
Privacy, Big Data, and the Public Good

Chapter 9: The New Deal on Data: A Framework for Institutional Controls

Lead Author: Daniel "Dazza" Greenwood
With: Arkadiusz Stopczynski, Brian Sweatt, Thomas Hardjono and Alex "Sandy" Pentland

Introduction

In order to realize the promise of a Big Data society and to reduce the potential risk to individuals, institutions are updating the operational frameworks which govern the business, legal, and technical dimensions of their internal organizations. In this chapter we outline ways to support the emergence of such a society within the framework of the New Deal on Data, and describe future directions for research and development. In our view, the traditional control points relied on as part of corporate governance, management oversight, legal compliance , and enterprise architecture must evolve and expand to match operational frameworks for big data. These controls must support and reflect greater user control over personal data, as well as large-scale interoperability for data sharing between and among institutions. The core capabilities of these controls should include responsive rule-based systems governance and fine-grained authorizations for distributed rights management.

The New Realities of Living in a Big Data Society

Building an infrastructure that sustains a healthy, safe, and efficient society is, in part, a scientific and engineering challenge which dates back to the 1800s when the Industrial Revolution spurred rapid urban growth. That growth created new social and environmental problems. The remedy then was to build centralized networks that delivered clean water and safe food, enabled commerce, removed waste, provided energy, facilitated transportation, and offered access to centralized health care, police , and educational services. These networks formed the backbone of society as we know it today.

These century-old solutions are, however, becoming increasingly obsolete and inefficient. We now face the challenges of global warming, uncertain energy, water, and food supplies, and a rising population and urbanization that will add 350 million people to the urban population by 2025 in China alone. 1 The new challenge is how to build an infrastructure that enables cities to be energy efficient, have secure food and water supplies, be protected from pandemics, and to have better governance. Big data can enable us to achieve such goals. Rather than static systems separated by function – water, food, waste, transport , education, energy – we can instead regard the systems as dynamic, data-driven networks . Instead of focusing only on access and distribution, we need networked and self-regulating systems, driven by the needs and preferences of citizens – a ‘nervous system’ that maintains the stability of government, energy, and public health systems around the globe. A control framework should be established which

enables data to be captured about different situations, those observations to be combined with models of demand and dynamic reaction, and the resulting predictions to be used to tune the nervous system to match those needs and preferences.

The engine driving this nervous system is big data: the newly ubiquitous digital data now available about so many aspects of human life. We can analyze patterns of human activity within the digital breadcrumbs we all leave behind as we move through the world: call records, credit card transactions, GPS location fixes, among others. 2 These data, which record actual activity, may be very different from what we put on Facebook or Twitter; our postings there are what we choose to tell people, edited according to the standards of the day and filtered to match the persona we are building. Although mining social networks can give great insight into human nature, 3 the value is limited for operational purposes. 4

The process of analyzing the patterns within these digital breadcrumbs is called 'reality mining.' 5 The Human Dynamics research group at MIT found that these patterns can be used to tell us if we are likely to get diabetes, 6 or whether we are the sort of person who will pay back loans. 7 By analyzing them across many people, we are discovering that we can begin to explain many things – crashes , revolutions, bubbles – that previously appeared unpredictable. 8 For this reason , the magazine MIT Technology Review named our development of reality mining one of the 10 technologies that will change the world. 9

The New Deal on Data

The digital breadcrumbs we leave behind are clues to who we are, what we do, and what we want. This makes personal data – data about individuals – immensely valuable, both for public good and for private companies. As the European Consumer Commissioner , Meglena Kuneva, said recently, "Personal data is the new oil of the Internet and the new currency of the digital world." 10 The ability to see details of so many interactions is also immensely powerful and can be used for good or for ill. Therefore, protecting personal privacy and freedom is critical to our future success as a society. We need to enable more data sharing for the public good; at the same time, we need to do a much better job of protecting the privacy of individuals.

A successful data-driven society must be able to guarantee that our data will not be abused – perhaps especially that government will not abuse the power conferred by access to such fine-grained data. There are many ways in which abuses might be directly targeted – from imposing higher insurance rates based on individual shopping history, 11 to creating problems for the entire society, by limiting user choices and enclosing users in information bubbles. 12 To achieve the potential for a new society, we require the New Deal on Data, which describes workable guarantees that the data needed for public good are readily available while at the same time protecting the citizenry. 13

The key insight behind the New Deal on Data is that our data are worth more when shared. Aggregate data – averaged, combined across population, and often distilled to high-level features – can be used to inform improvements in systems such as public health, transportation, and government. For instance, we have demonstrated that data about the way we behave and where we go can be used to minimize the spread of infectious disease. 14 Our research has also shown how digital breadcrumbs can be used to track the spread of influenza from person to person on an individual level. And the public good can be served as a result: if we can see it, we can also stop it. Similarly, if we are worried about global warming, shared, aggregated data can reveal how patterns of mobility relate to productivity . 15 This, in turn, equips us to design cities that are more productive and, at the same time, more energy efficient. However, to obtain these results and make a greener world, we must be able to see people moving around; this depends on having many people willing to contribute their data, if only anonymously and in aggregate. In addition, the Big Data transformation can help society find efficient means of governance by providing tools to analyze and understand what needs to be done, and to reach consensus on how to do it. This goes beyond simply creating more communication platforms; the assumption that more interaction between users will produce better decisions may be

very misleading. Although in recent years we have seen impressive uses of social networks for better organization in society, for example during political protests, 16 we are far from even starting to reach consensus about the big problems: epidemics, climate change, pollution – big data can help us achieve such goals.

