

# 1 **Operational Framework: Institutional Controls - The New Deal** 2 **on Data**

3 Daniel "Dazza" Greenwood<sup>1,\*</sup>, Arkadiusz Stopczynski<sup>1,2</sup>, Brian Sweatt<sup>1</sup>, Thomas Hardjono<sup>1</sup>,  
4 Alex Sandy Pentland<sup>1</sup>

5 **1 MIT**

6 **2 DTU**

7 **\* E-mail: dazza@civics.com**

## 8 **Contents**

9	<b>1 The New Realities of Living in a Big Data Society</b>	<b>1</b>
10	<b>2 The New Deal on Data</b>	<b>4</b>
11	<b>3 Personal Data: Emergence of a New Asset Class</b>	<b>6</b>
12	<b>4 Enforcing the New Deal on Data</b>	<b>10</b>
13	<b>5 Transitioning End-User Assent Practices</b>	<b>13</b>
14	<b>6 Big Data and Personal Data Institutional Controls</b>	<b>15</b>
15	<b>7 Scenarios of Use in Context</b>	<b>19</b>
16	7.1 Example Scenario: Research System for Computational Social Science . . . . .	24
17	7.2 Scenarios of Use Today, Tomorrow, and the Day After . . . . .	26
18	<b>8 Conclusions</b>	<b>28</b>

## 19 **1 The New Realities of Living in a Big Data Society**

20 To realize the promise and prospects of a Big Data society and avoid its security and confiden-  
21 tiality perils, institutions are updating operational frameworks governing business, legal, and

22 technical dimensions of their internal organization and interactions with the outside world. In  
23 this chapter we explore the emergence of the Big Data society, outline ways to support it in the  
24 context of institutional controls within the framework of the New Deal on Data, and describe  
25 future directions for research and development.

26 The control points traditionally relied upon as part of corporate governance, management  
27 oversight, legal compliance, and enterprise architecture must evolve and expand to match oper-  
28 ational frameworks for Big Data. An operational framework used for a Big Data driven organi-  
29 zation requires a balanced set of institutional controls. These controls must support and reflect  
30 greater user control over personal data, as well as large scale interoperability for data sharing be-  
31 tween and among institutions. Core capabilities of these controls include responsive rule-based  
32 systems governance and fine-grained authorizations for distributed rights management.

33 Sustaining a healthy, safe, and efficient society is a scientific and engineering challenge dating  
34 back to the 1800s when the Industrial Revolution spurred rapid urban growth, thereby creating  
35 huge social and environmental problems. The remedy then was to build centralized networks  
36 that delivered clean water and safe food, enabled commerce, removed waste, provided energy,  
37 facilitated transportation, and offered access to centralized health care, police, and educational  
38 services. These networks formed the backbone of society as we know it today.

39 These century-old solutions are, however, becoming increasingly obsolete and inefficient. We  
40 have cities jammed with traffic, world-wide outbreaks of disease that are seemingly unstoppable,  
41 and political institutions that are deadlocked and unable to act. We face the challenges of global  
42 warming, uncertain energy, water, and food supplies, and a rising population and urbanization  
43 that will add 350 million people to the urban population by 2025 in China alone [15].

44 It does not have to be this way. We can have cities that are energy efficient, have secure food  
45 and water supplies, are protected from pandemics, and enjoy much better governance. To reach  
46 these goals, however, we need to radically rethink our approach. Rather than static fixed systems  
47 separated by function — water, food, waste, transport, education, energy — we must consider  
48 them as dynamic, data-driven networks. Instead of focusing only on access and distribution, we

49 need networked and self-regulating systems, driven by the needs and preferences of the citizens.

50     Sustainable, future societies depend on our new technologies being used to create a *nervous*  
51 *system* maintaining the stability of government, energy, and public health systems around the  
52 globe. The digital feedback technologies of today are capable of creating a level of dynamic  
53 responsiveness required by our larger, more complicated, modern society. We must reinvent  
54 the systems of societies within a control framework: sensing the situation, combining these  
55 observations with models of demand and dynamic reaction, using the resulting predictions to  
56 tune the system to match the demands.

57     The engine driving this nervous system is Big Data: the newly ubiquitous digital data, now  
58 available about all aspects of human life. We can analyze patterns of human experience and  
59 idea exchange within the *digital breadcrumbs* we all leave behind as we move through the world:  
60 call records, credit card transactions, GPS location fixes, among others [25]. By recording our  
61 choices, these data tell the story of our lives. This may be very different from what we decide  
62 to put on Facebook or Twitter; our postings there are what we choose to tell people, edited  
63 according to the standards of the day and filtered to match the persona we are building. Mining  
64 social networks can give some great insights about human nature [4, 29, 44]; who we really are,  
65 however, is even more accurately determined by where we spend our time and which things we  
66 buy, rather than just what we say we do [28].

67     The process of analyzing the patterns within these digital breadcrumbs is called reality  
68 mining [14, 33], and through it we can learn an enormous amount about who we are. The  
69 Human Dynamics research group at MIT found that we can use them to tell if we are likely  
70 to get diabetes [34], or whether we are the sort of person who will pay back loans [36]. By  
71 analyzing these patterns across many people, we are discovering that we can begin to explain  
72 many things — crashes, revolutions, bubbles — that previously appeared to be random acts of  
73 God [31]. For this reason, the magazine Technology Review named our development of reality  
74 mining as one of the ten technologies that will change the world [18].

## 2 The New Deal on Data

The digital breadcrumbs we leave behind provide clues about 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” [24]. This new ability to see the details of every interaction can be used for good or for ill. Therefore, maintaining protection of personal privacy and freedom is critical to our future success as a society. We need to enable even more data sharing for the public good; at the same time, we need to do a much better job in protecting the privacy of the 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-grain data. The abuses may be directly targeted at users, for example, by offering them higher insurance rates based on their shopping history [17], or create problems for the entire society, such as limiting user choices and closing them into information bubbles [20]. To achieve the positive possibilities of a new society, we require the *New Deal on Data*, workable guarantees that the data needed for public good are readily available while at the same time protecting the citizenry [33].

The key insight motivating the idea of the New Deal on Data is that our data are worth more when shared, because these aggregated data — averaged, combined across population, and often distilled to high-level features — 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 [27,34]. Our research has reported how we were able to use these digital breadcrumbs to track the spread of influenza from person to person on an individual level. And if we can see it, we can also stop it.

Similarly, if we are worried about global warming, these shared, aggregated data can show us how patterns of mobility relate to productivity [32]. In turn, this provides us with the ability to design cities that are more productive and, at the same time, more energy efficient. However,

102 in order to obtain these results and make a greener world, we need to be able to see the people  
103 moving around; this depends on having many people willing to contribute their data, even if  
104 only anonymously and in aggregate.

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

111 Unfortunately, today most personal data are siloed off in private companies and therefore  
112 are largely unavailable. Private organizations collect the vast majority of the personal data in  
113 the form of mobility patterns, financial transactions, and phone and Internet communications.  
114 These data must not remain the exclusive domain of private companies, because then they are  
115 less likely to contribute to the common good. Thus, these private organizations must be key  
116 players in the New Deal on Data framework for privacy and data control. Likewise, these data  
117 should not become the exclusive domain of the government, as this will not serve the public  
118 interest of transparency; we should be suspicious of trusting the government with such power.  
119 The entities who should be empowered to share and make decisions about their data are the  
120 people themselves: users, participants, citizens.

