

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

3 Daniel "Dazza" Greenwood^{1,*}, Arkadiusz Stopczynski^{1,2}, Brian Sweatt¹, Thomas Hardjono¹,
4 Alex Sandy Pentland¹

5 **1 MIT**

6 **2 DTU**

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

8 **Contents**

9	1 The New Realities of Living in a Big Data Society	1
10	2 The New Deal on Data	4
11	3 Personal Data: Emergence of a New Asset Class	6
12	4 Enforcing the New Deal on Data	10
13	5 Transitioning End-User Assent Practices	13
14	6 Big Data and Personal Data Institutional Controls	14
15	7 Scenarios of Use in Context	17
16	7.1 Example Scenario: Research System for Computational Social Science	22
17	7.2 Scenarios of Use Today, Tomorrow, and the Day After	24
18	8 Conclusions	26

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; these types of rights have undergirded the European Union’s Privacy
 140 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. We can draw the definition of ownership from
 143 English common law on ownership rights of possession, use, and disposal:

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

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

The current personal data ecosystem is fragmented and inefficient. Too much leverage is currently being accorded to service providers that enroll and register end-users. These 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 other represent higher quality data that have been cross-correlated with other data points of the end-user.

For many participants, the risks and liabilities 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 in the ecosystem. The current focus of many Internet-based services is to capture personal data from the end-user and to sell this data to the advertising industry. Personal privacy concerns are thus inadequately addressed at best, or simply overlooked in the majority of cases. The current technologies and laws fall short of providing the legal and technical infrastructure needed to support a well-functioning digital economy.

Recently, we have shown how challenging, but also possible it is to open such institutional 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 (CDRs) 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 spread 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 [26]. 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. 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. 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].

The report of the World Economic Forum [45] suggests 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.

231 4 Enforcing the New Deal on Data

232 How can we enforce this New Deal? The threat of legal action is important, but not sufficient;
233 if you cannot see abuses then you cannot prosecute them. Moreover, who wants more lawsuits
234 anyway? Enforcement can be addressed in significant ways without prosecution of public statute
235 or regulation at all. In many fields, companies and governments rely upon rules governing com-
236 mon business, legal, and technical practices to create effective self-organization and enforcement.
237 This approach holds promise as a method by which institutional controls can form a reliable
238 operational framework for Big Data, privacy, and access.

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

249 When a trust network involves use of personal data, user permissions and corresponding
250 limits on use are fundamental to the trust model. In this context, the permissions, including
251 the provenance of the data, should require appropriate levels of audit. A well designed trust
252 network, elegantly integrating computer and legal rules, allows automatic auditing of data use
253 and allows individuals to change their permissions and withdraw data.

254 The mechanism for establishing and operating a trust network is to create system rules for
255 the applications, service providers, data, and the users themselves. System rules are sometimes
256 called operating regulations in the credit card context, trust frameworks in the identity federa-
257 tions 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 system rules design pattern allows participants in the network to be widely distributed across heterogeneous business ownership boundaries, legal governance structures, and technical security domains. Yet, the parties need not agree to conform to all or even most aspects of their basic roles, relationships, and activities in order to connect to systems of a trust network. Cross-domain trusted systems must, by their nature, focus enforceable rules narrowly upon commonly agreed upon items in order for that network to achieve its purpose.

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

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

The openPDS is a consumer version of a personal cloud trust network that we are now testing with a variety of industry and government partners. Soon, sharing your personal data

285 could become as safe and secure as transferring money between banks.

286 The Human Dynamics Lab has applied the system rules approach to development of inte-
287 grated business, technical architecture, and rules large scale institutional use of personal data
288 stores, available as an example under MIT's creative commons license at [https://github.com/](https://github.com/HumanDynamics/SystemRules)
289 `HumanDynamics/SystemRules`.