However, to enable the sharing of personal data and experiences, we need secure technology and regulation that allows individuals to safely and conveniently share personal information with each other, with corporations, and with government. Consequently, the heart of the New Deal on Data must be to provide both regulatory standards and financial incentives enticing owners to share data, while at the same time serving the interests of individuals and society at large. We must promote greater idea flow among individuals, not just within corporations or government departments.

Unfortunately, today most personal data are siloed in private companies and therefore largely unavailable. Private organizations collect the vast majority of personal data in the form of mobility patterns, financial transactions, and phone and Internet communications . These data must not remain the exclusive domain of private companies, because they are then less likely to contribute to the common good; private organizations must be key players in the New Deal on Data. Likewise, these data should not become the exclusive domain of the government. The entities who should be empowered to share and make decisions about their data are the people themselves: users, participants, citizens. We can involve both experts and use the wisdom of crowds – users themselves interested in improving society.

Personal Data: Emergence of a New Asset Class

One of the first steps to promoting liquidity in land and commodity markets is to guarantee ownership rights so that people can safely buy and sell. Similarly, a first step toward creating more ideas and greater flow of ideas – idea liquidity – is to define ownership rights. The only politically viable course is to give individual citizens key rights over data that are about them, the type of rights that have undergirded the European Union's Privacy Directive since 1995. 17 We need to recognize personal data as a valuable asset of the individual, which can be given to companies and government in return for services.

We can draw the definition of ownership from English common law on ownership rights of possession, use, and disposal:

- You have the right to possess data about yourself. Regardless of what entity collects the data, the data belong to you, and you can access your data at any time. Data collectors thus play a role akin to a bank, managing data on behalf of their 'customers'.

- You have the right to full control over the use of your data. The terms of use must be opt in and clearly explained in plain language. If you are not happy with the way a company uses your data, you can remove the data, just as you would close your account with a bank that is not providing satisfactory service.

- You have the right to dispose of or distribute your data. You have the option to have data about you destroyed or redeployed elsewhere.

Individual rights to personal data must be balanced with the need of corporations and governments to use certain data- account activity, billing information, and the like to run their day-to-day operations. The New Deal on Data therefore gives individuals the right to possess, control, and dispose of copies of these required operational data, along with copies of the incidental data collected about the individual, such as location and similar context. These ownership rights are not exactly the same as literal ownership under modern law; the practical effect is that disputes are resolved in a different , simpler manner than would be the case for land ownership disputes, for example.

In 2007, one author (AP) first proposed the New Deal on Data to the World Economic Forum. 18 Since then, this idea has run through various discussions and eventually helped to shape the 2012 Consumer Data Bill of Rights in the United States, along with a matching declaration on Personal Data Rights in the European Union.

The World Economic Forum (WEF) echoed the European Consumer Commissioner Meglena Kuneva in dubbing personal data the 'new oil' or new resource of the 21st century. 19 The 'personal data sector' of the economy today is in its infancy, its state akin to the oil industry during the late 1890s. Productive collaboration between government (building the state-owned freeways), the private sector (mining and refining oil, building automobiles), and the citizens (the user-base of these services) allowed developed nations to expand their economies by creating new markets adjacent to the automobile and oil industries.

If personal data, as the new oil, is to reach its global economic potential, productive collaboration is needed between all stakeholders in the establishment of a personal data ecosystem. A number of fundamental uncertainties exist, however, about privacy, property, global governance, human rights – essentially about who should benefit from the products and services built on personal data. 20 The rapid rate of technological change and commercialization in the use of personal data is undermining end-user confidence and trust.

The current personal data ecosystem is feudal, fragmented, and inefficient. Too much leverage is currently accorded to service providers that enroll and register end -users. Their siloed repositories of personal data exemplify the fragmentation of the ecosystem, containing data of varying qualities; some are attributes of persons that are unverified, while others represent higher quality data that have been cross-correlated with other data points of the end-user. For many individuals, the risks and liabilities of the current ecosystem exceed the economic returns. Besides not having the infrastructure and tools to manage personal data, many end-users simply do not see the benefit of fully participating. Personal privacy concerns are thus addressed inadequately at best, or simply overlooked in the majority of cases. Current technologies and laws fall short of providing the legal and technical infrastructure needed to support a well-functioning digital economy.