121 Through the years, the great goal of human societies was to find the efficient ways of gov-  
122 ernance. The Big Data transformation can contribute to this ultimate goal of providing the  
123 society with tools to analyze and understand what needs to be done, and to reach the consensus  
124 on how to do it. This goes beyond simple creation of more communication platforms; the as-  
125 sumption that more interactions between users will result in better decisions being made, may  
126 be very misleading. Although in the recent years we have seen some great examples of using  
127 social networks for better organization in society, for example during political protests [6, 19], we  
128 are not even close to the point where we can start reaching consensus about the big problems:

129 epidemics, climate change, pollution. We can improve the discussions by making them data  
 130 driven, involving both experts and wisdom of the crowds – users themselves interested in im-  
 131 proving the society. The problems we are dealing with as a now global society are more difficult  
 132 than ever. We are responsible for many of them, and being able to tackle them on a global scale  
 133 is necessary for our survival as a people.

### 134 **3 Personal Data: Emergence of a New Asset Class**

135 It has long been recognized that the first step to promoting liquidity in land and commodity  
 136 markets is to guarantee ownership rights so that people can safely buy and sell. Similarly, the  
 137 first step toward creating more new ideas and greater flow ideas — idea liquidity — is to define  
 138 ownership rights. The only politically viable course is to give individual citizens key rights over  
 139 data that are about them and in fact, these types of rights have undergirded the European  
 140 Union’s Privacy Directive since 1995 [13].

141 We need to recognize personal data as a valuable asset of the individual, which is given to  
 142 companies and government in return for services. The simplest approach to defining what it  
 143 means to own your own data is to draw an analogy with English common law on ownership  
 144 rights of possession, use, and disposal:

- 145 • You have the right to possess data about yourself. Regardless of what entity collects the  
 146 data, the data belong to you, and you can access your data at any time. Data collectors  
 147 thusly play a role akin to a bank, managing the data on behalf of their customers.
- 148 • You have the right to full control over the use of your data. The terms of use must be opt-  
 149 in and clearly explained in plain language. If you are not happy with the way a company  
 150 uses your data, you can remove the data, just as you would close your account with a bank  
 151 that is not providing satisfactory service.
- 152 • You have the right to dispose of or distribute your data. You have the option to have data  
 153 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 so on-to run their day-to-day operations. This 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.

Note that these ownership rights are not exactly the same as literal ownership under modern law, but 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 (Pentland) first proposed the New Deal on Data to the World Economic Forum [45]. Since then, this idea has run through various discussions and eventually helped shape the 2012 Consumer Data Bill of Rights in the United States, along with a matching declaration on Personal Data Rights in the EU. These new regulations hope to accomplish the combined trick of breaking data out of the current silos, thus enabling the public good, while at the same time giving individuals greater control over data about them. Of course, this is still a work in progress and the battle for individual control of personal data rages onward.

The World Economic Forum (WEF) has dubbed personal data as the “New Oil” or resource of the 21st century [45]. The discovery of oil and the subsequent development of the oil industry over the past 100 years has spurred not only the development of the automobile industry but, also the creation of the global transportation infrastructure, including the massive freeway networks we see today in the developed nations. The “personal data sector” of the economy today is still in its infancy, its state akin to the oil industry during the late 1890s, prior to the development of the Model-T Ford automobile. The productive collaboration between the Government (building the state owned freeways), the private sector (mining and refining oil, building automobiles), and the citizen (the user-base of these services) allowed the 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, there needs to be a productive collaboration between all the stakeholders in the establishment of a *personal data*

181 *ecosystem*. As mentioned in [45], a number of fundamental questions about privacy, property,  
 182 global governance, human rights — essentially around who should benefit from the products  
 183 and services built upon personal data — are major uncertainties shaping the opportunity. The  
 184 rapid rate of technological change and commercialization in using personal data is undermining  
 185 end user confidence and trust.

186 The current personal data ecosystem is fragmented and inefficient. Too much leverage is  
 187 currently being accorded to service providers that enroll and register end-users. These siloed  
 188 repositories of personal data exemplify the fragmentation of the ecosystem. These repositories  
 189 contain data of varying qualities. Some are attributes of persons that are unverified, while  
 190 other represent higher quality data that have been cross-correlated with other data points of the  
 191 end-user.

192 For many participants, the risks and liabilities exceed the economic returns. Besides not  
 193 having the infrastructure and tools to manage personal data, many end-users simply do not see  
 194 the benefit of fully participating in the ecosystem. The current focus of many Internet-based  
 195 service providers is to capture as much personal data from the end-user and to sell this data  
 196 into the advertising industry. Personal privacy concerns are thus inadequately addressed at  
 197 best, or simply overlooked in the majority of cases. The current technologies and laws fall short  
 198 of providing the legal and technical infrastructure needed to support a well-functioning digital  
 199 economy.

200 Recently, we have shown how challenging, but also feasible, it is to open such institu-  
 201 tional Big Data. In the Data For Development (D4D) Challenge <http://www.d4d.orange.com>,  
 202 the telecommunication operator Orange opened access to a large dataset of call detail records  
 203 (CDRs) from the Ivory Coast. Working with the data as part of a challenge, teams of researchers  
 204 came up with life-changing insights for the country. For example, one team developed a model  
 205 for how disease spread in the country and demonstrated that information campaigns based on  
 206 one-to-one phone conversations among members of social groups can be an effective counter-  
 207 measure [26]. In releasing and analyzing this data, the privacy of the people who generated



the data was protected not only by technical means, such as removal of Personally Identifiable Information (PIIs), but also by legal means, with the researchers signing an agreement they will not use the data for re-identification or other nefarious purposes. As we have seen in several cases, such as the Netflix Prize privacy disaster [30] and other similar privacy breaches [39], true anonymization is extremely hard. In the Unique in the Crowd [10], de Montjoye et al. showed that even though human beings are highly predictable [37], we are also very unique. Having access to one dataset may be enough to uniquely fingerprint someone based on just a few datapoints, and use this fingerprint to discover their true identity.

The report of the World Economic Forum [45] also suggest a way forward by recommending a number of areas where efforts could be directed:

- Alignment of key stakeholders: Citizens, the private sector and the public sector need to work in support of one another. Efforts such as NSTIC [40] — albeit still in its infancy — represent a promising direction for a global collaboration.
- Viewing “data as money”: There needs to be a new change in mindset where 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 it would be controlled, managed, exchanged and accounted for just like personal banking services operate today.
- End-user centricity: All entities in the ecosystem need to recognize that end-users are vital and independent stakeholders in the co-creation and value exchange of services and experiences. Efforts such as the *User Managed Access* (UMA) initiative [2] point in the right direction by designing systems that are user-centric and managed by the user.

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 substantial things when the 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 [34], underperforming students can be helped, and people with health risks can be treated before they get sick [9].

## 234 4 Enforcing the New Deal on Data

235 How can we enforce this New Deal? The threat of legal action alone is important, but insufficient,  
236 because if you cannot see abuses then you cannot prosecute them. Moreover, who wants more  
237 lawsuits anyway? Enforcement can be addressed in significant ways without prosecution of public  
238 statute or regulation at all. In many fields, companies and governments rely upon multi-party  
239 frameworks of agreed upon rules governing common business, legal, and technical practices to  
240 create effective self-organization and enforcement. These approaches hold promise as a method  
241 for using institutional controls to form a reliable operational framework balancing the needs for  
242 Big Data, privacy, and access.

