On the Design of Trustworthy Compute Frameworks for Self-organizing Digital Institutions

Thomas Hardjono[1], Patrick Deegan[2], and John Henry Clippinger[2]

[1] MIT Kerberos & Internet Trust Consortium
Massachusetts Institute of Technology, Cambridge, MA 02139, USA
hardjono@mit.edu
[2] ID3 & MIT Media Lab
Massachusetts Institute of Technology, Cambridge, MA 02139, USA
patrick@idcubed.org, john@idcubed.org

Abstract. This paper provides an overview of the *Open Mustard Seed* (OMS) project that seeks to develop a social interaction platform to facilitate *group affiliations* based on Reed's Law [1]. Reed posits that the value of a network soars when users are given the tools for free and responsible association for common purposes. The OMS as tool for common association supports the ability for people to form self-organizing groups following the notion of the *data commons* put forward by Elinor Ostrom [2]. The data commons in OMS consists of various personal data which the owner has agreed to contribute into what Ostrom calls the common-pool resource, and which is to be managed by the self-organized group or institution. This paper discusses some design considerations of the OMS platform from the perspective of the privacy and security of the personal data that participate in the common-pool resource. The technical core value of the OMS lies in its construction of the *Trusted Compute Cells*, which are intended to be recombinable and embeddable units of logic, computation and storage.

Keywords: Reed's Law, personal data, open data commons, social computing, virtualization, cloud computing.

1 Introduction: Authority and Governance in the Next Generation Internet

The Internet offers a new opportunity for individuals, communities and societies to interact based on self-organized network governance. Currently there is arguably inequitable access to resources on the Internet, where incumbent service providers and digital technology providers seek to resist open network dynamics and maintain the old business models that in the long term benefit only a fraction of the Internet population [3,4]. Current social networking platforms typically rely upon proprietary business models that collect and sell personal information about users, inducing social distrust in these business models.

G. Meiselwitz (Ed.): SCSM 2014, LNCS 8531, pp. 342–353, 2014.

The *Open Mustard Seed* (OMS) is a project at the ID3 organization [5] and the MIT Media Lab. The mission of the ID3 is to develop new social ecosystems consisting of trusted self-healing digital institutions [6,7]. This mission is being realized through the development of an open data platform to enable people to share all their personal data within a legally constituted *trust framework*. This framework will allow people to have their own personal data service that can securely store and process static and dynamic data about themselves. Governed by privacy by design principles, all agreements of the trust framework support open authentication, storage, discovery, payment, auditing, market making and monetized "app store" services. These aims are in alignment with the growing quantified-self movement occurring today via the Internet.

1.1 Group Forming Networks

Today the various social network services on the Internet can be considered still rudimentary and in their infancy in the face of the promise of Reed's Law [1]. Reed posits that the value in a network increases exponentially as interactions move from a "broadcasting model" that offers "best content" (in which value is described by the number of consumers n) to a network of "peer-to-peer transactions" (where the network's value is based on "most members", mathematically denoted as n^2). However, by far the most valuable networks are based on those that *facilitate group affiliations*. When users have tools for "free and responsible association for common purposes" the value of the network soars exponentially to 2^n. This is the foundation of Reed's *Group Forming Networks* (GFN).

The work of Reed points to the need for a new *network architecture* and tools that facilitates GFNs. Such a network architecture and software systems should allow the establishment of trust and social capital in a user-centric and scalable way. This leads, furthermore, to the promise of *self-organized network governance* as a manifestation of GFNs and which holds a great deal of appeal when it comes to "Big Data". Networked technologies in the sense of Reed's GFN could enable individuals to negotiate their own social contract(s) and meet their needs more directly and responsively. It would enable the emergence of new sorts of effective, quasi-autonomous governance and self-provisioning. And it could achieve these goals without necessarily or directly requiring government. Online communities working in well-designed software environments could act more rapidly, and with more legitimacy than conventional government institutions [6].

1.2 Data Commons and Digital Law

This scenario is inspired not just by Reed's analysis of how to reap value from networks, but by the extensive scholarship of Elinor Ostrom, the Nobel Laureate in economics in 2009. In this new network architecture, self-organizing groups identified by Ostrom [2] could emerge.

Ostrom identified key principles by which self-organized groups can manage common-pool resources in fair and sustainable ways. If data were to be regarded as a common-pool resource, Ostrom's research shows how it would be possible

for online groups to devise their own *data commons* to manage their personal data in their own interests. This opens the possibility for the data commons to be the basis for self-organizing digital institutions, where "law" would have a very different character from the kinds law we know today. The development of "digital law" in self-organizing digital institutions would enable users to devise new types of legal contracts that are computationally expressible and executable.

