Massachusetts
Institute of
Technology

MIT
Connection
Science

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Blockchain & Infrastructure (Identity, Data Security)

PART 3
MAY 2016

David Shrier, Weige Wu, Alex Pentland
Connection Science & Engineering
Massachusetts Institute of Technology

connection.mit.edu

Massachusetts Institute of Technology

MIT Connection Science

This paper is the third of a four-part series:

- Blockchain & Financial Services: 5th Horizon of Networked Innovation: May 3
- Blockchain & Transactions, Markets and Marketplaces: May 10
- Blockchain & Infrastructure (Identity, Data Security): May 17
- Blockchain & Policy: May 24

Disclaimer

The Massachusetts Institute of Technology may have financial or other relationships with one or more entities described in this document. No endorsement, implied or explicit, is intended by discussing any of the organizations or individuals mentioned herein, and is expressly disclaimed.

I. Introduction

We are living in a world that is rapidly undergoing a fundamental change: it is becoming driven by data. This is not just the Internet of Things (IoT) or ubiquitous mobile computing, this transformation is about all societal systems — traffic, health, government, logistics, marketing, power, defense — being qualitatively more quantified and efficient, but also more transparent and accountable. This changes not only the economics of systems, but their management and funding. It also blurs the lines between customer, citizen, company, and government. Everyone gets to see what is happening, and so everyone gets to have a role in shaping these new systems.

As a consequence, businesses in financial services, financial technology, software and security are struggling to understand what the changing landscape means and how they can participate. Not only is the technical environment changing quickly, but more importantly, as this new ecology develops the systems that support it will need to adapt rapidly as well. Recent data hacks such as Target and Ashley Madison reveal the dangers of a highly networked world in which our data is gathered and held in poorly-secured repositories.

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

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

Blockchain's highly resilient architecture and distributed nature make it an interesting platform to deliver this nervous system for society. In this paper, we will explore applications of blockchain to identity and data security.

II. Identity

Identity Authentication

The need for blockchain based identity authentication is particularly salient in the internet age. While there exists somewhat imperfect systems for establishing personal identity in the physical world, in the form of Social Security numbers, state liquor identification cards, drivers' licences and even passports or national identity cards, there is no equivalent system for securing either online authentication of our personal identities or the identity of digital entities. Facebook accounts, now often used as login for different digital applications, and media access control (MAC) addresses, may come close, yet both can hardly function as trustworthy forms of identification when they can be changed at will.

So while governments can issue forms of physical identification, online identities and digital entities do not recognize national boundaries and digital identity authentication appears at first look to be an intractable problem without an overseeing global entity. Yet it would be incredibly difficult, perhaps downright impossible, to establish a global entity overseeing digital identities given that there is common backlash against even national identity cards.[2] Blockchain technology may offer a way to circumvent this problem by delivering a secure solution without the need for a trusted, central authority.

Several blockchain startups are looking to use blockchain for online identity. A ShoCard, for example, is a digital identity that protects consumer privacy. ShoCard strives to be as easy to understand and use as showing a driver’s license; and simultaneously be so secure that a bank can rely on it. The key is that the ShoCard Identity Platform is built on a public blockchain data layer, so as a company it is not storing data or keys that could be compromised. According to ShoCard all identity data is encrypted, hashed and stored in the blockchain, where it cannot be tampered with or altered. A start-up in a similar vein that bridges the gap of both human and digital entities, is Uniquid. Uniquid allows for the authentication of devices, cloud services, and people.[3] Uniquid’s aim is to provide identity and access management of connected things, as well as humans, utilizing biometric information for the latter.

One implication of this trend for financial institutions is a growing need for improved identity authentication, particularly for compliance purposes. For compliance, blockchain technology may enable financial institutions to better verify customers during the onboarding process known as Know Your Client (KYC), and to better verify parties in a transaction and the transactions themselves to prevent fraudulent activities and more effectively comply with anti-money laundering (AML) regulation. Better AML/KYC systems can be used to help extend banking services to the world's 2 billion unbanked.