243 One current best practice is a system of data sharing called trust networks. Trust networks  
244 are a combination of networked computers and legal rules defining and governing expectations  
245 regarding data. With respect to data belonging to individuals, these networks of technical and  
246 legal rules keeps track of user permissions for each piece of personal data, and a legal contract  
247 that specifies both what you can and cannot do with the data and what happens if there is a  
248 violation of the permissions. For example, in such a system all personal data can have attached  
249 labels specifying what the data can and cannot be used for. These labels are exactly matched  
250 by the network's system rules and terms in legal contracts between all the participants, stating  
251 penalties for not obeying the permission labels. These rules can, and often do, reference or  
252 require audits of relevant systems and data use, demonstrating how traditional internal controls  
253 can be leveraged as part of the transition to more novel trust models.

254 Complete tracking and regulation of every aspect of a trust network is not the goal or  
255 even desirable in order to achieve effective enforcement. Rather, the rules for a trust network  
256 align enforcement with the highest priority issues and those upon which trust of participants is  
257 premised. The relevant issues for a given trust network arise from that systems underlying trust  
258 models and the contextual scenarios within which the networked data and the relationships of  
259 parties occur.

260 When a trust network involves use of personal data, then the user permissions and corre-

261 sponding limits on use are fundamental to the trust model. In this context, the permissions,  
262 including the provenance of the data, should require appropriate levels of audit. A well designed  
263 trust network, elegantly integrating computer and legal rules, allows automatic auditing of data  
264 use and allows individuals to change their permissions and withdraw data.

265 Having system rules applicable to the networks, applications, and data as well as all the ser-  
266 vices providers, other intermediaries, and the users themselves is the mechanism for establishing  
267 and operating a trust network. System rules are sometimes called operating regulations in the  
268 credit card context or known as trust frameworks in the identity federations context or trading  
269 partner agreements in a supply value chain context. There are many general examples of multi-  
270 party shared architectural and contractual rules that share the generic characteristic of creating  
271 binding obligations and enforceable expectations on all participants in scalable networks. An-  
272 other common characteristic of the system rules design pattern is that the participants in the  
273 network can be widely distributed across very heterogeneous business ownership boundaries,  
274 legal governance structures, and technical security domains. Yet, the parties need not agree  
275 to conform to all or most aspects of their basic roles, relationships, and activities in order to  
276 connect to systems of a trust network. Cross-domain trusted systems must, by their nature,  
277 focus mandatory and enforceable rules narrowly upon the critical items that must be commonly  
278 agreed in order for that network to achieve its purpose.

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

288       A system like this has made the interbank money transfer system among the safest systems  
289 in the world and the daily backbone for exchanges of trillions of dollars, but until recently such  
290 systems were only for the ‘big guys’. To give individuals a similarly safe method of managing  
291 personal data, the Human Dynamics research group at MIT, in partnership with the Insti-  
292 tute for Data Driven Design, co-founded by John Clippinger and one author (Pentland), have  
293 helped build open Personal Data Store (openPDS) [11]. See <http://openPDS.media.mit.edu>  
294 for project information and <https://github.com/HumanDynamics/openPDS> for the open source  
295 code.

296       The openPDS is a consumer version of a personal cloud trust network that we are now  
297 testing with a variety of industry and government partners. Soon, sharing your personal data  
298 could become as safe and secure as transferring money between banks.

299       The Human Dynamics Lab has applied the system rules approach to development of in-  
300 tegrated business, technical architecture, and rules large scale institutional use of personal  
301 data stores, available as an example under MIT’s creative commons license by MIT, at [https:](https://github.com/HumanDynamics/SystemRules)  
302 [//github.com/HumanDynamics/SystemRules](https://github.com/HumanDynamics/SystemRules).

303       When it comes to data intended to be accessible over networks — whether big, personal, or  
304 otherwise — the traditional container of an institution makes less and less sense. Institutional  
305 controls apply, by definition by or to some type of institutional entity such as a business, gov-  
306 ernmental, or religious organization. A combined view of the business, legal, and technical facts  
307 and circumstances surrounding Big Data is necessary to know what access, confidentiality, and  
308 other expectations exist. The relevant contextual aspects of Big Data of one institution is often  
309 profoundly different from that of another. As more and more organizations use and rely upon  
310 Big Data, a single formula for institutional controls will not work for increasingly heterogeneous  
311 business, legal, and technical environments in play. Many organizations are structured with clear  
312 leadership on business, legal, and technical issues functionally assigned to top level executive  
313 roles. Business issues are typically allocated to roles such as CEO, COO, or CFO, while leader-  
314 ship on legal issues is commonly assigned to roles like general counsel and regulatory compliance

315 and technical leads are often the roles of CIO, CTO, or CSO. Having top level leadership for  
 316 each of the business, legal, and technical aspects of a trust network is a critical success factor.

317 The capacity to apply the appropriate methods of enforcement for a trust network depend  
 318 upon a clear understanding and agreement among parties about the purpose of the trusted  
 319 system and the respective roles or expectations of those connecting as participants. Therefore,  
 320 an anchor is needed to a clear context of a Big Data operational framework and institutional  
 321 controls appropriate for access and confidentiality or privacy. The following section posits the  
 322 trust model and signature traits of such a context, through the lens of the New Deal on Data.

## 323 5 Transitioning End-User Assent Practices

324 The way users grant authorizations to their data is not a trivial matter. The flow of personal  
 325 information, such as location data, purchases and health records can be very complex. Every  
 326 tweet, geo-tagged picture, phone call, or purchase with credit card, provide the user's location  
 327 not only to the primary service, but also to all the applications and services that have been  
 328 authorized to access and reuse these data. The authorizations may come from the end-user  
 329 or be granted by the collecting service, based on an umbrella terms of service, allowing the  
 330 re-use of the data. Implementation of such flows was a crucial part of the Web 2.0 revolution,  
 331 realized with RESTful APIs, mashups, and authorization-based access. The way the personal  
 332 data travel between the services has however become arguably too complex for a user to handle  
 333 and manage.

334 Increasing the amount of data controlled by the user and granularity of this control is mean-  
 335 ingless if it cannot be exercised in an informed way. For many years, the End User License  
 336 Agreements (EULAs), long incomprehensible texts have been accepted blindly by the user,  
 337 trusting they have not agreed to anything that could harm them. The process of granting the  
 338 authorizations cannot be too complex, as it would prevent the user from understanding her deci-  
 339 sions. At the same time, it cannot be too simplistic, as it may not sufficiently convey the weight  
 340 of the privacy-related decisions. It is a challenge in itself, to build the end-user assent systems

341 that allow the user to understand and adjust their privacy settings. Complex EULAs do not  
342 promote the privacy of the users, effectively pushing them to press *I Agree* in every presented  
343 window.

344 This gap between the interface — single click — and the effect, can render the data owner-  
345 ship meaningless; the click may wrench people and their data into systems and rules that are  
346 antithetical to fair information practices, such as is prevalent with today’s end-user licenses in  
347 cloud services or applications. Managing the potentially long term and opposite dynamics fueled  
348 by old deal systems operating simultaneously with the new deal systems is an important design  
349 and migration challenge during the transition to a Big Data economy. During this transition  
350 and after the New Deal on Data is no longer new, personal data must continue to flow in order  
351 to be useful. Protecting the data of people outside of the user-controlled domain is very hard  
352 without a combination of cost effective and useful business practices, legal rules, and technical  
353 solutions.

