage requires the planner to determine what resources will be required in order for the means to reach the desired ends to be executed. In particular, it will be important at this stage to determine how much of the necessary resources – be they facilities and equipment, materials, energy, services, personnel, money, or expertise – will be needed in order to implement the means selected.[35]

V Design Implementation and Control

With all of the above steps addressed, it will then be important to decide upon how any agreed courses of action are to be implemented, and how they are to be monitored so to avoid them going awry and straying from the desired objectives. In particular it will be important for planners to determine precisely who is required to do what, where, and when, and devise methods of monitoring implemented planning decisions to determine whether they are producing the desired

32 Ackoff, 2001; Cordoba-Pachon, 2010, p. 30; Jackson & Keys, 1984, p. 480.
33 *Ibid*.
34 *Ibid*.
35 *Ibid*.

European Journal of Law Reform 2016 (18) 4
doi: 10.5553/EJLR/138723702016018004004

results and, if not, what corrective action ought to be taken.[36] Once completed, the idealized design and all of the conclusions drawn from the other stages of the process should be distributed for comment, criticism, and suggestions to as many relevant stakeholders who have not been involved in its preparation as possible. Where possible their inputs should then be incorporated into the design. When this is not done, an explanation should be provided to those who offered relevant comment, criticism, or suggestions.

E Applying Interactive Planning to Big Data's Emergent Policy Challenges

Having outlined why on paper systems methodologies, and particularly Ackoff's Interactive Planning methodology, may be well suited to addressing pressing problems arising in the big data and data protection policy area, in terms of both designing and implementing effective regulation, the next step is to test this general hypothesis. This section of the article intends to demonstrate the precise potential of Interactive Planning as a tool for policy formulation and regulatory design, using difficulties surrounding the principle of minimization caused by big data's emergence as a specific case study.

I Formulating the Mess

The principle of data minimization can be considered one of the European data protection framework's fundamental doctrines. The principle requires that the only personal data that ought to be collected, stored, and processed are those which are necessary to realize specified and legitimate goals, and that all such data held by a data controller should be destroyed once they are no longer relevant to the achievement of such goals. This is embodied by Article 6(c) of the Data Protection Directive, which specifies that data must not be 'excessive' to the purposes for which they are collected. The General Data Protection Regulation, which, as noted above, has been drafted specifically to bring the main tenets of the data protection framework into alignment with contemporary data-handling practices, retains the principle of data minimization as one of its substantive pillars. Notably, Article 5(c) of the Regulation explicitly states that any collections of personal data must be limited to 'what is necessary' required in order for a legitimate processing activity to be carried out. The upshot of these provisions is clear: data controllers should refrain from arbitrarily collecting and storing personal data carte blanche, and make collections only when strictly necessary for the completion of legitimate processing activities.

The origins and rationale behind the principle can be traced back to the 1970s, where concerns began to arise over the collection and processing of data relating to individuals in centralized government computers and databases. It was intended to prevent powerful organizations from building giant dossiers of information relating to individual persons that could then be used for nefarious purposes such as manipulation, profiling, and discrimination. The placing of limits

36 *Ibid.*

on the amount of data that can be gathered and stored was thought to offer safeguards against such potential harms. Data cannot, for instance, be lost, stolen, or misused, if they have not been collected in the first place. At the time the concept was formulated, computing technologies that had the software and processing power to handle large amounts of data were in their infancy. Similarly, there were no ways for individuals' personal data to be collected and distributed via an international super network like the Internet.[37] However, times have changed, and once again, as with a number of other prominent aspects of the European data protection framework, the emergence of big data has called into question the principle's practical utility and overall value.

In general terms there are now considerable doubts in respect of data minimization principle's practical utility, and whether it ought to be retained as a key regulatory principle, simply because it increasingly appears that it simply cannot be at all reconciled with the practical realities of data gathering and handling practices in the emerging big data environment. Simply put, given the fact that many big data analytics operations, and the benefits that are likely to stem therefrom, necessarily require the amassment of huge quantities of data, big data and the principle of data minimization simply cannot be aligned with one another. In any event, as noted above, the amount of data generated, gathered, compiled, and processed in the world in a year is growing at a phenomenal rate, and shows no sign of abating.[38] At the very least, therefore, there appears to be a considerable discrepancy between what is said on the statute book and what is happening in practice, with seemingly no prospect of the situation being reversed. Accordingly, it seems safe to conclude, that whichever way it is looked at, the principle simply does not work in practice in the big data environment. The 'mess' we are able to formulate here, therefore, is the fact that the operation of a fundamental tenet of the European data protection framework is manifestly sub-optimal, and possibly entirely unfit for purpose. This is severely problematic as systems of law and regulation that do not work as they should, or those which are not enforced in any meaningful way, will run the risk of appearing illegitimate and, in effect, not worth the paper they are written on. As has been widely noted elsewhere, a perceived lack of legitimacy is one of the primary reasons for which, historically, numerous high-profile and wide-ranging regulatory regimes have collapsed into failure.[39] This is clearly, therefore, a policy challenge that is in urgent need of response.

II Ends Planning

The preceding section outlined how the emergence of big data has raised significant questions as to the practical utility and worth of the principle of data mini-

37 L. Colana, 'Mo' Data, Mo' Problems? Personal Data Mining and the Challenge to the Data Minimization Principle', *The Future of Privacy Forum and Stanford Law School Centre for Internet & Society*, 2013, available at: <https://fpf.org/wp-content/uploads/Colonna-Mo-Data-Mo-Problems.pdf>.

38 *See supra* note 3.

39 R. Brownsword & M. Goodwin, *Law and the Technologies of the Twenty-First Century*, Cambridge, Cambridge University Press, 2012, p. 61.

European Journal of Law Reform 2016 (18) 4
doi: 10.5553/EJLR/138723702016018004004

mization, a central tenet of the European data protection framework. Having identified that this is the 'mess' that is in need of address, the next step is to attempt to design a desirable end or ends. In other words, before attempting to address *how* problems facing the data minimization principle might be addressed, we must ask the question, what are the end goals we must now strive towards, and what sort of big data regulatory environment do we want to have moving forward? In order to answer this question, it is first useful to take a brief trip down memory lane, and consider the emergence of data protection law as a distinct field of legal practice.