Recently, we have seen the challenges, but also the feasibility of opening private big data. In the Data for Development (D4D) Challenge (http://www.d4d.orange.com), the telecommunication operator Orange opened access to a large dataset of call detail records from the Ivory Coast. Working with the data as part of a challenge, teams of researchers came up with life-changing insights for the country. For example, one team developed a model for how disease spreads in the country and demonstrated that information campaigns based on one-to-one phone conversations among members of social groups can be an effective countermeasure. 21 Data release must be carefully done, however; as we have seen in several cases, such as the Netflix Prize privacy disaster 22 and other similar privacy breaches, 23 true anonymization is extremely hard – recent research by de Montjoye et al. and others 24 , 25 has shown that even though human beings are highly predictable, we are also unique. Having access to one dataset may be enough to uniquely fingerprint someone based on just a few data points, and this fingerprint can be used to discover their true identity. In releasing and analyzing the D4D data, the privacy of the people who generated the data was protected not only by technical means, such as removal of personally identifiable information (PII), but also by legal means, with the researchers signing an agreement that they would not use the data for re-identification or other nefarious purposes. Opening data from the silos by publishing static datasets – collected at some point and unchanging – is important, but it is only the first step. We can do even more when data is available in real time and can become part of a society`s nervous system. Epidemics can be monitored and prevented in real time, 26 underperforming students can be helped, and people with health risks can be treated before they get sick. 27

The report of the World Economic Forum 28 suggests a way forward by identifying useful areas on which to focus efforts:

- Alignment of key stakeholders Citizens, the private sector, and the public sector need to work in support of one another. Efforts such as NSTIC 29 in the United States – albeit still in its infancy – represent a promising direction for global collaboration.

- Viewing ‘data as money’ There needs to be a new mindset, in which an individual’s personal data items are viewed and treated in the same way as their money. These personal data items would reside in an ‘account’ (like a bank account) where they would be controlled, managed, exchanged, and accounted for just as personal banking services operate today.

- End-user centricity All entities in the ecosystem need to recognize end-users as vital and independent stakeholders in the co-creation and exchange of services and experiences. Efforts such as the User Managed Access (UMA) initiative 30 provide examples of system design that are user-centric and managed by the user.

Enforcing the New Deal on Data

How can we enforce this New Deal? The threat of legal action is important , but not sufficient; if you cannot see abuses, you cannot prosecute them. Enforcement can be addressed significantly without prosecution or public statute or regulation. In many fields, companies and governments rely on rules governing common business, legal, and technical (BLT) practices to create effective self-organization and enforcement. This approach holds promise as a method by which institutional controls can form a reliable operational framework for big data, privacy, and access.

One current best practice is a system of data sharing called a ‘trust network’, a combination of networked computers and legal rules defining and governing expectations regarding data. For personal data, these networks of technical and legal rules keep track of user permissions for each piece of data and act as a legal contract, specifying what happens in case of a violation. For example, in a trust network all personal data can have attached labels specifying where the data come from and what they can and cannot be used for. These labels are exactly matched by the terms in the legal contracts between all of the participants, stating penalties for not obeying them. The rules can – and often do – reference or require audits of relevant systems and data use, demonstrating how traditional internal controls can be leveraged as part of the transition to more novel trust models . A well-designed trust network, elegantly integrating computer and legal rules, allows automatic auditing of data use and allows individuals to change their permissions and withdraw data.

The mechanism for establishing and operating a trust network is to create system rules for the applications, service providers, data, and the users themselves. System rules are sometimes called ‘operating regulations’ in the credit card context, ‘trust frameworks’ in the identity federation context, or ‘trading partner agreements’ in a supply value chain context. Several multiparty shared architectural and contractual rules create binding obligations and enforceable expectations on all participants in scalable networks. Furthermore, the design of the system rules allows participants to be widely distributed across heterogeneous business ownership boundaries, legal governance structures, and technical security domains. However, the parties need not conform in all or even most aspects of their basic roles, relationships, and activities in order to connect to a trust network. Cross-domain trusted systems must – by their nature – focus enforceable rules narrowly on commonly agreed items in order for that network to achieve its purpose.

For example, institutions participating in credit card and automated clearing house networks are subject to profoundly different sets of regulations, business practices, economic conditions, and social expectations. The network rules focus on the topmost agreed items affecting interoperability, reciprocity, risk , and revenue allocation. The knowledge that fundamental rules are subject to enforcement action is one of the foundations of trust and a motivation to prevent or address violations before they trigger penalties. A clear example of this approach can be found in the Visa Operating Rules, which cover a vast global real-time network of parties agreeing to rules governing their roles in the system as merchants , banks, transaction processors, individual or business card holders, and other key system roles.

Such rules have made the interbank money transfer system among the safest systems in the world and the backbone for daily exchanges of trillions of dollars, but until recently those were only for the ‘big guys’. 31 To give individuals a similarly safe method of managing personal data, the Human

Dynamics group at MIT, in partnership with the Institute for Data Driven Design (co-founded by John Clippinger and one author (AP)) have helped to build an open Personal Data Store (openPDS). 32 The openPDS is a consumer version of a personal cloud trust network now being tested with a variety of industry and government partners. The aim is to make sharing personal data as safe and secure as transferring money between banks.

When dealing with data intended to be accessible over networks – whether big, personal, or otherwise – the traditional container of an institution makes less and less sense. Institutional controls apply, by definition, to some type of institutional entity such as a business, governmental, or religious organization. A synopsis of all the BLT facts and circumstances surrounding big data is necessary in order to know what access, confidentiality, and other expectations exist; the relevant contextual aspects of big data at one institution are often profoundly different from those at another. As more and more organizations use and rely on big data, a single formula for institutional controls will not work for increasingly heterogeneous BLT environments.