354 We envision Living Informed Consent, where the user is entitled to know what data is being  
355 collected about her by which entities, empowered to understand the implications of data sharing,  
356 and finally put in charge of the sharing authorizations. We suggest the readers ask themselves a  
357 question: *Which services know which city I am in today?*. Google? Apple? Twitter? Amazon?  
358 Facebook? Flickr? This small application we have authorized a few years ago to access our  
359 Facebook check-ins and forgot since then? This is an example of a fundamental question related  
360 to user privacy and assent, and yet finding the answer to it may be surprisingly difficult in today’s  
361 ecosystem. We can hope that most of the services treat the data responsibly and according to  
362 user authorizations. In the complex network of data flows however, it is relatively easy for the  
363 data to leak to careless or malicious services [7]. We need to build the solutions to help the user  
364 to make well informed decisions about data sharing.

## 365 6 Big Data and Personal Data Institutional Controls

366 The concept of “institutional controls” refers to safeguards and protections by use of legal, policy,  
 367 governance, and other non-strictly technical, engineering, or mechanical measures. The phrase  
 368 institutional controls in a Big Data context can perhaps best be understood by examining how  
 369 the concept has been applied to other domains. The most prevalent use of institutional controls  
 370 has been in the field of environmental regulatory frameworks.

371 A good example of how this concept supports and reflects the goals and objectives of en-  
 372 vironmental regulation can be found in the policy documents of the Environmental Protection  
 373 Agency (EPA). This following definition is instructive, and is part of the Institutional Control  
 374 Glossary of Terms [42]:

375 *Institutional Controls - Non-engineering measures intended to affect human activ-*  
 376 *ities in such a way as to prevent or reduce exposure to hazardous substances. They*  
 377 *are almost always used in conjunction with, or as a supplement to, other measures*  
 378 *such as waste treatment or containment. There are four categories of institutional*  
 379 *controls: governmental controls; proprietary controls; enforcement tools; and infor-*  
 380 *mational devices.*

381 Going deeper, the article by DeMeo and Doar [12] defines institutional controls thusly:

382 *Institutional controls are administrative and legal controls that help minimize the*  
 383 *potential for human exposure to contamination and/or protect the integrity of the*  
 384 *physical remedy. They can include recorded restrictive covenants, but land use laws*  
 385 *and regulations, deed restrictions, department consent orders, and conservation ease-*  
 386 *ments are all institutional controls.*

387 In domains of information technology, this approach is most commonly reflected as “enter-  
 388 prise controls” related to security. See, for example, the Juniper Networks enterprise security  
 389 report [23] stating: “Enterprise mobility technologies, especially those designed to retrofit en-  
 390 terprise controls on top of consumer mobile devices, are rapidly evolving. This was a message

391 we heard loud and clear in the study.” This study and analysis also reveals much about the  
392 internal controls needed to accommodate mobile device use by employees. In both capacities as  
393 employee, consumer, and other roles, the use of mobile devices triggers myriad legal, policy, and  
394 other implications for institutional controls.

395 In the legal domain, this concept frequently emerges under the moniker “regulatory compli-  
396 ance” or “legal compliance” anchored in legal and regulatory frameworks such as Health Insur-  
397 ance Portability and Accountability Act (HIPAA) and Sarbanes-Oxley (SOX). These statutory  
398 legal frameworks require covered organizations to establish integrated sets of governance, legal,  
399 transactional, security, and other internal controls to avoid violating the rules. The institutional  
400 controls are accomplished in tight integration with engineering and other measures in order  
401 to ensure compliance and to control legal and security risk. The use of institutional controls  
402 of this type are fundamental methods for achieving and maintaining the transition to a dig-  
403 ital, networked, and Big Data footing for any private company, government agency, or other  
404 organization.

405 Consider again the analogy of institutional controls in the context of environmental law, and  
406 how these types of measures can be applied in the Big Data, privacy, and access context to  
407 digital environments. Given the relatively mature and stable state of environmental regulation,  
408 there is much to be learned by examining this context of institutional controls. Environmental  
409 regulatory compliance with waste management cleanup requirements could include institutional  
410 controls restricting land use on adjacent property. In these situations, it is possible that the  
411 remediation strategy requires significant use of land outside the property boundaries of the  
412 cleanup site. In these cases, the regulators and the land owner responsible for the regulated  
413 property must find ways to ensure a common approach among multiple owners and across  
414 multiple property environments. Clauses on the relevant deeds, an enforceable consent order,  
415 or targeted regulations and zoning rules are examples of more severe institutional controls that  
416 can be employed to ensure consistent and effective actions are taken across ownership and real  
417 property boundaries.



418 See, for example, Florida Department of Environmental Protection (FDEP), Division of  
 419 Waste Management [16] which states that “...RMO III does contemplate contamination beyond  
 420 the Property boundaries, which would require agreement by the adjacent owners to put an RC  
 421 on their properties as well.”

422 The concept of an “institutional control boundary” is especially clarifying and powerful when  
 423 applied to the networked and digital boundaries of an institution. In the context of Florida’s  
 424 environmental regulation frameworks, the phrase is applied to describe the various types of  
 425 combinations risk management levels related to target cleanup standards and extend beyond  
 426 the area of a physical property boundary. Also see a recent University of Florida report on  
 427 Development of Cleanup Target Levels (CTLs) [8] stating “Risk Management Options Level  
 428 III, like Level II, allows concentrations above the default groundwater CTLs to remain on site.  
 429 However, in some rare situations, the institutional control boundary at which default CTLs must  
 430 be met can extend beyond the site property boundary.”

431 The EPA provides considerable information on the nature and use of institutional controls,  
 432 including situations when the situational scope extends to adjacent properties owned by third  
 433 parties. See, generally, *EPA Hazardous Waste Corrective Action Guidance on Institutional Con-*  
 434 *trols* [42]. Also see: *Institutional Controls Bibliography: Institutional Control, Remedy Selection,*  
 435 *and Post-Construction Completion Guidance and Policy, December 2005* [41].

436 When institutional controls would apply to “separately owned neighboring properties” a  
 437 number of issues arise that are very relevant to the problems associated with managing personal  
 438 and big data across legal, business and other systemic boundaries. Requiring the party respon-  
 439 sible for site cleanup to use “best efforts” to attain agreement by third parties to institute the  
 440 relevant institutional controls is perhaps the most direct and least prescriptive approach. When  
 441 direct negotiated agreement is not successful, then use of third party neutrals to resolve disagree-  
 442 ments regarding institutional controls can be required. If necessary, environmental regulation  
 443 can force an acquisition of neighboring land by compelling the party responsible to purchase the  
 444 other property or by purchase of the property directly by the EPA [43].

445 In the context of Big Data, privacy, and access, institutional controls are seldom, if ever,  
 446 the result of government regulatory frameworks such as are seen in the environmental waste  
 447 management oversight by the EPA. Rather, institutions applying measures constituting institu-  
 448 tional controls in the Big Data and related information technology and enterprise architecture  
 449 contexts will typically employ governance safeguards, business practices, legal contracts, tech-  
 450 nical security, reporting, and audit programs and various risk management measures.

451 Inevitably, institutional controls for Big Data will have to operate effectively across institu-  
 452 tional boundaries, just as environmental waste management internal controls must sometimes  
 453 be applied across real property boundaries and may subject multiple different owners to enforce-  
 454 ment actions corresponding to the applicable controls. Short of government regulation, the use  
 455 of system rules as a general model are one widely understood, accepted, and efficient method  
 456 for defining, agreeing, and enforcing institutional and other controls across business, legal, and  
 457 technical domains of ownership, governance, and operation.