290 When it comes to data intended to be accessible over networks — whether big, personal, or
291 otherwise — the traditional container of an institution makes less and less sense. Institutional
292 controls apply, by definition by or to some type of institutional entity such as a business, gov-
293 ernmental, or religious organization. A combined view of the business, legal, and technical facts
294 and circumstances surrounding Big Data is necessary to know what access, confidentiality, and
295 other expectations exist. The relevant contextual aspects of Big Data of one institution is often
296 profoundly different from that of another. As more and more organizations use and rely upon
297 Big Data, a single formula for institutional controls will not work for increasingly heterogeneous
298 business, legal, and technical environments in play. Many organizations are structured with clear
299 leadership on business, legal, and technical issues functionally assigned to top level executive
300 roles. Business issues are typically allocated to roles such as CEO, COO, or CFO, while leader-
301 ship on legal issues is commonly assigned to roles like general counsel and regulatory compliance
302 and technical leads are often the roles of CIO, CTO, or CSO. Having top level leadership for
303 each of the business, legal, and technical aspects of a trust network is a critical success factor.

304 The capacity to apply the appropriate methods of enforcement for a trust network depend
305 upon a clear understanding and agreement among parties about the purpose of the trusted
306 system and the respective roles or expectations of those connecting as participants. Therefore,
307 an anchor is needed to a clear context of a Big Data operational framework and institutional
308 controls appropriate for access and confidentiality or privacy.

309 5 Transitioning End-User Assent Practices

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

320 Increasing the amount of data controlled by the user and granularity of this control is mean-
321 ingless if it cannot be exercised in an informed way. For many years, the End User License
322 Agreements (EULAs), long incomprehensible texts have been accepted blindly by the user,
323 trusting they have not agreed to anything that could harm them. The process of granting the
324 authorizations cannot be too complex, as it would prevent the user from understanding her deci-
325 sions. At the same time, it cannot be too simplistic, as it may not sufficiently convey the weight
326 of the privacy-related decisions. It is a challenge in itself, to build the end-user assent systems
327 that allow the user to understand and adjust their privacy settings. Complex EULAs do not
328 promote the privacy of the users, effectively pushing them to press *I Agree* in every presented
329 window.

330 This gap between the interface — single click — and the effect, can render the data owner-
331 ship meaningless; the click may wrench people and their data into systems and rules that are
332 antithetical to fair information practices, such as is prevalent with today's end-user licenses in
333 cloud services or applications. Managing the potentially long term and opposite dynamics fueled
334 by old deal systems operating simultaneously with the new deal systems is an important design
335 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 the user-controlled domain 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 the readers ask themselves a question: *Which services know which city I am in today?*. Google? Apple? Twitter? Amazon? Facebook? Flickr? This small application we have authorized a few years ago to access our Facebook check-ins and forgot since then? This is an example of a fundamental question related to user privacy and assent, and yet finding the answer to it may be surprisingly difficult in today's ecosystem. We can hope that most of the services treat the data responsibly and according to user authorizations. In the complex network of data flows however, it is relatively easy for the data to leak to careless or malicious services [7]. We need to build the solutions to help the user to make well informed decisions about data sharing.

6 Big Data and Personal Data Institutional Controls

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

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). This following definition is instructive, and is part of the Institutional Control Glossary of Terms [42]:

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

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

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

386 When institutional controls would apply to “separately owned neighboring properties” a
 387 number of issues arise that are very relevant to the problems associated with managing personal

388 data across legal, business, and other systemic boundaries. Requiring the party responsible for
 389 site cleanup to use “best efforts” to attain agreement by third parties to institute the relevant
 390 institutional controls is perhaps the most direct and least prescriptive approach. When direct
 391 negotiated agreement is not successful, then use of third party neutrals to resolve disagreements
 392 regarding institutional controls can be required. If necessary, environmental regulation can force
 393 an acquisition of neighboring land by compelling the party responsible to purchase the other
 394 property or by purchase of the property directly by the EPA [43].

395 In the context of Big Data, institutional controls are seldom, if ever, the result of government
 396 regulatory frameworks such as are seen in the environmental waste management oversight by the
 397 EPA [8,12,16]. Rather, institutions applying measures constituting institutional controls in the
 398 Big Data and related information technology and enterprise architecture contexts will typically
 399 employ governance safeguards, business practices, legal contracts, technical security, reporting,
 400 and audit programs and various risk management measures.

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

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

- 413 • *New global data provenance network*: In order for personal data to be treated like bank
 414 accounts, the origin information regarding data items coming into the data store must be

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

440 approach to model business, legal, and technical requirements in a way that can objectively be
441 agreed in advance and tested against implemented systems and components. However, those
442 are typically atomic or very granular and operate deep within layers of assumed context. Higher
443 level contexts and corresponding scenarios of multiple use cases can describe fundamental ex-
444 pectations about matters like interests in property, rights to liberty, and honoring the social
445 compact.