Privacy-Preserving Identity on Permissioned Blockchains

Increased transparency does not necessarily mean the end of privacy. Some cryptographic identity schemes offer strong privacy protection through identity anonymity and unlinkability of transactions. A new model for privacy-preserving identities is needed if blockchain systems are to operate at a global scale. It must allow entities in the ecosystem to (a) verify the "quality" or security of an identity, (b) assess the relative "freedom" or independence of an identity from any given authority (e.g. government, businesses, etc.), and (c) assess the source of trust for a digital identity. Yet, a part of identity is derived from physically identifying a person, and part is from their behaviors. As we allow for behavioral identity models, how can systems address people who behave inconsistently – perhaps, a good person who behaves badly sometimes? As people adopt digital avatars or personae, what is the identity that is being validated?

MIT researchers have proposed ChainAnchor, a new means of establishing a trusted, yet privacy-preserving, identity. Designed for permissioned blockchains (such as those now being developed by several banks and trading platforms), the ChainAnchor architecture adds an identity and privacy-preserving layer above the blockchain. An anonymous identity verification step allows anyone to read and verify transactions from the blockchain but only anonymous verified identities can have transactions processed. Economic incentives, similar to those used in mining itself, help create resiliency in the system to defend against attacks and preserve the integrity of the identity network.

This system creates the potential for compliance with AML/KYC regulations without compromising the individual identities of counterparties in a transaction.

Transaction Monitoring

According to a 2014 survey of compliance professionals by KPMG International, only 58% of respondents stated that their organization's transaction monitoring system is able to monitor transactions across different businesses, and only 53% said they could monitor across different jurisdictions.[4] Within financial institutions, blockchain technology offers a better data infrastructure, allowing for better quality, more comprehensive and potentially even lower-cost records. It is worth noting here that financial institutions will likely prefer permissioned rather than permissionless blockchain; this means that one of the two features of blockchain, that there is no need for a central authority, is to some extent eroded. In a typical permissioned blockchain, a central organization or uniform certification utility decides who is allowed to participate, thereby partially compromising the completely decentralized nature of permissionless blockchains. However, permissioned blockchains still offer the advantage of strong consensus security and financial institutions are actively investigating advantages and disadvantages of permissioned and permissionless blockchain databases.

Ownership Rights

The strong consensus security offered by blockchain without the need for a central certifying authority renders it particularly suitable for the authentication of ownership rights. This includes digital property, intellectual property and physical property, including physical products and land. For example, Ascribe is a startup in this space. It describes itself as a "fundamentally new way to lock in attribution, securely share and trace where digital work spreads". Ascribe creates a permanent and unbreakable link between the creator and his or her creative work. By allowing ownership to be forever verified and tracked, Ascribe leverages blockchain technology to make it possible to transfer, cosign or loan digital creations similar to physical pieces of work. By preventing unauthorized access to creative work, Ascribe also helps creators monetize their work.[5]

BlockVerify, on the other hand, is an example of a startup that utilizes blockchain to attribute intellectual property through verifying the provenance of luxury goods, physical products, and, addressing the issue of counterfeit goods by verifying the legal status of pharmaceuticals, diamonds and electronics.[6] In the public domain, blockchain can have profound effects on state maintained records as well. The Economist cites an example of Mariana Catalina Izaguirre, a resident in Teguciagalpa, Honduras, whose house was demolished when the records at the country's Property Institute did not reflect the official title which she had to the land.[7] In countries where data maintenance is poor and corruption rampant, blockchain offers a reliable alternative to current registries – because the history of transactions on blockchain are immutable, corrupt individuals cannot alter the records. This sort of security happens because blockchain is decentralized, so that it does not rely on a single authority for its maintenance, and therefore a single case of mismanagement causing a point of failure does not affect the accuracy of the records.

However, technology solutions are incomplete without integration into the fabric of society. If the genesis block is hard to establish, because, for example, many cousins could put a claim on the same property, no technology can resolve the dispute.

III. Data Security

Conventional models of data security rely on creating harder and harder "walls" – adding multiple factors to authentication for access and stronger encryption. They typically rely on the same fundamental concept: once you enter the system, you can access the data. Compartmentalization is typically minimal. Edward Snowden used a combination of social engineering and a low-tech "spider" to crawl over 1.7 million documents.[8] With blockchain, there exists the potential to "scatter the stack", rendering the cost of any one breach or combination of breaches much lower. Combined with strong encryption methods and zero knowledge proofs, a much more secure method of storing and accessing data can be established, enhancing the ability of data managers to protect critical information.