458 The use of system rules and integrated participation agreements by developers and end-  
 459 users is a way to ensure intended operational frameworks conform to applicable institutional  
 460 controls. The example of Living Informed Consent described in this chapter, demonstrates how  
 461 institutional controls comprised of legal and definite workflow measures, in concert with technical  
 462 methods, can result in a higher level of performance, while appropriately balancing legitimate  
 463 interests of various parties regarding use and access to personal data.

464 Following the World Economic Forum recommendations of treating personal data stores in  
 465 the manner of bank accounts [45], there are a number of infrastructure improvements that need to  
 466 be realized, if the personal data ecosystem is to flourish and deliver new economic opportunities.  
 467 We believe the following infrastructure improvements are necessary for the coming personal data  
 468 ecosystem:

- 469 • *New global data provenance network*: In order for personal data to be treated like bank  
 470 accounts, the origin information regarding data items coming into the data store must be  
 471 maintained [22]. In other words, the provenance of all data items must be accounted for

by the IT infrastructure upon which the personal data store operates. The heterogeneous provenance databases must then be interconnected in order to provide a resilient and scalable platform for audit and accounting systems to track and reconcile the movement of personal data from the respective data stores.

- *Trust network for computational law:* In order for trust to be established between parties who wish to exchange personal data, we foresee that some degree of “computational law” technologies may have to be integrated into the design of personal data systems. Such technologies 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 [1, 2] are beginning to bring better evidentiary proof and enforceability of contracts into the technical protocol flows.
- *Development of institutional controls for digital institutions:* Currently there are a number of proposals for the creation of virtual currencies (e.g. BitCoin [5], Ven [38]) in which the systems have the potential to evolve into self-governing “digital institutions” [21]. Such systems and institutions that operate on them will necessitate the development of a new paradigm to understand the aspects of institutional control within their context.

## 7 Scenarios of Use in Context

Development of 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 Big Data exists. Although Big Data straddles multiple business, legal, and technical boundaries it will nonetheless have one or more institutions that are capable of, or in some situations required to, manage and control it. The public good referred to in the title of this book can be articulated through the use of system, service and software modeling, requirements setting, development, testing, and certification processes. Discrete use cases of actors and actions is one approach to model business, legal, and technical requirements in a way that can objectively be

497 agreed in advance and tested against implemented systems and components. However, those  
498 are typically atomic or very granular and operate deep within layers of assumed context. Higher  
499 level contexts and corresponding scenarios of multiple use cases can describe fundamental ex-  
500 pectations about matters like interests in property, rights to liberty, and honoring the social  
501 compact.

502       Consider that the applicable scenario within which the data exists can provide a method and  
503 mechanisms of sorts to establish the basic ownership, control, and other expectations of the key  
504 parties. For example, it may not be sufficient to describe the exchange of money and financial  
505 information because the nature of the transaction and their respective data and systems are not  
506 identified enough to predict the rights and obligations or other outcomes reasonably expected  
507 by individuals and organizations that engage in the activity of a financial exchange. The sale of  
508 used cars via an app, the conduct of a counseling session via Google Hangout, and the earning  
509 of a masters degree via an online university all represent scenarios wherein the use case of  
510 a financial exchange takes place. However, each of these scenarios occurs in contexts that are  
511 easily identifiable, involving the sale of goods and deeper access to financial information if the car  
512 is financed, or involving the practice of therapy by a licensed professional involving confidential  
513 mental health data or involving elearning services and protected educational records and possibly  
514 deeper financial information if the program is funded by scholarship or loans. Identifying the  
515 people (a consumer and a used car dealer) the transaction (purchase of a used car) the data  
516 (sales and title data, finance information, etc) and the systems (the third party app and it's  
517 relevant services or functions, state DMV services, credit card and bank services, etc) provide  
518 enough context to establish generally what existing consumer rights under the relevant state  
519 lemon laws, the Uniform Commercial Code and other applicable rules will govern when duties  
520 arise or are terminated, what must be promised, what can be repudiated, by whom data must  
521 be kept secure and other requirements or constraints on the use of personal data and Big Data.  
522 These and other factors vary when a transaction that is otherwise identical seeming operates  
523 within different scenarios, and even scenarios will differ depending upon which contexts apply.

524 The basic common law inspired ownership tenants of the New Deal on Data are general  
525 principles that guide and inform basic relationships and expectations. However, the dynamic  
526 bundle of recombinant rights and responsibilities constituting "ownership" interests in personal  
527 data and expectations pertaining to Big Data vary significantly from context to context and  
528 even from one scenario to another within a given general context. Institutional controls and  
529 other system requirements or safeguards are important methods to ensure context-appropriate  
530 outcomes consistent with clearly applicable system scenarios that set the contours and under-  
531 pinnings for a greater public good. The New Deal on Data can be achieved in part by sets of  
532 institutional controls involving governance, business, legal, and technical aspects of Big Data  
533 and interoperating systems. Reference to relevant scenarios reveal signature features of the New  
534 Deal on Data in various contexts and can serve as an anchor to evaluate what institutional  
535 controls are well aligned to achieve a balance of economic, privacy and other interests.

536 The types of requirements and rules governing participation by individuals and organizations  
537 in Trust Networks vary depending on the facts and circumstances related to the transactions,  
538 data types, relevant roles of people and other factors. Antecedent but relevant networks such  
539 as credit card systems, trading partner systems and exchange networks are instructive not only  
540 for their many common elements but also as important examples of how vastly different they  
541 are from one another depending upon contexts, scenarios, legal obligations, business models,  
542 technical processes and other signature patterns. Trust Networks that are formed to help manage  
543 Big Data in ways that appropriately respect personal data rights and other broader interests  
544 similarly will succeed to the extent they can tolerate or promote a wide degree of heterogeneity  
545 among participants for those business, legal and technical matters that need not be uniform  
546 or directly harmonized. In some situations, new business models and contexts will emerge that  
547 require fresh thinking and novel combinations of roles or types of relationships among transacting  
548 parties. In these cases, understanding the actual context and scenarios will serve as a critical  
549 anchor for establishment of acceptable and sustainable business, legal and technical rules and  
550 systems.

Which scenarios are relevant and what lower level use cases apply are knowable in detail only with reference to the relevant context of a factually based situation. Relevant scenario of use are comprised of people conducting transactions through systems in which personal data and Big Data exists or flows. It is possible to test whether frameworks for engagement successfully address Big Data, privacy and the public good by testing outcomes of relevant scenarios. Scenarios are capable of adequately defining these high level goals and objectives when they identify each of the following four elements:

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 data sets (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 is used by the people, for the transactions or with the data)?