446 Consider that the applicable scenario within which the data exists can provide a method and
447 mechanisms of sorts to establish the basic ownership, control, and other expectations of the key
448 parties. For example, it may not be sufficient to describe the exchange of money and financial
449 information because the nature of the transaction and their respective data and systems are not
450 identified enough to predict the rights and obligations or other outcomes reasonably expected
451 by individuals and organizations that engage in the activity of a financial exchange. The sale of
452 used cars via an app, the conduct of a counseling session via Google Hangout, and the earning
453 of a masters degree via an online university all represent scenarios wherein the use case of
454 a financial exchange takes place. However, each of these scenarios occurs in contexts that are
455 easily identifiable, involving the sale of goods and deeper access to financial information if the car
456 is financed, or involving the practice of therapy by a licensed professional involving confidential
457 mental health data or involving elearning services and protected educational records and possibly
458 deeper financial information if the program is funded by scholarship or loans. Identifying the
459 people (a consumer and a used car dealer) the transaction (purchase of a used car) the data
460 (sales and title data, finance information, etc) and the systems (the third party app and it's
461 relevant services or functions, state DMV services, credit card and bank services, etc) provide
462 enough context to establish generally what existing consumer rights under the relevant state
463 lemon laws, the Uniform Commercial Code and other applicable rules will govern when duties
464 arise or are terminated, what must be promised, what can be repudiated, by whom data must
465 be kept secure and other requirements or constraints on the use of personal data and Big Data.
466 These and other factors vary when a transaction that is otherwise identical seeming operates

467 within different scenarios, and even scenarios will differ depending upon which contexts apply.

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

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

494 systems.

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

- 502 1. Who are the people in the scenario (e.g. who are the parties involved and what are their
 503 respective roles and relationships)?
- 504 2. What are the relevant interactions (e.g. what transactions or other actions are conducted
 505 by or with the people involved)?
- 506 3. What are the relevant data and data sets (e.g. what types of data are created, stored,
 507 computed, transmitted, modified or deleted)?
- 508 4. What are the relevant systems (e.g. what services or other software is used by the people,
 509 for the transactions or with the data)?

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

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

528 Another common context involving personal data is governmental transactions with the
529 public. Government filings, registrations, permits and other such public sector transactions with
530 the individuals or organizations create a large volume and variety of personal data flow. Consider
531 the scenario whereby a person runs a small business and must comply with tax, employee
532 related, licensing and other rules by filing forms with multiple government agencies at the federal,
533 state and local levels. Individuals names, addresses, occupations, dates of birth, social security
534 numbers and many other types of personal information are common elements of such filings.
535 Similarly to the retail marketing scenario above, the parties to government filing transactions
536 also risk unauthorized access to the personal data by interception during transmission or by
537 breach of data storage systems. In addition, the costs associated with requiring the same data
538 by many different agencies and updating or correcting data are born by both the filer and the
539 regulator. What if the people who own or operate such businesses had access to the services
540 and functions of a personal data store for themselves individually and also for the corporate
541 entity they operated? Routine changes in status, such as a change of address or name, could
542 be accomplished in a secure manner once via their own data service and leveraged again and
543 again by the many faces of government requiring that data. When the authoritative source
544 of such information can be deemed to be housed within or logically connected to a person's
545 data store, then the laborious task of address verification and tedious forms and other processes
546 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

573 of Denmark, where freshmen students received mobile phones in order to study their networks
574 and social behavior in the important change moment of their lives, when joining the univer-
575 sity. The study, SensibleDTU (<https://www.sensible.dtu.dk/?lang=en>), uses not only data
576 collected from the mobile phones (location, Bluetooth-based proximity, call and sms logs etc.)
577 but also from social networks, questionnaires filled out by participants, etc. As the data is col-
578 lected in the context of the university, there is potentially an issue of students feeling obliged to
579 participate in the study or that the data may influence their grades. Here, we see the implemen-
580 tation of Living Informed Consent not only as a technical mean to put participants in control
581 of the collected data, but also to clearly and comprehensibly convey broader New Deal on Data
582 principles, such as the opt-in nature of the study, the boundaries of the data usage, and parties
583 accessing the data.