Decentralized Security

Underlying all of the above applications of blockchain technology is the importance of the data being securely held – in the sense that it cannot be tampered with. Data protection and privacy is another aspect of data security. The decentralized nature of blockchain may initially appear to be at odds with privacy; this is indeed a valid concern however there are some developments to reconcile the two. Enigma, for example, is a decentralized computation platform with guaranteed privacy, and an evolution upon the blockchain technology. Enigma's goal is to enable developers to build a 'privacy by design', end-to-end decentralized application without a trusted third party.[9]

Enigma is an extension of blockchain technology, because computation and data storage are not accomplished within the blockchain, instead the blockchain is an "operating system" for secure multiparty computations carried out by storage and computation nodes participating in the network. Data is split between different nodes, and different nodes cooperate to compute functions together without leaking information to the other nodes. In summary, "no single party ever has access to data in its entirety; instead, every party has a meaningless (i.e., seemingly random) piece of it."[10]

This essentially allows data to be used while its privacy is still guaranteed. Therefore, a program could be evaluated while the inputs are kept secret.[11] For example, it may be possible for the government to find out the characteristics of welfare recipients, and the type and amount of welfare support, without accessing the identities of the welfare recipients. Victims or whistleblowers can report crimes and have their claims verified without being identified by anyone.

Blockchain, distributed computation, and zero knowledge protocols, can help banks to solve numerous multi-jurisdiction data issues and capital calculations.

Besides Enigma, privacy is also a key concern within "traditional" blockchain technology. Storj is a peer-to-peer cloud storage network and claims to be the "most secure and private cloud".[12] Factom, the first usable blockchain technology to provide an unalterable record-keeping system, has partnered with medical records and services solutions provider, HealthNautica, to secure medical records and audit trails using the blockchain. By first cryptographically encoding private medical data, patient confidentiality is protected by ensuring that medical records are not revealed to third parties, including Factom, or transferred from their original location.[13]

IV. Towards a New Deal on Data

Blockchain holds the promise of enabling the "New Deal on Data": a greater degree of personal ownership, control, and monetization of personal data, within a framework that allows society to benefit from data aggregation. A simple example of the benefits of data aggregation is the traffic congestion information within Google Maps: by contributing location, speed of travel and other critical personal information, drivers gain the benefit of the common data pool in order to realize a shorter commute time and avoidance of traffic snarls. However, for this to happen, Google must aggregate personal location information about drivers. Imagine instead a system where you, the driver, have all of the benefits of pooled data but where you, not Google, owns and controls your own data. Based on quality and magnitude of contribution, you also may in future have the option to get paid for your effort of inputting data and aiding Google's commercial proposition.

The digital breadcrumbs we leave behind are clues to who we are, what we do, how we behave in different contexts, 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."[14]
The ability to see details of so many interactions is also immensely powerful and can be used for good or for ill. Therefore, protecting personal privacy and freedom is critical to our future success as a society. We need to enable more data sharing for the public good; at the same time, we need to do a much better job of protecting the privacy of individuals.

A successful data-driven society must be able to guarantee that our data will not be abused – perhaps especially that government will not abuse the power conferred by access to such fine-grained data. There are many ways in which abuses might be directly targeted – from imposing higher insurance rates based on individual shopping history,[15] to creating problems for the entire society, by limiting user choices and enclosing users in information bubbles.[16] To achieve the potential for a new society, we require the New Deal on Data, which describes workable guarantees that the data needed for public good are readily available while at the same time protecting the citizenry.[17]

The key insight behind the New Deal on Data is that our data is worth more when shared. Aggregate data – averaged, combined across population, and often distilled to high-level features – can be used to inform improvements in systems such as public health, transportation, and government. For instance, we have demonstrated that data about the way we behave and where we go can be used to minimize the spread of infectious disease.[18] Our research has shown how digital breadcrumbs can be used to track the spread of influenza from person to person on an individual level. And the public good can be served as a result: if we can see it, we can also stop it. Similarly, if we are worried about global warming, shared, aggregated data can reveal how patterns of mobility relate to productivity.[19] This, in turn, equips us to design cities that are more productive and, at the same time, more energy efficient. However, to obtain these results and make a greener world, we must be able to see people moving around; this depends on having many people willing to contribute their data, if only anonymously and in aggregate. In addition, the Big Data transformation can help society find efficient means of governance by providing tools to analyze and understand what needs to be done, and to reach consensus on how to do it. This goes beyond simply creating more communication platforms. The assumption that more interaction between users will produce better decisions may be very misleading. Although in recent years we have seen impressive uses of social networks for better organization in society, for example during political protests,[20] we are far from even starting to reach consensus about the big problems: epidemics, climate change, pollution – big data can help us achieve such goals.