The capacity to apply appropriate methods of enforcement for a trust network depends on clear understanding and agreement among the parties about the purpose of the system and the respective roles or expectations of those connecting as participants. Therefore , some contextual anchor is needed to have a clear basis for establishing an operational framework and institutional controls appropriate for big data.

Transitioning End-User Assent Practices

The way users grant authorization to share their data is not a trivial matter. The flow of personal information such as location data, purchases, and health records can be very complex. Every tweet, geotagged picture, phone call, or purchase with credit card provides the user's location not only to the primary service, but also to all the applications and services that have been authorized to access and reuse these data. The authorization may come from the end-user or be granted by the collecting service, based on umbrella terms of service that cover reuse of the data. Implementation of such flows was a crucial part of the Web 2.0 revolution, realized with RESTful APIs, mash-ups, and authorization-based access. The way personal data travels between services has arguably become too complex for a user to handle and manage.

Increasing the range of data controlled by the user and the granularity of this control is meaningless if it cannot be exercised in an informed way. For many years, a poor model has been provided by End User License Agreements (EULAs), long incomprehensible texts that are accepted blindly by users trusting they have not agreed to anything that could harm them. The process of granting meaningful authorization cannot be too complex, as it would prevent a user from understanding her decisions. At the same time, it cannot be too simplistic, as it may not sufficiently convey the weight of the privacy-related decisions it captures. It is a challenge in itself to build end-user assent systems that allow users to understand and adjust their privacy settings.

This gap between the interface – single click – and the effect can render data ownership meaningless; one click may wrench people and their data into systems and rules that are antithetical to fair information practices, as is prevalent with today's end-user licenses in cloud services or applications. Managing the long-term tensions fueled by 'old deal' systems operating simultaneously with the New Deal is an important design and migration challenge during the transition to a Big Data economy. During this transition and after the New Deal on Data is no longer new, personal data must continue to flow in order to be useful. Protecting the data of people outside of directly user-controlled domains is very hard without a combination of cost-effective and useful business practices, legal rules, and technical solutions.

We envision 'living informed consent', where the user is entitled to know what data is being collected about her by which entities, empowered to understand the implications of data sharing, and finally put in charge of the sharing authorizations. We suggest that readers ask themselves a question: Which services know which city I am in today? Google? Apple? Twitter? Amazon? Facebook? Flickr? Some app I authorized a few years ago to

access my Facebook check-ins and have since forgotten about? This is an example of a fundamental question related to user privacy and assent, and yet finding an accurate answer can be surprisingly difficult in today's ecosystem. We can hope that most services treat data responsibly and according to user authorizations. In the complex network of data flows, however, it is relatively easy for data to leak to careless or malicious services. 33 We need to build solutions that help users to make well-informed decisions about data sharing in this environment.

Big Data and Personal Data Institutional Controls

The concept of 'institutional controls' refers to safeguards and protections implemented through legal, policy, governance, and other measures that are not solely technical, engineering, or mechanical. Institutional controls in the context of big data can perhaps best be understood by examining how such controls have been applied to other domains, most prevalently in the field of environmental regulation. A good example of how this concept supports and reflects the goals and objectives of environmental regulation can be found in the policy documents of the Environmental Protection Agency (EPA), which gives the following definition in its Institutional Controls Glossary:

Institutional Controls – Non-engineering measures intended to affect human activities in such a way as to prevent or reduce exposure to hazardous substances. They are almost always used in conjunction with, or as a supplement to, other measures such as waste treatment or containment. There are four categories of institutional controls: governmental controls; proprietary controls; enforcement tools; and informational devices. 34

The concept of an 'institutional control boundary' is especially clarifying and powerful when applied to the networked and digital boundaries of an institution . In the context of Florida's environmental regulation, the phrase is applied when a property owner's risk management and clean-up responsibilities extend beyond the area defined by the physical property boundary. For example, a recent University of Florida report on clean-up target levels (CTLs) states, "in some rare situations, the institutional control boundary at which default CTLs must be met can extend beyond the site property boundary." 35

When institutional controls apply to "separately owned neighboring properties" a number of possibilities arise that are very relevant to management of personal data across legal, business, and other systemic boundaries. Requiring the party responsible for site clean-up to use "best efforts" to attain agreement from the neighboring owners to institute the relevant institutional controls is perhaps the most direct and least prescriptive approach. When direct negotiated agreement is unsuccessful, then use of third-party neutrals to resolve disagreements regarding institutional controls can be required. If necessary, environmental regulation can force the acquisition of neighboring land by compelling the party responsible to purchase the other property or by purchase of the property directly by the EPA. 36

In the context of big data, institutional controls are seldom, if ever, imposed through government regulatory frameworks such as are seen in environmental waste management oversight by the EPA. 37 Rather, institutions applying measures constituting institutional controls in the big data and related information technology and enterprise architecture contexts will typically employ governance safeguards, business practices, legal contracts, technical security, reporting , and audit programs and various risk management measures.

Inevitably, institutional controls for big data will have to operate effectively across institutional boundaries, just as environmental waste management must sometimes be applied across real property boundaries and may subject multiple different owners to enforcement actions corresponding to the applicable controls. Short of government regulation, the use of system rules as a general model is one widely understood, accepted, and efficient method for defining, agreeing, and enforcing institutional and other controls across BLT domains of ownership, governance, and operation.