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

595 Big Data, by its nature, represents a new set of business, legal, and technical capabilities and
596 requirements. Virtually all Big Data systems have yet to be designed, implemented, customized,
597 or deployed. Institutions that are the current early adopters of today's Big Data system will soon
598 replace those systems and the rest of the world will adopt Big Data systems in phases over time.
599 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 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; similarly in other contexts.

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 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 reuse of personal data is available in greater volumes and varieties.

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

Developing relevant context and scenarios defines a clear anchor for measuring whether a given use of Big Data and personal data is 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. Criteria applicable to business, legal, and technical aspects of a system or set of systems can be assessed, evaluated, and trace-

ably proven. Such criteria can provide a basic lowest common denominator requirements and constraints for work flow, transaction flow, data flow, and service flow within the relevant contexts and scenarios of use. The New Deal on Data provides a clear basis routed in common law and broad understandings of the social compact. Therefore, with the New Deal on Data the appropriate bundle of rights and expectations intended to cover privacy and other personal data interests in Big Data can be explicitly enumerated, debated, and eventually agreed in ways that fit relevant contexts.

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

An example of such a Living Lab is the ‘open data city’ just launched by one author (Pentland) with the city of Trento in Italy, along with Telecom Italia, Telefonica, the research university Fondazione Bruno Kessler, the Institute for Data Driven Design, and local companies. Importantly, this Living Lab has the approval and informed consent of all its participants. Not only do these participants consent to sharing of their data, they know that they are part of a gigantic experiment whose goal is to invent a better way of living. This can be a model followed by many types of systems within and beyond the social science research contexts. More detail on this Living Lab can be found at <http://www.mobileterritoriallab.eu/>.

The goal of this Living Lab is to develop new ways of sharing data to promote greater civic engagement and exploration. One specific goal is to build upon and test trust-network software such as our openPDS system. Tools such as openPDS make it safe for individuals to share

653 personal data (e.g., health data, facts about your children) by controlling where your data go
 654 and what is done with them.

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

662 8 Conclusions

663 Our societies today face unprecedented challenges. Solving these problems will require access
 664 to personal data, so we can understand how the society works, how we move around, what
 665 makes us productive, and how everything from ideas to diseases spread. The insights must be
 666 actionable, available in real-time, and engaging the population, creating the nervous system of
 667 the society. In this chapter we have reviewed how Big Data collected in institutional context
 668 can be used for the public good. In many cases, the data needed for creating better society is
 669 already collected and exists closed in silos of companies and governments. Using well designed
 670 and implemented sets of institutional controls, covering business, legal, and technical dimensions,
 671 we described how the silos can be opened. The framework for doing this — the New Deal on
 672 Data — postulates that the primary driver of the change must be by recognizing that ownership
 673 of personal data rests with the people about whom that data is about. This ownership, the right
 674 to use, transfer, and remove the data ensures that the data is available for public good, while
 675 at the same time protecting the privacy of the citizens.

676 The New Deal on Data is still new. Here we described our efforts in understanding the
 677 technical means of how it can be implemented, the legal framework around it, business rami-
 678 fications, 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.
5. Simon Barber, Xavier Boyen, Elaine Shi, and Ersin Uzun. Bitter to Better – how to make Bitcoin a better currency. In *Proceedings Financial Cryptography and Data Security Conference (Lecture Notes in Computer Science Volume 7397)*, pages 399–414, April 2012.
6. Ellen Barry. Protests in moldova explode, with help of twitter. *New York Times*, 8, 2009.
7. Nick Bilton. Girls around me: An app takes creepy to a new level. *The New York Times*, 2012.