Regulatory concerns pertaining to the use of information technologies and, in particular, the way in which they allow information relating to individuals to be handled and used, have materialized repeatedly, almost systematically, throughout history. The emergence of the first handheld cameras in the late nineteenth century, for instance, made it possible to capture peoples' images, and spawned concerns about media intrusions into their private lives.[40] Half a century later, leaps forward in computing technologies generated further worries about information technology's inherent harmful potential. The trend of increased apprehensiveness linked to the computerized processing of information experienced greater acceleration in the 1970s as mainframe computers became pervasively used by organizations in advanced economies.[41] The emergence of the Internet, and later the World Wide Web in the 1990s, and their inherent possibility for the monitoring and analysis of data relating to individuals further fanned the flames that had already been lit under the above-mentioned concerns, and was one of the contributing factors that led to the enactment of Europe-wide data protection laws designed, in part at least, to guard against the negative consequences that could potentially stem from the use of these technologies.[42]

What is noteworthy for the purposes of our present discussion, however, is the fact that not only were these issues significant in years gone by, but, due to the emergence of big data, they are now more salient and in need of discussion than ever before. This appears to be true for a number of reasons. As noted above, big data analytics operations often involve using the personal data of individuals to build sophisticated statistical models, which can then be exploited to derive information and construct profiles relating to the individuals to whom those data relate. While this may lead to benefits for said individuals, such analyses may cause significant aspects of their life and interests to be revealed, or inferred, possibly incorrectly, and the results may be used to make important decisions about

40 S. Wicker, *Cellular Convergence and the Death of Privacy*, Oxford, Oxford University Press, 2013, p. 58.

41 I. Brown & C. Marsden, *Regulating Code: Good Governance and Better Regulation in the Information Age*, London, MIT Press, 2013, p. 47.

42 The first example of this general trend was the adoption of the Convention on Data Protection (formerly known as the Convention for the Protection of Individuals with regard to the Automatic Processing of Personal Data, January 1981, CETS no. 108), which was later followed by Directive 95/46/EC (the Data Protection Directive) and, more recently Regulation (EU) 2016/679 (the General Data Protection Regulation).

them without their knowledge.[43] This in turn, could have the potential to lead to individuals being inadvertently or deliberately excluded from making use of certain services or from pursuing certain possibilities.[44] Not only do many big data analytics operations necessarily generalize, make broad predictive assumptions, and stereotype individuals based on their choices and circumstances, but they facilitate and encourage inequality by creating an environment where individuals are treated differently based on their profiles.[45] The ostracism and stigmatization this could allow, in respect of individuals belonging to minority groups in particular, could, therefore, increase and generate an explicit loss of freedom.[46] In addition to the possibilities for direct discrimination, targeted advertising and marketing endeavours that are often the result of big data analytics operations may also give rise to other indirect harms, namely the subtle influencing of individual behaviours. This is perhaps best illustrated by the now famous example of the US Target Superstore, which analysed its customers' purchasing behaviour to, among other things, identify customers in the early stages of pregnancy, and adjust their advertising and marketing endeavours in light of such discoveries.[47] This example precisely demonstrates how analytics and profiling-based advertising and marketing operations can be used to target specific individuals at critical or highly sensitive points in their lives, when their behaviour is in flux and new habits are formed.[48] Such is the invasiveness and potential impact of these activities that it has been suggested that they may have the potential to undermine individuals in respect of their shopping and purchasing decisions, which could have serious consequential effects in terms of important life decisions more generally.[49]

By looking at some of the major developments in information technology that have occurred over the course of the last century, as well as taking a specific look at some of big data's latent potential harms and concerns, we can see that personal data, and the technologies that utilize such data, have throughout history repeatedly been a prime concern to regulators and policy makers at both national and supranational levels. Crucially, we can see that the concerns sur-

43 R. Cumbley & P. Church, 'Is "Big Data" Creepy?', *Computer Law & Security Review*, Vol. 29, No. 5, 2013, pp. 601-609.

44 Certain types of big data analytics operations may in some circumstances, for instance, be used as a means of assessing an individual's credit worthiness. *See*, for instance J. Deville, 'Leaky Data: How Wonga Makes Lending Decisions', *Charisma: Consumer Market Studies*, 2013, available at: <www.charisma-network.net/finance/leaky-data-how-wonga-makes-lending-decisions>.

45 *See* Hildebrandt, 2008, pp. 55-70.

46 Magnani, 2013, pp. 63-86; International Working Group on Data Protection in Telecommunications, 'Working Paper on Big Data and Privacy: Privacy principles under pressure in the age of Big Data analytics', 2014; Executive Office of the President of the USA, 'Big Data: Seizing Opportunities, Preserving Values', 2014, available at: <https://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf>.

47 Forbes (last accessed September 2016) "How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did", <www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/#4c64265734c6>.

48 *See* Cumbley & Church, 2013, pp. 601-609.

49 C. Dwyer, 'Behavioural Targeting: A Case Study of Consumer Tracking on Levis.com', *in Proceedings of the Fifteenth Americas Conference on Information Systems*, San Francisco, California, 6-9 August 2009.

European Journal of Law Reform 2016 (18) 4
doi: 10.5553/EJLR/138723702016018004004

rounding the processing of personal data first voiced in mid- to late twentieth century are every bit as relevant in the present as they were in days gone by. Such are the extensive possibilities that are now inherent in the processing of personal data, the dangers of abuses, and the consequences of such abuses, are more extensive and potentially severe than at any other point in recorded history. The rationale for the principle of purpose limitation – the desire to guard against abuses stemming from the processing of personal data – therefore, remains defensible and highly salient. One desirable end that we can immediately identify from our deliberations thus far, therefore, is the necessity of constructing and maintaining a system of data protection regulation that affords individuals a high degree of protection against potential abuses that may stem from the processing of their personal data.

This, however, is only one side of the proverbial coin. As noted above, when attempting to deploy Interactive Planning, we must be reminded of the fact that one of the methodology's central facets is taking account of the viewpoints of a variety of relevant stakeholders. While the development of a system of law and regulation that ascribes individuals a high level of data protection and protects them from abuses stemming from the processing of their personal data will be desirable to some stakeholders, notably said individuals themselves, other stakeholders, notably private companies and other firms heavily involved in the analysis of big data, will inevitably have other, possibly competing, interests. In particular, large multinational firms whose business models are based on the gathering and in-depth analysis of data, such as Google, or research institutions whose research is reliant on the analysis of big data, are unlikely to be receptive to the introduction of high levels of individual data protection if any such measures have the impact of unduly and negatively impacting on their data analytics operations. Accordingly, the construction of a big data regulatory regime that excessively restricts economic uses of personal data is something that must be avoided if the views of all relevant stakeholders are to be taken into account.