Following on from the World Economic Forum's recommendation to treat personal data stores in the manner of bank accounts, 38 a number of infrastructure improvements need to be realized if the personal data ecosystem is to flourish and deliver new economic opportunities:

- New global data provenance network In order for personal data stores to be treated like bank accounts, origin information regarding data items coming into the data store must be maintained. 39 In other words, the provenance of all data items must be accounted for by the IT infrastructure on which the personal data store operates. The databases must then be interconnected in order to provide a resilient, scalable platform for audit and accounting systems to track and reconcile the movement of personal data from different data stores.

- Trust network for computational law For trust to be established between parties who wish to exchange personal data, some degree of 'computational law' technology may have to be integrated into the design of personal data systems. This technology should not only verify terms of contracts (e.g. terms of data use) against user-defined policies but also have mechanisms built in to ensure non-repudiation of entities who have accepted these digital contracts. Efforts such as the UMA initiative are beginning to bring better evidentiary proof and enforceability of contracts into technical protocol flows. 40

- Development of institutional controls for digital institutions Currently, a number of proposals for the creation of virtual currencies (e.g. BitCoin, 41 Ven 42) have underlying systems with the potential to evolve into self-governing 'digital institutions'. 43 Such systems and the institutions that operate on them will necessitate the development of a new paradigm to understand aspects of institutional control within their context.

Scenarios of Use in Context

Developing frameworks for big data that effectively balance economic , legal, security, and other interests requires an understanding of the relevant context and applicable scenarios within which the data exists.

A sound starting point from which to establish the applicable scenarios of use is to enumerate the institutions involved with a given set of big data, and develop a description of how or why they hold, access, or otherwise intermediate the data. Although big data straddles multiple BLT boundaries, one or more institutions are typically able to, or in some situations required to, manage and control the data. The public good referred to in the title of this book can be articulated as design requirements or even as certification criteria applicable to those institutions that operate the systems through which the big data is computed or flows.

It may be also be necessary to narrowly define certain aspects of the scenario in which the data exist in order to establish the basic ownership, control, and other expectations of the key parties. For example, describing a transaction as a financial exchange may not provide enough relevant detail to reveal the rights, obligations , or other outcomes reasonably expected by the individuals and organizations involved. The sale of used cars via an app, the conduct of a counseling session via Google Hangout, and the earning of a master's degree via an online university all represent scenarios in which the use case of a financial exchange takes place. However, each of these scenarios occurs in a context that is easily identifiable: the sale of goods and deeper access to financial information if the car is financed; the practice of therapy by a licensed professional accessing and creating confidential mental health data; or e-learning services and protected educational records and possibly deeper financial information if the program is funded by scholarship or loans. The scenarios can also identify the key elements necessary to establish existing consumer rights – the people (a consumer and a used car dealer), the transaction (purchase of a used car), the data (sales and title data, finance information, etc.), and the systems (the third-party app and its relevant services or functions, state DMV services, credit card and bank services, etc.). The rights established by relevant state lemon laws, the Uniform Commercial Code, and other applicable rules will determine when duties arise or are terminated, what must be promised, what can be repudiated, by whom data must be kept secure, and other requirements or constraints on the use of personal data and big data. These and other factors

differ when a transaction that seems identical operates within a different scenario, and even scenarios will differ depending on which contexts apply. The following four elements are critical for defining high-level goals and objectives:

1. Who are the people in the scenario (e.g. who are the parties involved and what are their respective roles and relationships)?

2. What are the relevant interactions (e.g. what transactions or other actions are conducted by or with the people involved)?

3. What are the relevant data and datasets (e.g. what types of data are created, stored, computed, transmitted, modified, or deleted)?

4. What are the relevant systems (e.g. what services or other software are used by the people, for the transactions, or with the data)?

Inspired by common law, the New Deal on Data sets out general principles of ownership that both guide and inform basic relationships and expectations. However, the dynamic bundle of recombinant rights and responsibilities constituting 'ownership' interests in personal data and expectations pertaining to big data vary significantly from context to context, and even from one scenario to another within a given general context. Institutional controls and other system safeguards are important methods to ensure that there are context-appropriate outcomes that are consistent with clearly applicable system scenarios as well as the contours and foundations for a greater public good. The New Deal on Data can be achieved in part by sets of institutional controls involving governance, business, legal, and technical aspects of big data and interoperating systems. Reference scenarios can be used to reveal signature features of the New Deal on Data in various contexts and can serve as anchors in evaluating what institutional controls are well aligned to achieve a balance of economic, privacy, and other interests.

The types of requirements and rules governing participation by individuals and organizations in trust networks vary depending on the facts and circumstances of the transactions, data types, relevant roles of people, and other factors. Antecedent but relevant networks such as credit card systems, trading partner systems, and exchange networks are instructive not only for their many common elements but also as important examples of how vastly different they are from one another in their contexts, scenarios, legal obligations, business models , technical processes, and other signature patterns . Trust networks that are formed to help manage big data in ways that appropriately respect personal data rights and other broader interests will similarly succeed to the extent they can tolerate or promote a wide degree of heterogeneity among participants for BLT matters that need not be uniform or directly harmonized. In some situations, new business models and contexts will emerge that require fresh thinking and novel combinations of roles or types of relationships among transacting parties. In these cases, understanding the actual context and scenarios is critical in customizing acceptable and sustainable BLT rules and systems. Example scenarios can describe deeper fact-based situations and circumstances in the context of social science research involving personal data and big data. 44 The roles of people, their interactions, the use of data, and the design of the corresponding systems reflect and support the New Deal on Data in ways that deliberately provide greater immediate value to stakeholders than is typically expected.