- 700 8. Center for Environmental & Human Toxicology University of Florida. Development of
 701 Cleanup Target Levels (CTLs) For Chapter 62-777, F.A.C. Technical report, Division of
 702 Waste Management Florida Department of Environmental Protection, February 2005.
- 703 9. Paul Lukowicz Bert Arnrich Cornelia Setz Gerhard Troster David Tacconi, Oscar Mayora
 704 and Christian Haring. Activity and emotion recognition to support early diagnosis of
 705 psychiatric diseases. pages 100–102. IEEE, 2008.
- 706 10. Yves-Alexandre de Montjoye, César A Hidalgo, Michel Verleysen, and Vincent D Blondel.
 707 Unique in the crowd: The privacy bounds of human mobility. *Scientific reports*, 3, 2013.
- 708 11. Yves-Alexandre de Montjoye, Samuel S Wang, Alex Pentland, Dinh Tien Tuan Anh, An-
 709 witaman Datta, Kevin W Hamlen, Lalana Kagal, Murat Kantarcioglu, Vaibhav Khadilkar,
 710 Kerim Yasin Oktay, et al. On the trusted use of large-scale personal data. *IEEE Data*
 711 *Eng. Bull.*, 35(4):5–8, 2012.
- 712 12. Ralph A. DeMeo and Sarah Meyer Doar. Restrictive covenants as institutional controls
 713 for remediated sites: Worth the effort? *The Florida Bar Journal*, 85(2), 2011.
- 714 13. EU Directive. 95/46/ec of the european parliament and of the council of 24 october 1995
 715 on the protection of individuals with regard to the processing of personal data and on the
 716 free movement of such data. *Official Journal of the EC*, 23:6, 1995.
- 717 14. Nathan Eagle and Alex Pentland. Reality mining: sensing complex social systems. *Per-*
 718 *sonal and ubiquitous computing*, 10(4):255–268, 2006.
- 719 15. Jonathan Woetzel et al. Preparing for china’s urban billion. 2009.
- 720 16. Florida Department of Environmental Protection - Division of Waste Management. Insti-
 721 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)
 722 [_topics/publications/wc/csf/icpg.pdf](http://www.dep.state.fl.us/waste/quick_topics/publications/wc/csf/icpg.pdf), June 2012.
- 723 17. Kim Gittleson. How big data is changing the cost of insurance. *BBC News*, 2013.

- 724 18. Kate Greene. Reality mining. *Technology Review*, 2008.
- 725 19. Lev Grossman. Iran protests: Twitter, the medium of the movement. *Time Magazine*,
726 17, 2009.
- 727 20. Aniko Hannak, Piotr Sapiezynski, Arash Molavi Kakhki, Balachander Krishnamurthy,
728 David Lazer, Alan Mislove, and Christo Wilson. Measuring personalization of web search.
729 In *Proceedings of the 22nd international conference on World Wide Web*, pages 527–538.
730 International World Wide Web Conferences Steering Committee, 2013.
- 731 21. Thomas Hardjono, Patrick Deegan, and John Clippinger. On the Design of Trustworthy
732 Compute Frameworks for Self-Organizing Digital Institutions. In *Proceedings of the 16th*
733 *International Conference on Human-Computer Interaction*, 2014.
- 734 22. Thomas Hardjono, Daniel Greenwood, and Alex Pentland. Towards a trustworthy digital
735 infrastructure for core identities and personal data stores. In *Proceedings of the ID360*
736 *Conference on Identity*. University of Texas, April 2013.
- 737 23. Juniper Networks. Secure Data Access Anywhere and Anytime: Current Landscape and
738 Future Outlook of Enterprise Mobile Security. A forrester consulting thought leadership
739 paper commissioned by att and juniper networks, Forrester Research, October 2012.
- 740 24. Meglena Kuneva. Roundtable on Online Data Collection, Targeting and Profiling . [http:](http://europa.eu/rapid/press-release_SPEECH-09-156_en.htm)
741 [//europa.eu/rapid/press-release_SPEECH-09-156_en.htm](http://europa.eu/rapid/press-release_SPEECH-09-156_en.htm), 2009.
- 742 25. David Lazer, Alex Sandy Pentland, Lada Adamic, Sinan Aral, Albert Laszlo Barabasi,
743 Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann,
744 et al. Life in the network: the coming age of computational social science. *Science (New*
745 *York, NY)*, 323(5915):721, 2009.
- 746 26. Antonio Lima, Manlio De Domenico, Veljko Pejovic, and Mirco Musolesi. Exploiting
747 cellular data for disease containment and information campaigns strategies in country-