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

Personal Data as a New Asset Class

One of the first steps to promoting liquidity in land and commodity markets is to guarantee ownership rights so that people can safely buy and sell. Similarly, a first step toward creating more ideas and greater flow of ideas – idea liquidity – is to define ownership rights. The only politically viable course is to give individual citizens key rights over data that is about them, the type of rights that have undergirded the European Union's Privacy Directive since 1995.[21] We need to recognize personal data as a valuable asset of the individual, which can be given to companies and government in return for services.

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

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

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

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

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

In 2007, one of the authors, Alex Pentland, first proposed the New Deal on Data to the World Economic Forum.[22] Since then, this idea has run through various discussions and eventually helped to shape the 2012 Consumer Data Bill of Rights in the United States, along with a matching declaration on Personal Data Rights in the European Union.

The World Economic Forum (WEF) echoed the European Consumer Commissioner Meglena Kuneva in dubbing personal data the "new oil" or new resource of the 21st century.[23] The "personal data sector" of the economy today is in its infancy, its state akin to the oil industry during the late 1890s. Productive collaboration between government (building the state-owned freeways), the private sector (mining and refining oil, building automobiles), and the citizens (the user-base of these services) allowed developed nations to expand their economies by creating new markets adjacent to the automobile and oil industries.

If personal data, as the new oil, is to reach its global economic potential, productive collaboration is needed between all stakeholders in the establishment of a personal data ecosystem. A number of fundamental uncertainties exist, however, about privacy, property, global governance, human rights – essentially about who should benefit from the products and services built on personal data.[24] The rapid rate of technological change and commercialization in the use of personal data is undermining end-user confidence and trust.

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

Recently, we have seen the challenges, but also the feasibility of opening up private big data. In the Data for Development (D4D) Challenge (http:// www.d4d.orange.com), the telecommunication operator Orange opened access to a large dataset of call detail records from the Ivory Coast. Working with the data as part of a challenge, teams of researchers came up with life-changing insights for the country. For example, one team developed a model for how disease spreads in the country and demonstrated that information campaigns based on one-to-one phone conversations among members of social groups can be an effective countermeasure.[25] Data release must be carefully done, however; as we have seen in several cases, such as the Netflix Prize privacy disaster[26] and other similar privacy breaches,[27] true anonymization is extremely hard – recent research by de Montjoye et al. and others[28,29] has shown that even though human beings are highly predictable, we are also unique. Having access to one dataset may be enough to uniquely fingerprint someone based on just a few data points, and this fingerprint can be used to discover their true identity.

In releasing and analyzing the D4D data, the privacy of the people who generated the data was protected not only by technical means, such as removal of personally identifiable information (PII), but also by legal means, with the researchers signing an agreement that they would not use the data for re-identification or other nefarious purposes. Opening data from the silos by publishing static datasets – collected at some point and unchanging – is important, but it is only a beginning. We can do even more when data is available in real time and can become part of a society's nervous system. Epidemics can be monitored and prevented in real time,[30] underperforming students can be helped, and people with health risks can be treated before they get sick.[31]

The report of the World Economic Forum[32] suggests a way forward by identifying useful areas on which to focus efforts:

- Alignment of key stakeholders. Citizens, the private sector, and the public sector need to work in support of one another. Efforts such as NSTIC[33] in the United States – albeit still in its infancy – represent a promising direction for global collaboration.
- Viewing “data as money”. There needs to be a new mindset, in which an individual’s personal data items are viewed and treated in the same way as their money. These personal data items would reside in an “account” (like a bank account) where they would be controlled, managed, exchanged, and accounted for just as personal banking services operate today.
- End-user centricity. All entities in the ecosystem need to recognize end-users as vital and independent stakeholders in the co-creation and exchange of services and experiences. Efforts such as the User Managed Access (UMA) initiative[30] provide examples of system design that are user-centric and managed by the user.