The New Deal on Data is designed to provide good value to anyone creating, using, or benefiting from personal data, but the vision need not be adopted in its entirety before its value becomes apparent. Its principles can be adopted on a large scale in increments – an economic sector, transaction type, or data type at a time. Adopting the New Deal on Data in successive phases helps to address typical objections to change based on cost, disruption , or overregulation. Policy incentives can further address these objections, for example by allowing safe harbor protections for organizations operating under the rules of a trust network.

Predesigned use cases can provide benchmarks for determining whether given uses of personal data are consistent with measurable criteria. Such criteria can be used to establish compliance with the rules of a trust network and for certification by government for the right to safe harbor or other

protections. Because the New Deal on Data is rooted in common law and the social compact, the appropriate set of rights and expectations covering privacy and other personal data interests can be enumerated, debated, and agreed upon in ways that fit the given use cases.

Conclusions

Society today faces unprecedented challenges and meeting them will require access to personal data, so we can understand how society works, how we move around, what makes us productive, and how everything from ideas to diseases spread. The insights must be actionable and available in real time , thus engaging the population, creating the nervous system of the society. In this chapter we have reviewed how big data collected in institutional contexts can be used for the public good. In many cases, although the data needed to create a better society has already been collected, it sits in the closed silos of companies and governments. We have described how the silos can be opened using well-designed and carefully implemented sets of institutional controls, covering business, legal, and technical dimensions . The framework for doing this – the New Deal on Data – postulates that the primary driver of change must be recognizing that ownership of personal data rests with the people that data is about. This ownership – the right to use, transfer, and remove the data – ensures that the data is available for the public good, while at the same time protecting the privacy of citizens.

The New Deal on Data is still new. We have described here our efforts to understand the technical means of its implementation, the legal framework around it, its business ramifications, and the direct value of the greater access to data that it enables. It is clear that companies must play the major role in implementing the New Deal, incentivized by business opportunities, guided by legislation, and pressured by demands from users. Only with such orchestration will it be possible to modernize the current system of data ownership and put immense quantities and capabilities of collected personal data to good use.

Notes

1. Jonathan Woetzel et al., “Preparing for China’s Urban Billion” (McKinsey Global Institute, March 2009), http:// www.mckinsey.com/ insights/ urbanization/ preparing_for_urban_billion_in_china .

2. David Lazer, Alex Sandy Pentland, Lada Adamic, Sinan Aral, Albert Laszlo Barabasi, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, and Myron Gutmann, “Life in the Network: The Coming Age of Computational Social Science,” Science 323 (2009): 721– 723.

3. Sinan Aral and Dylan Walker, “Identifying Influential And Susceptible Members Of Social Networks,” Science 337 (2012): 337– 341; Alan Mislove, Sune Lehmann, Yong-Yeol Ahn, Jukka-Pekka Onnela, and J. Niels Rosenquist, Pulse of the Nation: U.S. Mood throughout the Day Inferred from Twitter (website), http:// www.ccs.neu.edu/ home/ amislove/ twittermood/ (accessed November 22, 2013); Jessica Vitak, Paul Zube, Andrew Smock, Caleb T. Carr, Nicole Ellison, and Cliff Lampe, “It’s Complicated: Facebook Users ’ Political Participation in the 2008 Election,” Cyberpsychology, Behavior, and Social Networking 14 (2011): 107– 114.

4. Alexis Madrigal, “Dark Social: We Have the Whole History of the Web Wrong,” The Atlantic, October 12, 2013, http:// www.theatlantic.com/ technology/ archive/ 2012/ 10/ dark-social-we-have-the-whole-history-of-the-web-wrong/ 263523/ .

5. Nathan Eagle and Alex Pentland, “ Reality Mining: Sensing Complex Social Systems,” Personal and Ubiquitous Computing 10 (2006): 255– 268; Alex Pentland, “Reality Mining of Mobile Communications: Toward a New Deal on Data,” The Global Information Technology Report 2008– 2009 (Geneva: World Economic Forum, 2009), 75– 80.

6. Alex Pentland, David Lazer, Devon Brewer, and Tracy Heibeck, “Using Reality Mining to Improve Public Health and Medicine,” Studies in Health Technology and Informatics 149 (2009): 93– 102.

7. Vivek K. Singh, Laura Freeman, Bruno Lepri, and Alex Sandy Pentland, “Classifying Spending Behavior using Socio-Mobile Data,” HUMAN 2 (2013): 99– 111.

8. Wei Pan, Yaniv Altshuler, and Alex Sandy Pentland, “Decoding Social Influence and the Wisdom of the Crowd in Financial Trading Network,” in 2012 International Conference on Privacy, Security, Risk and Trust (PASSAT), and 2012 International Conference on Social Computing (SocialCom), 203– 209.