Retail marketing is a common context within which personal data is important. Personal data is critical to many different scenarios in the context of retail marketing. Consider the scenario whereby a merchant conducts an online promotion for an app or service by using a purchased direct marketing database of consumers who have expressed interest in similar products. Data such as the names, email addresses, phone numbers and other personal information can be used to lower costs and increase revenue by better targeting promotional messages and increasing sales. However, there are risks to the merchant and consumer alike, including the potential of a data breach and resulting identity theft and fraud. There is also risk that some consumers will feel annoyed or violated when their personal information is used in this manner without their prior knowledge or consent. The information available from such third party

576 marketing lists and databases may be out of data and lead to the wast of marketing dollars and  
577 the failure to inform potentially interested consumers of a product they might have purchased if  
578 the solicitation had gone to their current email or appropriate network. Imagine that the same  
579 consumers had individual personal data stores and were able to "intent-cast" their interest in  
580 the product. This can be done without revealing all the other personal data of that person. The  
581 The openPDS system could be configured to provide permission based answers to questions such  
582 as whether the consumer is over the age of 18 or lives in a city, suburb or rural area. Sectors  
583 such as real estate could be transformed by such intent-casting by qualified buyers.

584 Another common context involving personal data is governmental transactions with the  
585 public. Government filings, registrations, permits and other such public sector transactions with  
586 the individuals or organizations create a large volume and variety of personal data flow. Consider  
587 the scenario whereby a person runs a small business and must comply with tax, employee  
588 related, licensing and other rules by filing forms with multiple government agencies at the federal,  
589 state and local levels. Individuals names, addresses, occupations, dates of birth, social security  
590 numbers and many other types of personal information are common elements of such filings.  
591 Similarly to the retail marketing scenario above, the parties to government filing transactions  
592 also risk unauthorized access to the personal data by interception during transmission or by  
593 breach of data storage systems. In addition, the costs associated with requiring the same data  
594 by many different agencies and updating or correcting data are born by both the filer and the  
595 regulator. What if the people who own or operate such businesses had access to the services  
596 and functions of a personal data store for themselves individually and also for the corporate  
597 entity they operated? Routine changes in status, such as a change of address or name, could  
598 be accomplished in a secure manner once via their own data service and leveraged again and  
599 again by the many faces of government requiring that data. When the authoritative source  
600 of such information can be deemed to be housed within or logically connected to a person's  
601 data store, then the laborious task of address verification and tedious forms and other processes  
602 required by each government entity could be avoided. The saving of direct and indirect costs,

the regaining of time spent by each agency and business and avoidance of delays and uncertainty are of significant value to all parties (See: <http://kansasbusinesscenter.com> and see the data files at <https://github.com/kansasbusinesscenter>)

The scenario below describes deeper fact-based situations and circumstances in the context of social science research and studies involving personal data and Big Data. Note how 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 immediate and increasing value to the stakeholders than is typical or expected typically.

## 7.1 Example Scenario: Research System for Computational Social Science

In order to achieve low-risk high-value research outcomes efficiently, design and deployment of the coming global wave of Big Data systems should apply relevant research, such as that identified in this chapter and the book generally.

Computational Social Science (CSS) studies are based on data collected often with an extremely high resolution and scale [25]. Using computational power combined with mathematical models, such data can be used to provide insights into human nature. Much of the data collected, for example mobility traces are sensitive and private; most individuals would feel uncomfortable sharing them publicly.

The data collection in the CSS context is based on the informed consent of the participants. Countries have different bodies regulating such studies, for example Institutional Research Boards (IRBs) in the US. Although certain minimal requirements for implementing informed consent in these contexts exist [35], they may often be not very well suited for the large-scale studies, where the amount and sensitivity of the data calls for sophisticated privacy controls. As the scale of the studies grows, in terms of the number of participants, collected bits per user, and duration, the EULA-style informed consent is no longer sufficient and makes it hard to claim that participants in fact expressed informed consent.

One author (Stopczynski) has recently deployed a 1,000 phones study at Technical University



of Denmark, where freshmen students received mobile phones in order to study their networks and social behavior in the important change moment of their lives, when joining the university. The study, called SensibleDTU (<https://www.sensible.dtu.dk/?lang=en>), uses not only data collected from the mobile phones (location, Bluetooth-based proximity, call and sms logs etc.) but also from social networks, questionnaires filled out by participants, behavior in economic games and so on. As the data is collected in the context of the university, there is potentially an issue of students feeling obliged to participate in the study or that the data may influence their grades. In this context, we see the implementation of Living Informed Consent not only as a technical mean to put participants in control of the data we collect, but also to clearly and comprehensibly convey broader New Deal on Data principles such as the opt-in nature of the study, the boundaries of the data usage, and parties accessing the data. It is important for science and research to develop further solutions and options ensuring contextually appropriate rules can be applied by Big Data systems. For rules to be effectively applied, systems must not only be able to establish which rules apply but also support the right functional capabilities and have appropriate information structure, format, and meta-data.

As the study will last for several years, hopefully allowing us to see the life of a student from the very first friendships made until the graduation party, the consent must remain alive. It is again a matter of balance: we do not want the participants to feel under constant surveillance — data is used mostly in aggregated form — but at the same time to remember that the data is being collected and used. We are still trying to understand how to achieve this equilibrium: how often should we remind the users about the collection? Should they re-authorize applications from time to time? We see a great hope in the applications we create for the users to provide certain services, simple such as life-logging where they can see how active they are, what are their top places etc. and more advanced, such as artistic visualizations of their social networks. Making the user aware of the data by transforming them into value, can greatly benefit the privacy, making users constantly aware what is being collected, but also what kind of value they can get out of it.

Big Data, by its nature, represents a new set of business, legal, and technical capabilities and requirements. The key observation is that virtually all Big Data systems have yet to be designed, implemented, customized, or deployed. Institutions that are the current early adopters of today's Big Data system will soon replace those systems and the rest of the world will adopt Big Data systems in phases over time. Based upon this observation, it follows that design improvements made now or soon will have much greater impact than can be had after mass-scale adoption has occurred.

## 7.2 Scenarios of Use Today, Tomorrow, and the Day After

The New Deal on Data is designed to provide good value to all stakeholders creating, using or benefiting from personal data, but the entire vision need not be adopted before value starts to flow. The mentioned social science research study scenario, demonstrates how researchers and study participants alike derive value from New Deal on Data principles today. As more researchers use the type of systems described above, the value is predicted to increase based upon a network effect. The same dynamic is expected in other contexts as well.

Adopting New Deal on Data principles on a large scale can be accomplished iteratively, such as one economic sector, transaction type or data type at a time. A reasonable success metric for adoption of large scale visions such as the New Deal on Data is whether change management has been designed to achieve enough value at every phase for every key stakeholder group to make the change worth the effort. Value to all parties participating in the New Deal on Data increases as direct or indirect use and re-use of personal data is available in greater volumes and varieties. Such volume and variety of personal data increases as more parties and transaction types and data sets and systems adopt and interoperate within the New Deal on Data.

By staging and phasing adoption of the New Deal on Data typical objections to change based on grounds of cost, disruption, or over regulation can be addressed. Policy incentives can further address these objections, such as allowing safe harbor protections for conduct of organizations operating under the rules of a trust network. Policy makers can resolve other difficulties by

682 combinations of strategic transition management methods like allowing safe harbor compliance  
683 delays, or approving alternative adoption paths and granting other non-substantive waivers to  
684 ease any burdens of migrating to new business methods.

685     Developing relevant context and scenarios defines a clear anchor for measuring whether a  
686 given use of Big Data and personal data is consistent with measurable criteria. Such criteria  
687 can be used to establish compliance with the rules of a Trust Network and for certification by  
688 government for the right to safe harbor or other protections. Criteria applicable to business,  
689 legal, and technical aspects of a system or set of systems can be assessed, evaluated, and trace-  
690 ably proven. Such criteria can provide a basic lowest common denominator requirements and  
691 constraints for work flow, transaction flow, data flow, and service flow within the relevant con-  
692 texts and scenarios of use. The New Deal on Data provides a clear basis routed in common law  
693 and broad understandings of the social compact. Therefore, with the New Deal on Data the  
694 appropriate bundle of rights and expectations intended to cover privacy and other personal data  
695 interests in Big Data can be explicitly enumerated, debated, and eventually agreed in ways that  
696 fit relevant contexts.