When taking all of the above factors into consideration, it would appear that the 'end' that is most desirable in the context of designing data protection law and policy in a world of big data is the construction of a regulatory environment in which individuals are afforded meaningful safeguards and protection against abuses stemming from the processing of their personal data, but simultaneously, an environment that does not unduly restrict and impede legitimate specified uses of those data by others whether such uses are of a commercial, economic, or research-related nature. Having recognized this, we must now identify the barriers that are standing in the way of the achievement of this goal. This, however, appears to be a relatively straightforward task. Quite simply, in the big data environment, such is the enormity of the volume of information generated, gathered, and analysed, nobody knows when, in what form, and with whom, individuals share their personal data. This is unsatisfactory to all relevant stakeholders. To the individual, this uncertainty is reflective of a growing inability to control and exercise autonomy over their personal data, as well as being reflective of the inability of regulators to provide them with meaningful data protection rights, which they can invoke against other parties who may seek to use their personal

data in nefarious ways. If data protection law is to provide individuals with relief against potential abuses, for instance, it is surely a prerequisite that the said individuals must be able to identify those against whom they can pursue a remedy. To private firms and other organizations engaged in big data analytics operations, on the other hand, this uncertainty has fostered a culture of mistrust between them and the individuals whose personal data they seek. This, in turn, has prevented them from reaping the full benefits of their data gathering and processing activities.[50] It would seem, therefore, that the 'barrier' that must be removed if the interests of both individuals and other data-using parties are to be given effect to, is the current uncertainty and dearth of knowledge regarding data flows and information streams that are pervasive in the big data environment caused by the huge volume of data generated, gathered, and collected.

III Means Planning

As noted above, the most significant barrier that stands in the way of the achievement of the desirable 'end' of constructing a big data regulatory framework in which individuals enjoy a high level of data protection appears to be the fact that, due to the enormity of all the data collections and processing activities that occur in the big data environment, nobody knows precisely who shares which data with whom, and when and under what circumstances sharing occurs. The data minimization principle, despite the apparent flaws in its practical application, was designed precisely to prevent the manifestation of this sort of situation. One obvious way in which the above-mentioned problems associated with the data minimization principle could potentially be addressed, therefore, would be to try and resurrect the principle, and breathe new life into it, so that it would be able to function more smoothly in conjunction with contemporary data-handling practices. A revised version of the principle that was worded more tightly, for instance, might represent one possible way in which this could be attempted. It is difficult to see, however, how such endeavours could ever be successful. Surely no amount of clever reformulation of the principle could ever reconcile the very notion of restricting or reducing the amount or volume of data that are gathered with the enormous quantities of personal data that are currently collected in the world today, a trend that has effectively become established as a norm both online and in other key areas of big data's application. Alternatively, regulators could seek to breathe new life into the principle by way of introducing improved standards in an attempt to encourage minimal data collections, or more robustly enforce the principle by way of introducing stricter penal sanctions to be levied against par-

50 It has been shown, for instance, that the development internet economy of the G20 countries, which is been estimated to be worth more than US$4.2 trillion alone by 2016, has been hindered significantly by a widespread unwillingness of individuals to share their personal data. World Economic Forum, 'Rethinking Personal Data: Strengthening Trust', 2012, available at: <www3.weforum.org/docs/WEF_IT_RethinkingPersonalData_Report_2012.pdf>. *See also* J. Cohen, 'What Privacy Is For', *Harvard Law Review 1904*, Vol. 126, 2013, pp. 1918-1927; M. Taddicken, 'Privacy, Surveillance, and Self-Disclosure in the Social Web: Exploring the User's Perspective in Focus Groups', *in* C. Fuchs *et al.* (Eds.), *Internet and Surveillance: The Challenges of Web 2.0 and Social Media*, New York, Routledge, 2011, p. 266.

European Journal of Law Reform 2016 (18) 4
doi: 10.5553/EJLR/138723702016018004004

ties whose collections are deemed to be excessive. Again, however, this would likely end in failure. Against the background of such proposals we must be reminded of the fact that enormous data gatherings in the big data environment have essentially become established as a cultural and societal norm. Any regulatory initiative that has the effect of clashing with that norm will surely be met with resistance, not only from parties such as big data analytics firms, whose day-to-day activities rely on the acquisition and analysis of data, but possibly also from individuals.[51] What is significant in this regard is, as noted above, regulatory endeavours, which clash with established cultural and societal norms due to difficulties in ensuring compliance and enforcement, frequently end in failure.[52] For these reasons, so far as using regulatory endeavours to protect individuals from abuses by actually limiting personal data collections is concerned, quite frankly, the proverbial wheels may have been turned too far for them to be turned back.

However, while the goal of limiting data collections may be unattainable, as noted above, the rationale behind the purpose of limitation principle remains desirable, and there is likely to be more than one way of skinning the proverbial cat. Simply put, there are a number of prospective ways in which the data minimization principle's rationale of guarding against abuses stemming from large-scale data collections can be put into action. One promising and novel way in which this might be done would be to shift the existing regulatory framework to a *sui generis* model of data protection, which targeted certain uses of personal data rather than their collection. For instance, such a regime might prohibit certain data processing activities in certain circumstances (e.g., if the processing of certain types of personal data in a certain context, such as consumer profiling, was likely to lead to discriminatory or otherwise harmful outcomes), or mandate that certain processing activities may only be permitted if adequate legal or technical safeguards were put in place to ensure any injurious consequences could be avoided. In the event that their personal data were used in a way that fell outside the scope of what the law permitted, individuals would be able to identify when and where their personal data were used in ways that were undesirable or unlawful, and seek immediate legal redress. Regulators, therefore, could step up their monitoring of personal data usage without having to worry about the collection of those data, meaning that in effect, the principle of data minimization could effectively be retired. Accordingly, a shift to a *sui generis* model of data protection could theoretically allow for the problems associated with the data minimization principle's shortcomings to be dissolved rather than resolved and thus, accord-

51 Research has shown that, to many individuals, personal data has become a bargaining chip or tradeable commodity that must be parted with in exchange for an ever-expanding range of amenities and services, both online and offline. To such individuals, who wish to make the most of such services with minimal hassle, a more robust and rigorously enforced data minimization principle could conceivably be an unwelcome development. *See* D. Trottier & D. Lyon, 'Key Features of Social Media Surveillance', *in* C. Fuchs *et al.* (Eds.), *Internet and Surveillance: The Challenges of Web 2.0 and Social Media*, New York, Routledge, 2011, p. 96.