- 748 wide epidemics. School of computer science university of birmingham technical report
749 csr-13-01, University of Birmingham, May 2013.
- 750 27. Anmol Madan, Manuel Cebrian, David Lazer, and Alex Pentland. Social sensing for
751 epidemiological behavior change. In *Proceedings of the 12th ACM international conference*
752 *on Ubiquitous computing*, pages 291–300. ACM, 2010.
- 753 28. AC Madrigal. Dark social: We have the whole history of the web wrong. *The Atlantic*,
754 2013.
- 755 29. Alan Mislove, Sune Lehmann, Yong-Yeol Ahn, Jukka-Pekka Onnela, and J Niels Rosen-
756 quist. Pulse of the nation: Us mood throughout the day inferred from twitter. *Accessed*
757 *November, 22(2011):2011*, 2010.
- 758 30. Arvind Narayanan and Vitaly Shmatikov. Robust de-anonymization of large sparse
759 datasets. In *Security and Privacy, 2008. SP 2008. IEEE Symposium on*, pages 111–125.
760 IEEE, 2008.
- 761 31. Wei Pan, Yaniv Altshuler, and Alex Sandy Pentland. Decoding social influence and
762 the wisdom of the crowd in financial trading network. In *Privacy, Security, Risk and*
763 *Trust (PASSAT), 2012 International Conference on and 2012 International Confernece*
764 *on Social Computing (SocialCom)*, pages 203–209. IEEE, 2012.
- 765 32. Wei Pan, Gourab Ghoshal, Coco Krumme, Manuel Cebrian, and Alex Pentland. Urban
766 characteristics attributable to density-driven tie formation. *Nature communications*, 4,
767 2013.
- 768 33. ALEX PENTLAND. Reality mining of mobile communications: Toward a new deal on
769 data. *The Global Information Technology Report 2008–2009*, page 1981, 2009.
- 770 34. Alex Pentland, David Lazer, Devon Brewer, and Tracy Heibeck. Using reality mining to
771 improve public health and medicine. *Stud Health Technol Inform*, 149:93–102, 2009.

- 772 35. R. Pietri. Privacy in computational social science, 2013. DTU supervisor: Sune Lehmann
773 Jørgensen, sljo@dtu.dk, DTU Compute.
- 774 36. Vivek K Singh, Laura Freeman, Bruno Lepri, and Alex Sandy Pentland. Classifying
775 spending behavior using socio-mobile data. *HUMAN*, 2(2):pp-99, 2013.
- 776 37. Chaoming Song, Zehui Qu, Nicholas Blumm, and Albert-László Barabási. Limits of
777 predictability in human mobility. *Science*, 327(5968):1018–1021, 2010.
- 778 38. Stan Stalnaker. The Ven currency, 2013. <http://www.ven.vc>.
- 779 39. Latanya Sweeney. Simple demographics often identify people uniquely. *Health (San Fran-*
780 *cisco)*, pages 1–34, 2000.
- 781 40. The White House. National Strategy for Trusted Identities in Cyberspace: Enhancing On-
782 line Choice, Efficiency, Security, and Privacy. The White House, April 2011. Available on
783 http://www.whitehouse.gov/sites/default/files/rss_viewer/NSTICstrategy_041511.pdf.
- 784 41. United States Environmental Protection Agency. Institutional Controls Bibliography.
785 <http://www.epa.gov/superfund/policy/ic/guide/biblio.pdf>, December 2005.
- 786 42. United States Environmental Protection Agency. RCRA Corrective Action Institu-
787 tional Controls - glossary. [http://www.epa.gov/epawaste/hazard/correctiveaction/](http://www.epa.gov/epawaste/hazard/correctiveaction/resources/guidance/ics/glossary1.pdf)
788 [resources/guidance/ics/glossary1.pdf](http://www.epa.gov/epawaste/hazard/correctiveaction/resources/guidance/ics/glossary1.pdf), 2007.
- 789 43. United States Environmental Protection Agency. Institutional Controls: A Guide to Plan-
790 ning, Implementing, Maintaining, and Enforcing Institutional Controls at Contaminated
791 Sites. Technical Report OSWER 9355.0-89 EPA-540-R-09-001, EPA, December 2012.
- 792 44. Jessica Vitak, Paul Zube, Andrew Smock, Caleb T Carr, Nicole Ellison, and Cliff Lampe.
793 It’s complicated: Facebook users’ political participation in the 2008 election. *CyberPsy-*
794 *chology, behavior, and social networking*, 14(3):107–114, 2011.

- 795 45. World Economic Forum. Personal Data: The Emergence of a New
796 Asset Class, 2011. Available on <http://www.weforum.org/reports/>
797 [personal-data-emergence-new-asset-class](http://www.weforum.org/reports/personal-data-emergence-new-asset-class).