697     We must move beyond the closed, laboratory-based question-and-answering process that we  
698 currently use, and begin to manage our society in a new way. We must begin to test connections  
699 in the real world far earlier and more frequently than we have ever had to do before, using the  
700 methods the Human Dynamics research group have developed with our collaborators for the  
701 Friends and Family [3] or the SensibleDTU (<https://www.sensible.dtu.dk>) study. We need  
702 to construct Living Laboratories — communities willing to try a new way of doing things or, to  
703 put it bluntly, to be guinea pigs — in order to test and prove our ideas. This is new territory  
704 and so it is important for us to constantly try out new ideas in the real world in order to see  
705 what works and what does not.

706     An example of such a Living Lab is the ‘open data city’ just launched by one author (Pent-  
707 land) with the city of Trento in Italy, along with Telecom Italia, Telefonica, the research uni-  
708 versity Fondazione Bruno Kessler, the Institute for Data Driven Design, and local companies.

709 Importantly, this Living Lab has the approval and informed consent of all its participants. Not  
 710 only do these participants consent to sharing of their data, they know that they are part of a  
 711 gigantic experiment whose goal is to invent a better way of living. This can be a model followed  
 712 by many types of systems within and beyond the social science research contexts. More detail  
 713 on this Living Lab can be found at <http://www.mobileterritoriallab.eu/>.

714 The goal of this Living Lab is to develop new ways of sharing data to promote greater civic  
 715 engagement and exploration. One specific goal is to build upon and test trust-network software  
 716 such as our openPDS system. Tools such as openPDS make it safe for individuals to share  
 717 personal data (e.g., health data, facts about your children) by controlling where your data go  
 718 and what is done with them.

719 The specific research questions we are exploring depend upon a set of “personal data ser-  
 720 vices” designed to enable users to collect, store, manage, disclose, share, and use data about  
 721 themselves. These data can be used for the personal self-empowerment of each member, or  
 722 (when aggregated) for the improvement of the community through data commons that enable  
 723 social network incentives. The ability to share data safely should enable better idea flow among  
 724 individuals, companies, and government, and we want to see if these tools can in fact increase  
 725 productivity and creative output at the scale of an entire city.

## 726 8 Conclusions

727 Our societies today face unprecedented challenges. Solving these problems will require access  
 728 to personal data, so we can understand how the society works, how we move around, what  
 729 makes us productive, and how everything from ideas to diseases spread. The insights must be  
 730 actionable, available in real-time, and engaging the population, creating the nervous system of  
 731 the society. In this chapter we have reviewed how Big Data collected in institutional context  
 732 can be used for the public good. In many cases, the data needed for creating better society is  
 733 already collected and exists closed in silos of companies and governments. Using well designed  
 734 and implemented sets of institutional controls, covering business, legal, and technical dimensions,

we described how the silos can be opened. The framework for doing this — the New Deal on Data — postulates that the primary driver of the change must be by recognizing that ownership of personal data rests with the people about whom that data is about. This ownership, the right to use, transfer, and remove the data ensures that the data is available for public good, while at the same time protecting the privacy of the citizens.

The New Deal on Data is still new. Here we described our efforts in understanding the technical means of how it can be implemented, the legal framework around it, business ramifications, and the direct value that can be derived from researchers, companies, governments, and users having more access to the data. It is clear that companies must play the major role in the implementation of the New Deal, incentivized by business opportunities and pressured by the legislation and demand of the users. Only with such orchestration will it be possible to change the current feudal system of data ownership and finally put the immense quantities and capabilities of collected personal data to good use.

## References

1. Binding obligations on User-Managed Access (UMA) participants. Technical Specifications draft-maler-oauth-umatrust-01, Kantara Initiative, July 2013.
2. User-Managed Access (UMA) profile of OAuth2.0. Technical Specifications draft-hardjono-oauth-umacore-08, Kantara Initiative, December 2013.
3. Nadav Aharony, Wei Pan, Cory Ip, Inas Khayal, and Alex Pentland. Social fmri: Investigating and shaping social mechanisms in the real world. *Pervasive and Mobile Computing*, 7(6):643–659, 2011.
4. Sinan Aral and Dylan Walker. Identifying influential and susceptible members of social networks. *Science*, 337(6092):337–341, 2012.

- 758 5. Simon Barber, Xavier Boyen, Elaine Shi, and Ersin Uzun. Bitter to Better – how to  
759 make Bitcoin a better currency. In *Proceedings Financial Cryptography and Data Security*  
760 *Conference (Lecture Notes in Computer Science Volume 7397)*, pages 399–414, April 2012.
- 761 6. Ellen Barry. Protests in moldova explode, with help of twitter. *New York Times*, 8, 2009.
- 762 7. Nick Bilton. Girls around me: An app takes creepy to a new level. *The New York Times*,  
763 2012.
- 764 8. Center for Environmental & Human Toxicology University of Florida. Development of  
765 Cleanup Target Levels (CTLs) For Chapter 62-777, F.A.C. Technical report, Division of  
766 Waste Management Florida Department of Environmental Protection, February 2005.
- 767 9. Paul Lukowicz Bert Arnrich Cornelia Setz Gerhard Troster David Tacconi, Oscar Mayora  
768 and Christian Haring. Activity and emotion recognition to support early diagnosis of  
769 psychiatric diseases. pages 100–102. IEEE, 2008.
- 770 10. Yves-Alexandre de Montjoye, César A Hidalgo, Michel Verleysen, and Vincent D Blondel.  
771 Unique in the crowd: The privacy bounds of human mobility. *Scientific reports*, 3, 2013.
- 772 11. Yves-Alexandre de Montjoye, Samuel S Wang, Alex Pentland, Dinh Tien Tuan Anh, An-  
773 witaman Datta, Kevin W Hamlen, Lalana Kagal, Murat Kantarcioglu, Vaibhav Khadilkar,  
774 Kerim Yasin Oktay, et al. On the trusted use of large-scale personal data. *IEEE Data*  
775 *Eng. Bull.*, 35(4):5–8, 2012.
- 776 12. Ralph A. DeMeo and Sarah Meyer Doar. Restrictive covenants as institutional controls  
777 for remediated sites: Worth the effort? *The Florida Bar Journal*, 85(2), 2011.
- 778 13. EU Directive. 95/46/ec of the european parliament and of the council of 24 october 1995  
779 on the protection of individuals with regard to the processing of personal data and on the  
780 free movement of such data. *Official Journal of the EC*, 23:6, 1995.