9. Kate Greene, “Reality Mining,” MIT Technology Review, March/ April 2008, http:// pubs.media.mit.edu/ pubs/ papers/ tr10pdfdownload.pdf .

10. Meglena Kuneva, European Consumer Commissioner, “Keynote Speech,” in Roundtable on Online Data Collection, Targeting and Profiling, March 31, 2009, http:// europa.eu/ rapid/ press-release_SPEECH-09-156_en.htm .

11. Kim Gittleson, “How Big Data Is Changing The Cost Of Insurance,” BBC News, November 14, 2013, http:// www.bbc.co.uk/ news/ business-24941415 .

12. Aniko Hannak, Piotr Sapiezynski, Kakhki Arash Molavi, Balachander Krishnamurthy, David Lazer, Alan Mislove, and Christo Wilson, “Measuring Personalization of Web Search,” in Proc. 22nd International Conference on World Wide Web (WWW 2013), 527– 538.

13. Pentland, “Reality Mining of Mobile Communications.”

14. Anmol Madan, Manuel Cebrian, David Lazer, and Alex Pentland, “Social Sensing for Epidemiological Behavior Change,” in Proc. 12th ACM International Conference on Ubiquitous Computing (Ubicomp 2010), 291– 300; Pentland et al. “Using Reality Mining to Improve Public Health and Medicine.”

15. Wei Pan, Gourab Ghoshal, Coco Krumme, Manuel Cebrian, and Alex Pentland, “Urban Characteristics Attributable to Density-Driven Tie Formation,” Nature Communications 4 (2013): article 1961.

16. Lev Grossman, “Iran Protests: Twitter, the Medium of the Movement,” Time Magazine, June 17, 2009; Ellen Barry, “Protests in Moldova Explode, with Help of Twitter,” The New York Times, April 8, 2009.

17. “Directive 95/ 46/ EC of the European Parliament and of the Council of 24 October 1995 on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data,” Official Journal L281 (November 23, 1995): 31– 50.

18. World Economic Forum, “Personal Data: The Emergence of a New Asset Class,” January 2011, http:// www.weforum.org/ reports/ personal-data-emergence-new-asset-class .

19. Ibid.

20. Ibid.

21. Antonio Lima, Manlio De Domenico, Veljko Pejovic, and Mirco Musolesi, “Exploiting Cellular Data for Disease Containment and Information Campaign Strategies in Country-Wide Epidemics,” School of Computer Science Technical Report CSR-13-01, University of Birmingham, May 2013.

22. Arvind Narayanan and Vitaly Shmatikov, “Robust De-Anonymization of Large Sparse Datasets,” in Proc. 2008 IEEE Symposium on Security and Privacy (SP), 111– 125.

23. Latanya Sweeney, “Simple Demographics Often Identify People Uniquely,” Data Privacy Working Paper 3, Carnegie Mellon University, Pittsburgh, 2000.

24. de Montjoye, Yves-Alexandre, Samuel S. Wang, Alex Pentland, “On the Trusted Use of Large-Scale Personal Data,” IEEE Data Engineering Bulletin 35, no. 4 (2012): 5– 8.

25. Chaoming Song, Zehui Qu, Nicholas Blumm, and Albert-Laszlo Barabasi, “Limits of Predictability in Human Mobility,” Science 327 (2010): 1018– 1021.

26. Pentland et al., “Using Reality Mining to Improve Public Health and Medicine.”

27. David Tacconi, Oscar Mayora, Paul Lukowicz, Bert Arnrich, Cornelia Setz, Gerhard Troster, and Christian Haring, “Activity and Emotion Recognition to Support Early Diagnosis of Psychiatric Diseases,” in Proc. 2nd International ICST Conference on Pervasive Computing Technologies for Healthcare, 100– 102.

28. World Economic Forum, “Personal Data.”

29. The White House, “National Strategy for Trusted Identities in Cyberspace: Enhancing Online Choice, Efficiency, Security, and Privacy,” Washington, DC, April 2011, http:// www.whitehouse.gov/ sites/ default/ files/ rss_viewer/ NSTICstrategy_041511. pdf .

30. Thomas Hardjono, “User-Managed Access UMA Profile of OAuth2.0 ,” Internet draft, 2013, http:// docs.kantarainitiative.org/ uma/ draft-uma-core.html.

31. A Creative Commons licensed example set of integrated business and technical system rules for the institutional use of personal data stores is available at https://github.com/HumanDynamics/SystemRules.

32. See http:// openPDS.media.mit.edu for project information and https://github.com/HumanDynamics/openPDS for the open source code.

33. Nick Bilton, “Girls around Me: An App Takes Creepy to a New Level,” The New York Times, Bits (blog), March 30, 2012, http:// bits.blogs.nytimes.com/2012/03/30/girls-around-me-ios-app-takes-creepy-to-a-new-level .

34. U.S. Environmental Protection Agency, RCRA Corrective Action Program, “Institutional Controls Glossary,” Washington, DC, 2007, http://www.epa.gov/epawaste/hazard/correctiveaction/resources/guidance/ics/glossary1.pdf .