52 C. Reed, *Making Laws for Cyberspace*, Oxford, Oxford University Press, 2012, pp. 20-22. *See also* Brownsword & Goodwin, 2012, p. 61.

ingly, there is an emerging consensus among various commentators that such a move would appear to have considerable promise.[53]

A shift to a model of data protection of this sort, however, would not be straightforward. The most obvious challenge is that a model of data protection premised on regulating data uses rather than one reliant on fair information principles as is the case with the existing model, is that it would still necessarily require individuals to have a knowledge and understanding of other parties holding their personal data, and how and why those data were being processed. If, for instance, individuals are not able to discern when, where, and by whom their personal data is being processed, they would never meaningfully be able to invoke new explicit rules regarding data processing activities and seek a remedy in the event their personal data were being misused. Accordingly, the question that now arises is, if restricting the volume of data collections that occur in the big data environment is not a feasible way in which the understanding of individuals can be enhanced in respect of which parties hold their personal data, as well as how they use those data, what other avenues can be identified as a means of achieving the same objective?

Another occasionally mooted suggestion is the introduction of increased transparency obligations, backed by law, that would be imposed on parties that make use of personal data.[54] Legally requiring such parties to present their processing activities to persons whose personal data were involved in commonly understood language, for instance, ought to contribute to the development of a more robust understanding of how personal data are used in the big data environ-

53 Numerous observers have, for instance, thrown their weight behind the notion that a shift to a *sui generis* model of data protection in the manner outlined above may help address some of the most severe problems currently facing the data protection framework. *See*, for instance B. Koops, 'The Trouble with European Data Protection Law', *International Data Privacy Law*, 2014; Colana, 2013.

54 The General Data Protection Regulation for instance, places a number of new transparency obligations on data controllers. First, recital 58 of the Regulation states that the principle of transparency requires that information addressed to the public or individuals should be both easily accessible and easy to understand, and that such information should be articulated in clear and plain language. So to give effect to this ideal, Art. 12(1) then makes it clear that data controllers must provide transparent and easily accessible policies pertaining to the processing activities which they intend to subject their users' personal data, and to the exercise of their user's legal rights in respect of that processing. Art. 14(1) provides that where data relating to an individual are collected from the individual, the controller will provide that individual with, at least, the following: the identity and contact details of the controller, the purposes of the processing for which the personal data are collected, the period for which the personal data will be stored, the recipients or categories of recipients of the personal data, and other information regarding data transfers to third countries and the contact details of relevant supervisory authorities. Art. 14(1)(a) then provides the same for data that have not been obtained from the individual. *See also* B. Koops, 'On Decision Transparency: How to Enhance Data Protection After the Computational Turn', *in* M. Hildebrandt & J. de Vries (Eds.), *Privacy, Due Process and the Computational Turn: The Philosophy of Law Meets the Philosophy of Technology*, Oxford, Routledge, 2013, p. 199; Y. Liu, 'User Control of Personal Information Concerning Mobile-App: Notice and Consent?', *Computer Law & Security Review*, Vol. 30, No. 5, 2014, pp. 521-529; Electronic Frontier Foundation, 'It's Time for Transparency Reports to Become the New Normal', 2013, available at: <https://www.eff.org/en-gb/deeplinks/2013/01/its-time-transparency-reports-become-new-normal>.

European Journal of Law Reform 2016 (18) 4
doi: 10.5553/EJLR/138723702016018004004

ment, allowing for individuals to more effectively invoke their data protection rights if the need arose. The immediate, and likely fundamental, problem that can be identified in relation to such initiatives, however, is the fact that there is reason to suspect that it may, in fact, be virtually impossible, given their inherent complexity, to explain the intricacies and likely consequences of many big data analytics operations in plain language notices that are understandable to ordinary people.[55] If this is the case, any attempt to rectify the above-mentioned uncertainty and lack of lucidity in the big data environment by way of heightened transparency rules and obligations will be a complete non-starter.

However, a more promising way forward might be found in the source of all the trouble, that is to say, technology itself. Though technologies themselves tend not to be habitually thought of as widely accepted ways by which desired regulatory objectives can be secured,[56] the last few years have seen the emergence of, and huge developments in, what have been termed Personal Information Management Services.[57] These can broadly be described as technological tools designed to help individuals collect, control, monitor, and use their personal data for their own purposes and make better decisions in respect of the use of those data by others. Given their purpose and rationale, it has been suggested that these types of services have the potential to have far-reaching consequences in respect of how individuals and other parties seeking their personal data interact with one another and, in particular, make significant inroads into reducing problems linked to the inability of the individual to keep track of their personal data in the big data environment. Though Personal Information Management Services come in a variety of forms, two types that are particularly salient in the immediate context are personal data stores and transparency-enhancing technologies.

Personal data stores are technological platforms that allow individuals themselves, as opposed to private organizations and other third parties, to store, manage, and deploy their data in a secure and structured way. They typically provide individuals with a means by which they can visualize their personal data, and functional mechanisms by which they can exercise determinations in respect of with whom particular aspects of those data are shared, at which times, how they are used, and for what purposes. These features and functions are encompassed by a definition offered by O'Hara and Van Kleek, who explain personal data stores accordingly:

55 M. Jensen, 'Challenges of Privacy Protection in Big Data Analytics', *2013 IEEE Congress on Big Data*, 2013, pp. 235-238; S. Barocas & H. Nissenbaum, 'Big Data's End Run around Anonymity and Consent', *in* J. Lane *et al*. (Eds.), *Privacy, Big Data and the Public Good*, New York, Cambridge University Press, 2014, p. 59.

56 There is, in fact, some debate as to whether technological tools should ever be used as a means of securing desired regulatory objectives. *See*, for instance K. Yeung, 'Towards an Understanding of Regulation by Design', *in* R. Brownsword and K. Yeung (Eds.), *Regulating Technologies: Legal Futures, Regulatory Frames and Technological Fixes*, Oxford, Hart, 2008.