Such an innovation would make institutional corruption and insider collusion far easier to detect and eliminate. Arcane systems of law – once based on oral traditions and printed texts – could make the great leap to computable code, providing powerful new platforms for governance. Law that is dynamic, evolvable and outcome-oriented would make the art of governance subject to the iterative innovations of Moore's Law. Designs could be experimentally tested, evaluated by actual outcomes, and made into better iterations [6].

1.3 Data Driven Societies

Fair access to data shared within "data commons" – as a manifestation of Ostrom's common-pool resources – will have tremendous economic impact, as societies today are increasingly reliant on data as the basis for economic interactions and decisions.

Today "Big Data" offers a way to examine the detailed patterns occurring within the billions of individual exchanges occurring in the Internet and other digital medium. Data such as the billions of telephone call records, credit card transactions and GPS location fixes allow us to precisely measure patterns of interaction between people. These individual exchanges lead to the realization that social influence is the most important phenomenon emerging from these exchanges [8,9].

People are highly influenced by the actions of others. It is the patterns that have to do with the flow of information between people that can provide us with the best insight. These patterns range from telephone calls, social media "tweets" to purchasing behaviors. These flows of information are central not only to the functioning of efficient systems, but key also to innovation. The spread and combination of information is the basis for innovation.

The patterns of information flow underscore the promise of data driven governance and policy. The use of Big Data to examine the fine-grain patterns of information exchanges promises greater transparency, control and stability in market behaviors as well improved social outcomes. Thus the vision of the *data driven society* assumes that we have continual access to Big Data. However, such access must be fair to all and must protect the personal privacy of individuals.

In the remainder of the current paper we provide a semi-technical discussion regarding the OMS platform design (Section 2) which seeks to provide a new infrastructure to let people build their own highly distributed social ecosystem for reliably governing shared resources or data commons, including controlling access to personal data. The OMS could be viewed as a component of a new kind of "social stack" of protocols, software and legal trust frameworks for self-organized digital institutions. Section 3 discusses the groups based on contextual

affinities within OMS. The paper is closed in Section 4 with a description of future work.

2 Design of Open Mustard Seed

In this section we discuss the two main building blocks of the OMS, namely the TCF and TCC constructs.

2.1 OMS Building Blocks: TCC and TCF

The design of OMS distinguishes two types of constructions that support the creation and management of digital representations of individuals, groups and institutions. These are the *Trusted Compute Frameworks* (TCF) and *Trusted Compute Cells* (TCC).

The TCC can be considered as a *cell* unit that can be replicated, enjoined with other cells and enhanced with capabilities that are context-specific. The TCF is a larger unit of computational capability that is designed to operate in the virtual environment atop a virtual machines layer.

Figure 1 attempts to illustrate a generic virtualization stack with a TCF environment containing the TCCs. Figure 1 (a) illustrates a TCF with multiple TCCs, where the TCF and the TCCs are viewed as a portable constructs that are moveable from one virtualizatin stack to another. Figure 1 (b) shows abstractly both TCF#2 and TCF#3 running multiple TCC cells with relationships or links among them (within the same TCF and across TCFs). A summary of the functions inside the TCC is shown in Figure 2.

Using the TCF and TCC constructs the OMS project seeks to explore the possibility of a TCF design that can support millions of TCCs, where each TCC represents an individual or a community. In this way the OMS platform can be used not only peer-to-peer interactions, but also peer-to-community and peer-to-business relationships.

2.2 Trusted Compute Frameworks (TCF)

The TCF is a portable compute unit which can be spun-up (and shut-down) by its owner at a TCF-compliant cloud provider (or self-operated infrastructure). The TCF is portable in that it can be relocated from one TCF-compliant cloud provider to another using a trustworthy migration protocol.

One useful way to view the TCF is as a *virtual resource container* within which one or more TCC operates. The primary concern of the TCF is (a) to support the secure and uninterrupted operations of the TCCs and (b) to ensure the TCF as compute unit can operate atop the virtualization stack (e.g. hypervisor layer, security monitor layer, hardware abstraction layer, etc) operated by the cloud provider.