35. University of Florida , Center for Environmental & Human Toxicology, “Development of Cleanup Target Levels (CTLs) for Chapter 62-777, F.A.C.,” Technical report, Florida Department of Environmental Protection, Division of Waste Management, February 2005, http:// www.dep.state.fl.us/ waste/ quick_topics/ publications/ wc/ FinalGuidanceDocumentsFlowCharts_April2005/ TechnicalReport2FinalFeb2005 (Final3-28-05). pdf.

36. U.S. Environmental Protection Agency, “Institutional Controls: A Guide to Planning, Implementing, Maintaining, and Enforcing Institutional Controls at Contaminated Sites,” OSWER 9355.0-89, Washington, DC, December 2012, http:// www.epa.gov/ superfund/ policy/ ic/ guide/ Final% 20PIME% 20Guidance% 20December% 202012. pdf .

37. Ralph A . DeMeo and Sarah Meyer Doar, “Restrictive Covenants as Institutional Controls for Remediated Sites: Worth the Effort?” The Florida Bar Journal 85, no. 2 (February 2011); Florida Department of Environmental Protection, Division of Waste Management, “Institutional Controls Procedures Guidance,” Tallahassee, June 2012, http:// www.dep.state.fl.us/ waste/ quick_topics/ publications/ wc/ csf/ icpg.pdf ; University of Florida, “Development of Cleanup Target Levels.”

38. World Economic Forum, “Personal Data.”

39. Thomas Hardjono, Daniel Greenwood, and Alex Pentland, “Towards a Trustworthy Digital Infrastructure for Core Identities and Personal Data Stores,” in Proc. ID360 Conference on Identity, 2013.

40. Hardjono, “User-Managed Access UMA Profile of OAuth2.0”; Eve Maler and Thomas Hardjono, “Binding Obligations on User-Managed Access (UMA) Participants,” Internet draft, 2013, http:// docs.kantarainitiative.org/ uma/ draft-uma-trust.html .

41. Simon Barber, Xavier Boyen, Elaine Shi, and Ersin Uzun, “Bitter to Better – How to Make Bitcoin a Better Currency,” in Proc. Financial Cryptography and Data Security Conference (2012), LNCS 7397, 399– 414.

42. Stan Stalnaker, “About [Ven Currency],” http:// www.ven.vc (accessed January 16, 2014).

43. Thomas Hardjono , Patrick Deegan, and John Clippinger, “On the Design of Trustworthy Compute Frameworks for Self-Organizing Digital Institutions,” in Proc. 16th International Conference on Human-Computer Interaction (2014), forthcoming; Lazer et al., “Life in the Network.”

44. See e.g. the study SensibleDTU (https:// www.sensible.dtu.dk/? lang = en). This study of 1,000 freshman students at the Technical University of Denmark gives students mobile phones in order to study their networks and social behavior during an important change in their lives. It uses not only data collected from the mobile phones (such as location, Bluetooth-based proximity, and call and sms logs), but also from social networks and questionnaires filled out by participants.

-- * -- * ---

Attribution:

This material was been published as Chapter 9 in the book "Privacy, Big Data, and the Public Good" edited by Julia Lane, Victoria Stodden, Stefan Bender, and Helen Nissenbaum, and is reproduced by permission of Cambridge University Press. For more information on this book, please see:

- Cambridge University Press': Books Online: http://dx.doi.org/10.1017/CBO9781107590205 and http://www.cambridge.org/9781107637689 and
- The Amazon.com book page: http://www.amazon.com/Privacy-Big-Data-Public-Good/dp/1107637686

Discuss the New Deal on Data

2 Comments NewDealOnData Dazza Greenwood

Recommend Share Sort by Best

Join the discussion…

Dwight Arthur • 2 years ago

Hi, Dan. This discussion touches on the issues of informed end-user assent, but does not touch on the risk that data will be used in ways outside of the assent that was given. It mentions identifying the party responsible for securing the data but does not consider the consequences when reasonably rigorous security provisions fail. The discussion is centered on the bright side of the framework: how should it work when it is working, and does not give enough attention to the dark side: how should it fail when it is failing. Reference is made to the BLT environment around payment systems, where a considerable component of the structure treats the identification, avoidance, mitigation and recovery of various risks. BLT environments around big data have a far less mature understanding of risk, and yet the risks are far more severe than in payment systems. A failure of a payment system yields financial losses; as long as adequate reserves are maintained the injured parties can all be made whole. When personal information is inappropriately disclosed, there not yet any shared understanding of what a remedy would be. How much reserves would be adequate to compensate the subjects of a data breach? Where security measures were reasonable, in the absence of negligence or malfeasance, what liability applies to the custodian of improperly disclosed data? It should be clear by now that improper disclosure cannot merely be prohibited or avoided, it must be anticipated and the BLT environment should address the consequences.

• Reply • Share ›

Dazza Greenwood Mod • 2 years ago

The Computational Law academic and research team at MIT Media Lab are currently embroiled in a discussion of similarities and differences between Fair Information Practices (FIPS) and the New Deal on Data. This Discussion space is generally about the New Deal on Data, but we'd also love to hear any views specifically on how FIPS and the New Deal on Data relate to each other and to the goal of achieving a stable foundation for personal data and individual privacy?

• Edit • Reply • Share ›

This page is maintained by Dazza Greenwood/a> and published by CIVICS.com