57 A 2014 report by the market analyst Ctrl-Shift estimated the value of the PIMS market to be in excess of £16bn, with new forms of relevant services emerging every week. Ctrl-Shift, 'Personal Information Management Services: An analysis of an emerging market', 2014, p. 8, available at: <https://www.ctrl-shift.co.uk/research/product/90>.

> A personal data store is a set of capabilities built into a software programme or service that allows an individual to manage and maintain his or her digital information, artefacts and assets, longitudinally and self-sufficiently, so it may be used practically when and where it can for the individual's benefit as perceived by the individual, and shared with others directly, without relying on external third parties.[58]

Personal data stores can come in a variety of guises. Some take the form of cloud-based intermediary platforms that collect and integrate data from numerous other online services on behalf of individuals. Others can be installed locally on devices or servers operated by the individuals themselves. Some personal data stores are intended to be of general purpose, while others are more specific and confine themselves to dealing with particular types of data or offer a specific type of functionality.[59] Most, however, rely on the principles and standards of the architecture of the World Wide Web due to the way in which it provides a flexible platform for the development of highly interoperable network-based systems.[60] This reliance is in no small part due to recent advancements in technologies and protocols, which have made it easier than ever to integrate social features into diverse Web applications than at any other point in the Web's history.[61] From a technical perspective a personal data store typically operates in the following way. After an individual decides to install, or otherwise access and make use of, a personal data store service, they will then be provided with their own personal data store, which will then house their personal data in one central location. The user, as the central point of control, with a touch of a button, will then be able to make a determination as to when, which, and how much of their personal data they wish to share with other parties.[62] Accordingly, because of this functionality the individual will be far better equipped to monitor the location of their data, the identity of the parties that hold them, and what they are being used for.

It is precisely because of the heightened level of control personal data store services purport to offer that they have been identified as being capable of providing an alternative and more sophisticated approach to previously observed models of data management technologies; this control has led to them increasingly being talked about in both academic and commercial environments as a means through which the above-mentioned dearth of knowledge surrounding informa-

58 M. Van Kleek & K. O'Hara, 'The Future of Social is Personal: The Potential of the Personal Data Store', *in* D. Miorandi *et al*. (Eds.), *Social Collective Intelligence: Combining the Powers of Humans and Machines to Build a Smarter Society*, Berlin, Springer-Verlag, 2014, pp. 125-158.

59 R. Binns, 'Personal Data Empowerment and the Ideal Observer', *in* K. O'Hara *et al*. (Eds.), *Digital Enlightenment Yearbook 2014: Social Networks and Social Machines, Surveillance and Empowerment*, Amsterdam, IOS Press, 2014.

60 L. Dragan, M. Luczak-Roesch & N. Shadbolt, 'Understanding Personal Data as a Space – Learning from Dataspaces to Create Linked Personal Data', *Services and Applications over Linked APIs and Data (SALAD2014)*, Anissaras, Crete, Greece, 26 May 2014.

61 A. Sambra *et al*., 'Building Decentralized Applications for the Social Web', *Proceedings of the 25th International Conference Companion on World Wide Web*, 2016, pp. 1033-1034.

62 T. Kirkham *et al*., 'The Personal Data Store Approach to Personal Data Security', *IEEE Security and Privacy Magazine*, Vol. 11, No. 5, 2013, p. 13.

European Journal of Law Reform 2016 (18) 4
doi: 10.5553/EJLR/138723702016018004004

tion streams and data flows in the big data environment can be remedied.[63] Others have gone even further, suggesting that not only do personal data stores represent a likely or possible way of doing this, but that they may be the *only* way to liberate individuals from a 'toxic' environment where unseen and unaccountable data barons have a monopoly on the control and understanding of their personal data.[64] While the idea of individuals being given a controlling stake in the administration of their personal data is far from new, it is increasingly propounded that only recently we have arrived at a point where the notion has become technologically feasible.[65] While many personal data store services are still in development, they are already widely believed to have considerable promise so far as allowing individuals to develop an understanding of how their personal data are collected and used by others in the big data environment is concerned.[66] To take a few notable prominent examples, MyDex,[67] CPDS,[68] TAMIAS,[69] Avoco PDS,[70] and the Hub of all Things (HAT)[71] – all provide online platforms from which individuals are able to manually enter aspects of their personal data, visualize those data, and negotiate fine-grained data sharing agreements, allowing them to determine precisely which third parties are granted access to specific aspects of their personal data on a one-by-one basis. In a similar vein, openPDS/SafeAnswers provides

63 Notably, in a 2012 report entitled "Rethinking Personal Data: Strengthening Trust," the World Economic Forum identified personal data stores as one of the primary potential means by which individuals could be put in charge of their own data and provided with a greater sense of control. World Economic Forum, 2012, *supra* note 50.

64 *See* Wired.co.uk, 'Personal Data Stores Will Liberate Us from a Toxic Privacy Battleground', available at: <www.wired.co.uk/news/archive/2012-05/30/ideas-bank-personal-data-stores> (last accessed September 2016).

65 The concept of information intermediary-type services were first envisioned in the mid-1990s at least, and projects focusing on the development of individual personal data control have been in development ever since. *See* M. Becker, 'The Consumer Data Revolution: The Reshaping of Industry Competition and a New Perspective on Privacy', *Journal of Direct, Data and Digital Marketing Practice*, Vol. 15, No. 3, 2014, pp. 213-218.

66 Van Kleek & O'Hara, 2014, pp. 125-158.

67 MyDex, 'The Case for Personal Information Empowerment: The Rise of the Personal Data Store', 2010, available at: <https://mydex.org/wp-content/uploads/2010/09/The-Case-for-Personal-Information-Empowerment-The-rise-of-the-personal-data-store-A-Mydex-White-paper-September-2010-Final-web.pdf>; *See also* W. Heath *et al*., 'Digital Enlightenment, Mydex, and Restoring Control', *in* M. Hildebrandt *et al*. (Eds.), *Digital Enlightenment Yearbook 2013: The Value of Personal Data*, Amsterdam, IOS Press, 2013.