The TCF implements a number of functions related to supporting itself as a virtual resource container:

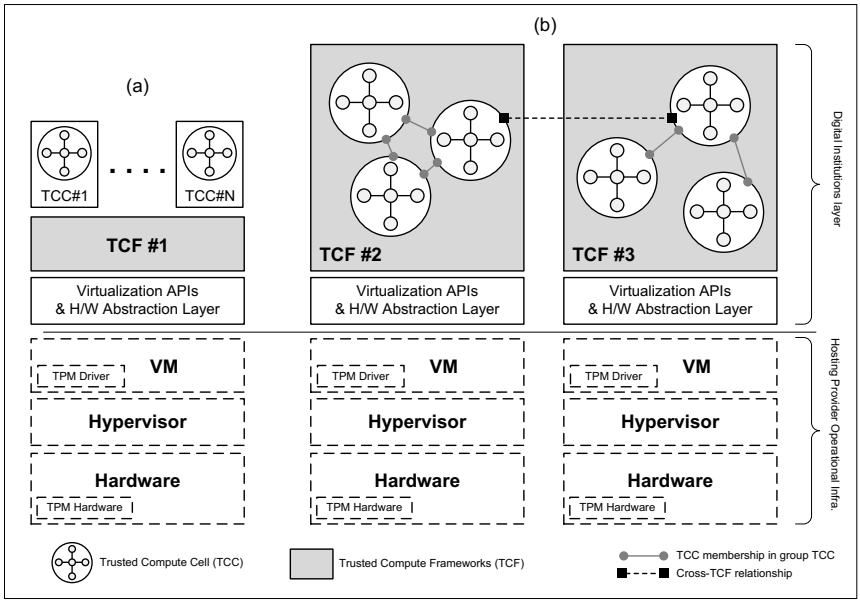

Fig. 1. Overview of TCC and TCF

- *TCF administration*: As a compute unit operating atop a virtualization stack, there are administrative tasks pertaining to the operations of the TCF itself. These include, but not limited to secure boot-up and shut-down under the owner's control, migration and the secure archiving of one or more TCC inside a TCF.
- *VM provisioning & management*: When a TCF is to be launched, a virtual machine (VM) must first be provisioned that suits the desired TCF. These include processes that interact with the underlying layers (e.g. hypervisor layer), processes for memory management, processes related to security management, and others.
- *Framework bootstrapping*: Inside the TCF, there are several processes that need to be started and managed related to the support of the TCC. These include shared databases, API end-points, registries, and so on. Some of these processes will be utilized by the applications that are run by the TCC.
- *Portal, policy & applications management*: Since the TCF by design supports the importation and the running of applications as part of the TCC these applications must be instrumented and managed through the TCF. It is envisioned that much of the social network supporting applications will operate inside the TCC, allowing the TCC to support virtual individuals, groups and institutions.
- *Security & self-protection*: As an infrastructure supporting TCCs, the TCF must provide security and resiliency against possible attacks (e.g. DDOS attacks from external sources, interference from adjacent VMs in a multi-tenant environment, etc).

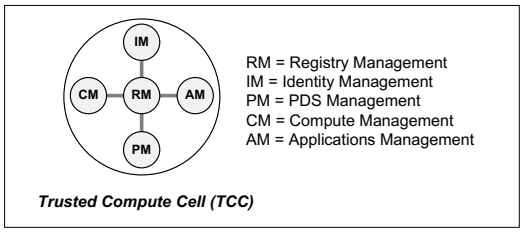

Fig. 2. Overview of functions of the Trusted Compute Cell (TCC)

2.3 Trusted Compute Cells (TCC)

The Trusted Compute Cell (TCC) is best seen from the perspective of the *social functions* it seeks to provide (as a service) to its owner. When the owner of a TCC is an individual that represents himself or herself in the virtual space, the TCC acts among others as an identity manager, personal data manager, registry of his or her connections (to other TCCs), applications execution manager and other functions.

When a TCC is created to serve as an organizational unit (e.g. social group or digital institution), the TCC has the capability to provide services that pertain to groups and group-behaviors. In this case the TCC establishes a group-identity, and also performs membership management, collective data store management, shared applications management and other group-supporting services.

In designing the TCC, the OMS project seeks to use the TCC as a cell unit from which larger "organisms" and social constructs can be created in the digital world. From the perspective of technological functions, the capabilities of the TCC are grouped under five (5) categories (see Figure 3):

1. *Identity Management*:
   The function of identity management includes authentication, authorization, audit and log, core-identity and persona management [11,12], group identity management, assertions and claims management [13], single-sign-on (SSO) establishment, and others.
2. *Personal Data Store (PDS) Management*:
   The PDS system [3,14] is a component inside the TCC which collects data (or receives streams of data) coming from the owner's devices, either generated by the device (e.g. GPS data) or proxied by the device (e.g. device pulling down copies of the owner's postings on external social network sites). The PDS system also exposes a number of APIs to external readers or consumers of the de-personalized data, such as analytics organizations and data brokers that make available the de-personalized data to the market [4,12]. An important sub-component of the PDS system is the *dynamic rule engine* which performs the role of a filtering gateway for access requests to the TCC owner's data in the PDS. The rule engine receives queries and returns answers to the querier, all the while ensuring that the responses follows the

data access policies set by the owner. As such the rule engine acts as a *Policy Enforcement Point* (PEP) for access requests to data in the PDS system.