- 781 14. Nathan Eagle and Alex Pentland. Reality mining: sensing complex social systems. *Per-*  
782 *sonal and ubiquitous computing*, 10(4):255–268, 2006.
- 783 15. Jonathan Woetzel et al. Preparing for china’s urban billion. 2009.
- 784 16. Florida Department of Environmental Protection - Division of Waste Management. Insti-  
785 tutional Controls Procedures Guidance. [http://www.dep.state.fl.us/waste/quick\](http://www.dep.state.fl.us/waste/quick\_topics/publications/wc/csf/icpg.pdf)  
786 [\\_topics/publications/wc/csf/icpg.pdf](http://www.dep.state.fl.us/waste/quick\_topics/publications/wc/csf/icpg.pdf), June 2012.
- 787 17. Kim Gittleson. How big data is changing the cost of insurance. *BBC News*, 2013.
- 788 18. Kate Greene. Reality mining. *Technology Review*, 2008.
- 789 19. Lev Grossman. Iran protests: Twitter, the medium of the movement. *Time Magazine*,  
790 17, 2009.
- 791 20. Aniko Hannak, Piotr Sapiezynski, Arash Molavi Kakhki, Balachander Krishnamurthy,  
792 David Lazer, Alan Mislove, and Christo Wilson. Measuring personalization of web search.  
793 In *Proceedings of the 22nd international conference on World Wide Web*, pages 527–538.  
794 International World Wide Web Conferences Steering Committee, 2013.
- 795 21. Thomas Hardjono, Patrick Deegan, and John Clippinger. On the Design of Trustworthy  
796 Compute Frameworks for Self-Organizing Digital Institutions. In *Proceedings of the 16th*  
797 *International Conference on Human-Computer Interaction*, 2014.
- 798 22. Thomas Hardjono, Daniel Greenwood, and Alex Pentland. Towards a trustworthy digital  
799 infrastructure for core identities and personal data stores. In *Proceedings of the ID360*  
800 *Conference on Identity*. University of Texas, April 2013.
- 801 23. Juniper Networks. Secure Data Access Anywhere and Anytime: Current Landscape and  
802 Future Outlook of Enterprise Mobile Security. A forrester consulting thought leadership  
803 paper commissioned by att and juniper networks, Forrester Research, October 2012.

- 804 24. Meglena Kuneva. Roundtable on Online Data Collection, Targeting and Profiling . [http://europa.eu/rapid/press-release\\_SPEECH-09-156\\_en.htm](http://europa.eu/rapid/press-release_SPEECH-09-156_en.htm), 2009.
- 805
- 806 25. David Lazer, Alex Sandy Pentland, Lada Adamic, Sinan Aral, Albert Laszlo Barabasi,  
807 Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann,  
808 et al. Life in the network: the coming age of computational social science. *Science (New*  
809 *York, NY)*, 323(5915):721, 2009.
- 810 26. Antonio Lima, Manlio De Domenico, Veljko Pejovic, and Mirco Musolesi. Exploiting  
811 cellular data for disease containment and information campaigns strategies in country-  
812 wide epidemics. School of computer science university of birmingham technical report  
813 csr-13-01, University of Birmingham, May 2013.
- 814 27. Anmol Madan, Manuel Cebrian, David Lazer, and Alex Pentland. Social sensing for  
815 epidemiological behavior change. In *Proceedings of the 12th ACM international conference*  
816 *on Ubiquitous computing*, pages 291–300. ACM, 2010.
- 817 28. AC Madrigal. Dark social: We have the whole history of the web wrong. *The Atlantic*,  
818 2013.
- 819 29. Alan Mislove, Sune Lehmann, Yong-Yeol Ahn, Jukka-Pekka Onnela, and J Niels Rosen-  
820 quist. Pulse of the nation: Us mood throughout the day inferred from twitter. *Accessed*  
821 *November*, 22(2011):2011, 2010.
- 822 30. Arvind Narayanan and Vitaly Shmatikov. Robust de-anonymization of large sparse  
823 datasets. In *Security and Privacy, 2008. SP 2008. IEEE Symposium on*, pages 111–125.  
824 IEEE, 2008.
- 825 31. Wei Pan, Yaniv Altshuler, and Alex Sandy Pentland. Decoding social influence and  
826 the wisdom of the crowd in financial trading network. In *Privacy, Security, Risk and*  
827 *Trust (PASSAT), 2012 International Conference on and 2012 International Conferenece*  
828 *on Social Computing (SocialCom)*, pages 203–209. IEEE, 2012.



- 829 32. Wei Pan, Gourab Ghoshal, Coco Krumme, Manuel Cebrian, and Alex Pentland. Urban  
830 characteristics attributable to density-driven tie formation. *Nature communications*, 4,  
831 2013.
- 832 33. ALEX PENTLAND. Reality mining of mobile communications: Toward a new deal on  
833 data. *The Global Information Technology Report 2008–2009*, page 1981, 2009.
- 834 34. Alex Pentland, David Lazer, Devon Brewer, and Tracy Heibeck. Using reality mining to  
835 improve public health and medicine. *Stud Health Technol Inform*, 149:93–102, 2009.
- 836 35. R. Pietri. Privacy in computational social science, 2013. DTU supervisor: Sune Lehmann  
837 Jørgensen, sljo@dtu.dk, DTU Compute.
- 838 36. Vivek K Singh, Laura Freeman, Bruno Lepri, and Alex Sandy Pentland. Classifying  
839 spending behavior using socio-mobile data. *HUMAN*, 2(2):pp–99, 2013.
- 840 37. Chaoming Song, Zehui Qu, Nicholas Blumm, and Albert-László Barabási. Limits of  
841 predictability in human mobility. *Science*, 327(5968):1018–1021, 2010.
- 842 38. Stan Stalnaker. The Ven currency, 2013. <http://www.ven.vc>.
- 843 39. Latanya Sweeney. Simple demographics often identify people uniquely. *Health (San Fran-*  
844 *cisco)*, pages 1–34, 2000.
- 845 40. The White House. National Strategy for Trusted Identities in Cyberspace: Enhancing On-  
846 line Choice, Efficiency, Security, and Privacy. The White House, April 2011. Available on  
847 [http://www.whitehouse.gov/sites/default/files/rss-viewer/NSTICstrategy\\_041511.pdf](http://www.whitehouse.gov/sites/default/files/rss-viewer/NSTICstrategy_041511.pdf).
- 848 41. United States Environmental Protection Agency. Institutional Controls Bibliography.  
849 <http://www.epa.gov/superfund/policy/ic/guide/biblio.pdf>, December 2005.
- 850 42. United States Environmental Protection Agency. RCRA Corrective Action Institu-  
851 tional Controls - glossary. [http://www.epa.gov/epawaste/hazard/correctiveaction/](http://www.epa.gov/epawaste/hazard/correctiveaction/resources/guidance/ics/glossary1.pdf)  
852 [resources/guidance/ics/glossary1.pdf](http://www.epa.gov/epawaste/hazard/correctiveaction/resources/guidance/ics/glossary1.pdf), 2007.

- 853 43. United States Environmental Protection Agency. Institutional Controls: A Guide to Plan-  
854 ning, Implementing, Maintaining, and Enforcing Institutional Controls at Contaminated  
855 Sites. Technical Report OSWER 9355.0-89 EPA-540-R-09-001, EPA, December 2012.
- 856 44. Jessica Vitak, Paul Zube, Andrew Smock, Caleb T Carr, Nicole Ellison, and Cliff Lampe.  
857 It's complicated: Facebook users' political participation in the 2008 election. *CyberPsy-*  
858 *chology, behavior, and social networking*, 14(3):107–114, 2011.
- 859 45. World Economic Forum. Personal Data: The Emergence of a New  
860 Asset Class, 2011. Available on [http://www.weforum.org/reports/](http://www.weforum.org/reports/personal-data-emergence-new-asset-class)  
861 [personal-data-emergence-new-asset-class](http://www.weforum.org/reports/personal-data-emergence-new-asset-class).