68 M. Chessa & P. Loiseau, 'CPDS: the Cooperative Personal Data Store for Managing Social Network Data', 2015, available at: <http://fontenellebio.fr/michelachessa/wp-content/uploads/2015/07/CoNext2015.pdf>.

69 J. Lorchat *et al*., 'TAMIAS: A Distributed Storage Built on Privacy and Identity', *Internet Initiative Japan*, 2012, available at: <https://tamias.iijlab.net/wp-content/uploads/2014/07/tnc2012_misc_Jean1.pdf>.

70 Avoco Identity, 'Avoco PDS: Personal Data Store', available at: <www.avocoidentity.com/avoco-platform/avoco-pds-personal-data-store/> (last accessed September 2016).

71 Hub of all Things, 'What is the HAT', available at: <http://hubofallthings.com/what-is-the-hat/> (last accessed July 2016).

comparable services for an individual's personal metadata, which tends to be observed and inferred rather than volunteered by the infidel themselves.[72]

In contrast to personal data stores and other similar technological tools that purport to give individuals greater *influence* in respect of how their personal data are shared with, and used by, others, transparency-enhancing technologies are not concerned with enhancing individual control. Instead, their objective is to provide the user with a greater *understanding* of how their personal data are stored, exchanged, processed, and used by others.[73] In other words, they are tools capable of providing the individual with clear visibility in respect of the use of their personal data.[74] Like personal data stores, transparency-enhancing technologies can come in a variety of forms. Some types, commonly known as data provenance tools,[75] for instance, provide the individual with greater insight into the data they have disclosed to another party. Others, on the other hand, attempt to provide a better comprehension of the way other parties attempt to use the personal data of an individual.[76] It is not difficult to see, therefore, how both varieties of these types of tool could have the potential to help alleviate the currently observable dearth of knowledge and understanding of information and data flows in the big data environment. Again, as with personal data stores, however, the potential of transparency-enhancing technologies is not merely a paper possibility. There are, in fact, a range of established and emerging tools of this type whose effectiveness and potential has been borne out by empirical findings. For instance, the so-called 'sticky policies' can allow for personal data to be effectively watermarked and which could prevent certain types of processing activities, and allow for the usage of such data to be monitored even after the individual has agreed to part with them.[77] Similarly, dynamic taint analysis tools, which allow individuals to track their personal data in real time, such as TaintDroid and DiOS, have been shown to be able to successfully monitor and trace personal data individuals share through smartphones and mobile applications, even after such data have left the individual's device and been passed on to other parties.[78] In a similar mould, XRay, a personal data tracking system developed by Columbia Engineer-

72 Y.-A. de Montjoye *et al*., 'openPDS: Protecting the Privacy of Metadata through SafeAnswers', *PLoS One*, Vol. 9, No. 7, 2014, p. e98790.

73 M. Janic & J. Wijbenga, 'Transparency Enhancing Tools (TETs): An Overview', *3rd workshop on Socio-Technical Aspects in Security and Trust (STAST)*, 2013, pp. 18-25.

74 M. Hansen, 'Marrying Transparency Tools with User-Controlled Identity Management', *Proceedings of Third International Summer School organised by IFIP*, Karlstad, Sweden, 2007; H. Hedbom, 'A Survey on Transparency Tools for Enhancing Privacy', *in* V. Matyáš *et al.* (Eds.), *The Future of Identity in the Information Society*, Berlin, Springer, 2009, pp. 67-82.

75 P. Buneman *et al*., 'Data Provenance: Some Basic Issues', *in* S. Kapoor & S. Prasad (Eds.), *FST TCS 2000: Foundations of Software Technology and Theoretical Computer Science*, Springer, New York, 2000.

76 M. Janic & J. Wijbenga, 2013, pp. 18-25.

77 S. Pearson & M. Mont, 'Sticky Policies: An Approach for Managing Privacy Across Multiple Parties', *Computer*, Vol. 44, No. 9, 2011. pp. 60-68.

78 W. Enck *et al*., 'TaintDroid: An Information-Flow Tracking System for Realtime Privacy Monitoring on Smartphones', *ACM Transactions on Computer Systems (TOCS)*, Vol. 32, No. 5, 2014, p. 5; A. Kurtz *et al*., 'DiOS: Dynamic Privacy Analysis of iOS Applications', *Friedrich-Alexander-Universitat Erlangen-Nurnberg, Department of Science Technical Reports*, 2014, ISSN 2191-5008.

doi: 10.5553/EJLR/138723702016018004004

ing, allows individuals to monitor which of their personal data, such as data found in emails, web searches, and viewed products, are being used to target them in a particular way (i.e., through targeted advertisements, pricing schemes, or the provision, or lack of provision, of particular goods or services).[79] While XRay is at the time of writing a prototype tool, empirical findings strongly suggest that it is of considerable promise, and demonstrate that the development and construction of scalable transparency architectures that can be used across huge networks, such as the entire World Wide Web, is most definitely achievable.[80]

Though many technological platforms and tools in the mould of those considered here are, as noted above, still at their developmental stages, there is a growing body of research that suggests that their progress and success to date is encouraging, and that they will be capable of dispelling some of the uncertainties currently surrounding data flows and information streams in the big data environment. Accordingly, the widespread deployment of such tools would likely benefit all major stakeholders in the big data environment in which there existed a *sui generis* model of data protection law. For instance, to the individual, the widespread adoption and deployment of these technologies would provide them with a means by which they could more effectively exercise their data protection rights against parties responsible for inappropriate uses of their personal data. To other major parties, like big data analytics firms and research organizations, the widespread adoption and deployment of these technologies would, due to the heightened sense of security and data protection they provided to individuals, help generate a sense of trust between them and the individuals whose personal data they sought, resulting in increased levels of data sharing and, accordingly allowing them to more fully reap the benefits of their data analyses. For these reasons there is a growing body of opinion that suggests they ought to be incorporated into prospective data protection regimes looking forward.[81] This article fully endorses this sentiment. If we are to accept that technological tools are a likely means by which desired regulatory objectives in the big data environment can be delivered, however, the question which then arises, therefore, is in what way should they be incorporated into prospective big data regulatory environments, and what role should the law play in their incorporation? One possibility that can immediately be dismissed is the idea of the law mandatorily imposing obligations on organizations that engage in big data analytics operations to incorporate specific technological tools and make them available to the individuals whose personal data they make use of. As has been noted extensively elsewhere, for instance, technology and the development thereof moves at a far greater pace

79 M. Lecuyer *et al*., 'XRay: Enhancing the Web's Transparency with Differential Correlation', *Proceedings of the 23rd USENIX Security Symposium*, San Diego, California, 2014.