When thinking about opportunity in the financial business space, entrepreneurs may wish to consider the potential of creating these new forms of data brokers – “data exchanges” that re-empower the individual and provide new revenue opportunities.

Securing the Trust Network

Blockchain holds the potential to unlock the prime requisite for a New Deal on Data: creating viable trust networks.

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

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

By bringing the code to the data, as blockchain systems do, we can now embed the rules around data access and data governance directly within the network. The ability to realize the potential of creating greater authority of an individual over their own data is at hand.

REFERENCES

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

[2] ShoCard. 2015. Homepage. Accessed 2 21, 2016. http://www.shocard.com.

[3] Uniquid. n.d. Homepage. Accessed 2 21, 2016. http://www.uniquid.co.

[4] KPMG. 2014. Global Anti-Money Laundering Survey 2014. KPMG International Co-operative.

[5] ascribe GmbH. 2016. ascribe for Artists & Creators. Accessed 2 21, 2016. http://www.ascribe.io.

[6] Blockverify. n.d. Homepage. Accessed 2 21, 2016. www.blockverify.io.

[7] The Economist. 2015. The great chain of being sure about things. 10 31. Accessed 2 21, 2017. http://www.economist.com/news/briefing/21677228-technology-behind-bitcoin-lets-people-who-do-not-know-or-trust-each-other-build-dependable.

[8] Sanger DE and E Schmitt “nowden Used Low-Cost Tool to Best N.S.A.” *New York Times,* http://www.nytimes.com/2014/02/09/us/snowden-used-low-cost-tool-to-best-nsa.html?_r=0.

[9] Zyskin, Guy, Oz Nathan, and Alex 'Sandy' Pentland. n.d. "Enigma: Decentralized Computation Platform with Guaranteed Security." White paper.

[10] *Ibid.*

[11] *Ibid.*

[12] Storj. 2016. Homepage. Accessed 2 21, 2016. http://www.storj.io.

[13] Factom. 2014. Healthnautica, Factom announce partnership. 4 23. Accessed 22 21, 2016. http://www.factom.com/healthnautica-factom-announce-partnership/.

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

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

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

[17] Pentland A, “Reality Mining of Mobile Communications.” (2009) *Social Computing and Behavioral Modeling.*

[18] Madan A, Cebrian M, Lazer D, Pentland A, “Social Sensing for Epidemiological Behavior Change,” in Proc. 12th ACM International Conference on Ubiquitous Computing (Ubicomp 2010), 291– 300; Pentland et al. “Using Reality Mining to Improve Public Health and Medicine.”

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

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

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

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

[23] *Ibid.*

[24] *Ibid.*

[25] Lima A, De Domenico M, Pejovic V, Musolesi M, "Exploiting Cellular Data for Disease Containment and Information Campaign Strategies in Country-Wide Epidemics," School of Computer Science Technical Report CSR-13-01, University of Birmingham, May 2013.

[26] Narayanan A, Shmatikov V, "Robust De-Anonymization of Large Sparse Datasets," in Proc. 2008 IEEE Symposium on Security and Privacy (SP), 111– 125.

[27] Latanya Sweeney, "Simple Demographics Often Identify People Uniquely," Data Privacy Working Paper 3, Carnegie Mellon University, Pittsburgh, 2000.

[28] de Montjoye Y, Wang A, Pentland A, "On the Trusted Use of Large-Scale Personal Data," IEEE Data Engineering Bulletin 35, no. 4 (2012): 5– 8.

[29] Song C, Qu Z, Blumm N, Barabasi A, "Limits of Predictability in Human Mobility," *Science* 327 (2010): 1018– 1021.

[30] Pentland A, Lazer D, Brewer D, Heibeck T, "Using Reality Mining to Improve Public Health and Medicine." *Stud Health Technol Inform.* (2009) 149:93-102.

[31] Tacconi D, Mayora O, Lukowicz P, Arnrich B, Setz C, Troster G, Haring C, "Activity and Emotion Recognition to Support Early Diagnosis of Psychiatric Diseases," in Proc. 2nd International ICST Conference on Pervasive Computing Technologies for Healthcare, 100– 102.

[32] World Economic Forum, "Personal Data."

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