80 *Ibid*.

81 *See, e.g.* S. Spiekermann & A. Novotny, 'A Vision for Global Privacy Bridges: Technical and Legal Measures for International Data Markets', *Computer Law & Security Review*, Vol. 31, No. 2, 2015, pp. 181-200.

than the law will ever be able to.[82] The implication being, of course, that the introduction of any legal requirement that imposed the use of specific technological tools on organizations engaged in big data analytics would run the risk of becoming rapidly outdated. For these reasons alone, such a course should not be pursued.

An alternative way forward that appears to be manifestly more promising, is the rollout of new standards, and a list of regulator-endorsed best available tools and practices, backed by law. The new standards could, for instance, denote scoring systems in respect of certain types of big data analytics operations and other associated data processing activities, with high scores being awarded to processing activities that involved particularly sensitive types of personal data, such as religious affiliations and health data, or to processing activities that were capable of having serious consequences for the individual whose personal data were involved, with comparatively lower scores being given to more routine processing activities, the consequences of which were deemed to be less severe. Following this, organizations and parties undertaking big data analytics operations could then have a variety of obligations placed on them depending on the aggregate score awarded to them as a result of the processing activities for which they were responsible, including, in certain situations, the mandatory deployment and incorporation of technological tools that allow individuals to either exercise greater control over their personal data, or develop a greater understanding of how they were being used. In a similar vein, the score awarded would correlate with a score awarded to technological tools considered and ranked in a list of best available tools and practices, agreed upon by a consensus of all European data protection authorities. Accordingly, to use a hypothetical example, a big data analytics firm whose data processing operations involved sensitive personal data, or were likely to lead to significant consequences for the individuals whose data were involved, would likely be awarded a high score, and thus might be required to incorporate relevant technological tools, be they of the personal data store, transparency enhancing, or other variety, of their choice, from the list of best tools and practices which aligned with that score. In so doing, this would in effect provide individuals sharing their data with such organizations with a means by which they could effectively monitor their personal data and allow them to invoke legal safeguards and rules in the event their personal data were misused. So to ensure that the relevant standards and best available techniques were obeyed and effectively enforced, the law would impose severe penalties against those who ignored them, as well as against those found to have illegally acquired, possessed, used, or sold personal data. So to be made meaningful, firms and organizations that score highly would also be subject to routine audits and monitoring, for which data protection authorities of each Member State would be responsible, to ensure that they were fulfilling their responsibilities. Alongside this, the rollout of digital literacy educational initiatives would also be required, so that individuals

82 J. Barlow, 'Selling Wine without Bottles', *in* P. Ludlow (Ed.), *High Noon on the Electronic Frontier: Conceptual Issues in Cyberspace*, London, MIT Press, 1999, pp. 9-24. *See also* Brownsword & Goodwin, 2012, p. 61.

European Journal of Law Reform 2016 (18) 4
doi: 10.5553/EJLR/138723702016018004004

in the big data regulatory environment were not only made aware of the existence of the technological empowerment tools that may aid in enhancing the protection of their personal data, but so that they are able to operate them competently and, in so doing, prevent manifestation of security vulnerabilities.[83]

The result of these combined endeavours ought to be the construction of a system of data protection law and regulation that allows the alleviation of the current dearth of knowledge and understanding of data sharing and handling practices in the big data environment. Accordingly, they could pave the way for a successful shift to a *sui generis* model of data protection, which, as noted above, would target certain uses of personal data, rather than their collection, negating the need for a data minimization principle. In so doing, by following these steps, it should be possible to enact a system of data protection law and regulation that resembled the desired 'ends' identified above inasmuch as it would afford individuals with meaningful safeguards against potential abuses, allowing them to enjoy a heightened level of data protection while, concurrently, the free flow of personal data in the internal market would not be unduly restricted.

IV Resource Planning

As noted above, the resource planning stage requires the determination of what resources will be required in order for the identified means to reach the desired ends. As also noted above, when undertaking the ends and means phases of any deployment of Interactive Planning, financial constraints should generally not act as an impediment to the identification and achievement of desired goals. While an exact figure of the likely costs that would be incurred as a result of the construction of a big data regulatory environment as envisaged above would be extremely complicated to calculate, and such a calculation was beyond the scope of the research that this article is based on, it is still possible to advance some comment on the resources that would likely be required.

General administrative uncertainty aside, it seems certain that the construction of a *sui generis* model of data protection as a part of a wider big data regulatory environment, in the manner outlined above, would be fairly resource intensive. This is true in respect of both technological development and regulatory oversight. For instance, in order for technological empowerment tools to continue to be developed at a rate that keeps pace with the evolution of data sharing practices and analytical techniques, so that they remain functional and fit for pur-

83 A recent story regarding security concerns associated with Network Attached Storage systems, regarding large amounts of personal data being leaked online, revealed that such problems were exacerbated by individuals not possessing the knowledge to correctly configure their devices correctly, and serves as a pertinent example, and a stark warning, of this sort of possibility. *See* BBC, 'Personal Data Stores Found Leaking Online', available at: <www.bbc.co.uk/news/technology-28707117> (last accessed September 2016). In a similar vein, it has been suggested, that personal data stores would have to pass the test of convenience ('the Mum test') if they were to ever be prevalently used. *See* W. Heath, 'Personal Data Stores', Society for Computers and Law, 2014, available at: <www.scl.org/site.aspx?i=ed38100>. *See also* H. Jenkins, *Confronting the Challenges of Participatory Culture: Media Education for the 21st Century*, Cambridge, MA, MIT Press, 2009.

pose, funding will evidently be required.[84] Any public education initiatives that are required in order for the relevant technological tools to be operated effectively will obviously also require funding. The administrative upheaval of moving from a model of data protection based on fair information principles to a one that is of a *sui generis* character would also surely incur substantial costs, many of which would probably be unforeseeable. At the same time, it is widely acknowledged at present that data protection authorities in many EU Member States are already perceived to be worryingly underfunded.[85] As has been noted elsewhere, however, the construction of any wide-ranging regulatory regime that places rigorous monitoring responsibilities on such bodies, or extends those that are already in place, will inevitably require their funding to be drastically stepped up.[86]

V Design Implementation and Control

Having outlined the main features of the regulatory model designed in the article's preceding sections, the final stage of the Interactive Planning process now requires that the way in which the desired design is to be implemented and controlled also be outlined. Though, as with the resource planning stage of the Interactive Planning procedure, the process of shifting to a *sui generis* model of data protection from the model based on fair information principles in place currently would undoubtedly require a great deal of administrative and logistical upheaval, such as the enactment of new legislation, with many of the associated challenges not being foreseeable in advance, and are thus beyond the scope of this article's research. Nevertheless, as with the resource planning stage, it is once again possible to advance some specific comments in respect of design implementation and control.

First of all, as outlined in the preceding sections, a shift to a *sui generis* model of data protection in the manner envisaged above would require the drafting of fresh and novel standards, and the compilation and maintenance of a list of best available technological tools and practices. These should be the result of discussions between the European Commission, national data protection authorities, technological experts, and other relevant regulatory bodies like the Article 29

84 One of the primary challenges that must be wrestled with if technological empowerment tools are to represent a suitable mechanism by which individuals can take charge of their personal data and protect them from abuses is the fact that their technical capabilities are likely to change drastically as time progresses. It will therefore be imperative that they are designed in ways that accommodate such changes as they arise, so that they can be effectively 'future proofed'. M. Van Kleek & K. O'Hara, 'The Future of Social is Personal: The Potential of the Personal Data Store', *in* D. Miorandi *et al*. (Eds.), *Social Collective Intelligence: Combining the Powers of Humans and Machines to Build a Smarter Society*, Berlin, Springer-Verlag, pp. 125-158.

85 As a case in point, for instance, the United Kingdom Information Commissioner's Office has recently cast doubt on how it will fund its operations in the future. Computer Weekly, 'ICO Reports Progress in Data Protection, But Funding Remains a Concern', available at: <www.computerweekly.com/news/4500249237/ICO-reports-progress-in-data-protection-but-funding-remains-a-concern> (last accessed September 2016). *See also* Financial Times, 'Data Protection Agencies Gain Power from Google Defeat', available at: <www.ft.com/cms/s/0/157eecaa-db82-11e3-b112-00144feabdc0.html> (last accessed September 2016).

86 Koops, 2014.

European Journal of Law Reform 2016 (18) 4
doi: 10.5553/EJLR/138723702016018004004

Working Party. The maintenance and updating of these standards must be a continuous process, and thus it will be vital that they are periodically reviewed and considered against the background of the development of data-handling practices in the big data environment, so to ensure that they remain fit for purpose, and do not fall into obsolescence. In a similar vein, it will be vital that research and development in relation to technological empowerment tools is also continued, so that they too can keep pace with the ever-changing practical realities of the big data environment, and remain fit for purpose for inclusion on the above-mentioned list of best available technological tools and practices. The manifestation of any potential security or user safety issues must also be continually monitored.

Once in place, it will be vital that the data protection authorities of member states identify and carry out regulator auditing exercises on private companies and other organizations that are responsible for analytical operations that are adjudged to be deserving of regulatory attention. The purposes of these exercises must be to ensure that the audited parties are fully meeting the obligations imposed by them by the relevant standards and best practices. Again, ascertaining the level of obligations imposed on organizations that involve themselves in big data analytics operations must be a continuous process to ensure any such obligations are commensurate to the data processing activities for which they are responsible.

F Conclusions and Thoughts for Further Research

This article has, by considering the emergent challenges posed by big data to the smooth operation of the principle of purpose limitation, attempted to make the argument that systems thinking methodologies and, in particular, Interactive Planning, as popularized by Ackoff, are capable of playing a key role in the development of data protection law and policy in the big data environment. To this end, the result of the above analyses is the presentation of a policy response to the challenges posed by big data to data protection law's principle of data minimization, which, it is argued, is well-researched, credible, and conceivable. By using an Interactive Planning approach, it was possible to consider the underlying challenges posed to the principle by big data's emergence, consider the views of a number of relevant stakeholders, devise a desirable end goal, devise a means by which barriers restricting the achievement of this goal could be removed, and devise a means by which the identified plan could be implemented. The primary conclusion of the analyses undertaken was that the European data protection framework might best be served by shifting to a model of data protection that focuses on certain *uses* of personal data, underpinned by personal empowerment technologies and new legal rules, and move away from the current model, which is based on fair information principles and focuses on the *collection* of such data. In this new model of data protection, individuals would be afforded with meaningful technological and legal safeguards against potential abuses stemming from the processing of their personal data, allowing them to enjoy a heightened level of data protection while, concurrently, the free flow of personal data in the internal

market would not be unduly restricted. In so doing, the article sketched a plausible and potentially promising prospective way in which above-mentioned problems facing the principle of data minimization can be *dissolved*, rather than *resolved*, and a desirable future pursued. As noted at the article's outset, however, it was not the purpose of this article to present a definitive solution to this particular policy challenge, nor any others posed by the emergence of big data. Instead, the objective of this article was to highlight the potential of Interactive Planning as a methodological means of improving data protection law and policy in the big data environment during both development and implementation stages. By using the principle of data minimization and its associated challenges as a particular case study and, by subjecting these challenges to the Interactive Planning methodology, allowing for the development of potentially promising prospective ways forward, it is tentatively argued that this objective has been achieved.

The principle of data minimization is, however, just one particular aspect of data protection law where the emergence of big data requires policy responses, and arguably a re-gearing of long established regulatory principles and maxims. There are other pressing policy areas associated with big data's emergence and the resultant effects on European data protection law. Notably, and as alluded to above, for instance, there are serious concerns associated with the suitability of continued reliance on at least two other facets central to the data protection framework, namely: informed consent and the concept of personal data itself. In light of this, there is clearly ample scope to examine how systems methodologies, including those other than Interactive Planning, can help with emergent policy challenges in these areas, and possibly beyond. In relation to the particular 'solutions' proposed in this article, there is also clearly ample scope for more research to be undertaken in relation to the types of personal empowerment technologies considered above, and precisely how it is they may be capable of contributing to the achievement of regulatory objectives in the big data environment.

European Journal of Law Reform 2016 (18) 4
doi: 10.5553/EJLR/138723702016018004004