3. *Applications Management*:
   Applications within the OMS architecture will be executed in the context of the calling (and managing) TCC. The owner of a TCC can stand-up an application for his or her sole use, or stand-up an application that will be shared by a group or community. A shared application can then be made accessible (to other TCCs who are community members) through its published APIs. As such, the management and instrumentation of applications is a core requirement of TCCs.
4. *Compute Power Management*:
   Related to applications management is the need for compute power to be expanded or reduced in an elastic manner depending on the current demand of the TCC. Elastic compute capability is particularly relevant in the case of community-shared applications, which may be shared by hundreds to millions of TCCs.
5. *Registry & Cell Management*:
   The registry in the TCC is the component that keeps track of identities, relationships, access policies, the TCC's memberships (to communities or institutions), and others. The registry also aids in the day-to-day management of the TCC by its owner. The registry acts as a *Policy Administration Point* (PAP) where the owner of a TCC can set policies regarding access to applications in the TCC (which is relevant in community-shared applications) and access to the owner's data in the PDS.

2.4 Security and Privacy Considerations

There are a number of security and privacy requirements for a TCF/TCC design and implementation. These arise from the need to protect the user's personal data in the PDS inside the TCC and from the need for the TCF as a virtualized resource container to operate in the manner for which it was designed, regardless of the cloud provider's platform on which it is running. Some key security and privacy requirements [10,15,16] include *unambiguous identification* of each TCC instance, *unhindered operation* of a TCC instance and its enveloping TCF, and *truthful attestations* reported by a TCC instance regarding its internal status.

There are a number of new and emerging trustworthy computing technologies that can be used to address some of the security and privacy requirements of the TCC and TCF design. For example, a hardware-based root of trust could be used as the basis for truthful attestations regarding not only the TCF (and the TCCs it supports), but also for the entire virtualization stack. The wide availability of hardware such as *Trusted Platform Module* (TPM) [10] on both client and server hardwares can be used as a starting point to address the security needs of the TCF and TCC. Cloud providers that seek to provide high assurance services could make use of these technologies to increase the security of their virtualization infrastructure [16]. Features such as "trusted boot" of a TCF could

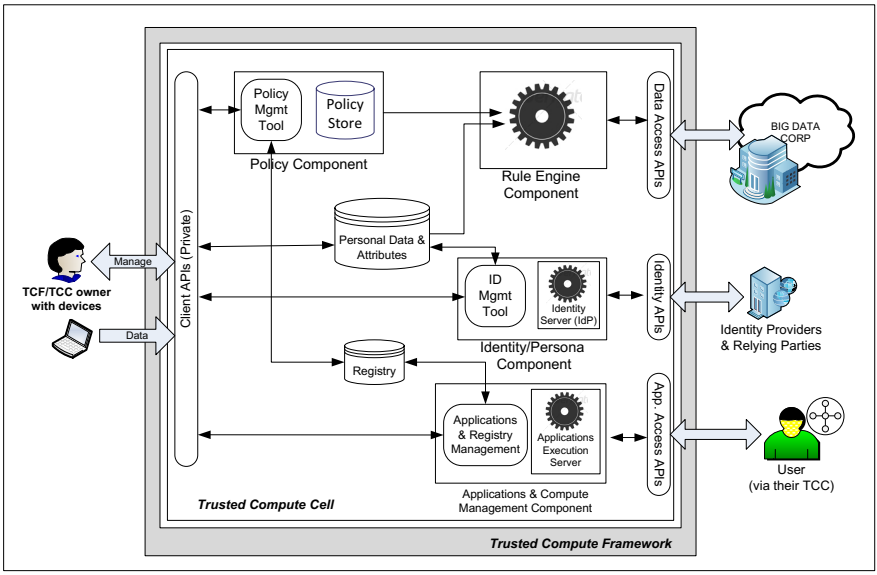

Fig. 3. Overview of Components of the Trusted Compute Cell (TCC)

be deployed more widely if these trustworthy computing hardware were deployed by cloud providers.

A number of features of the TPM hardware could be used today to increase the security of the TCF and TCC. For example, the "sealing" capability of the TPMv2.0 hardware could be used to provide data-at-rest security to a TCF. In such a scenario, when in-rest (not in operation) a TCF could be encrypted and the keys then be bound to a given hardware platform (e.g. bound to the TPM hardware belonging to the cloud provider or the TPM hardware in the owner's portable device). In this way, the launching of the TCF can only be cryptographically possible with the presence of the TCF-owner (i.e. a human owner). Similarly, a secure "TCF migration" protocol could be envisaged based on the migration protocol designed for the TPM hardware [17]. Such a migration protocol would allow a TCF-owner to safely move their TCF from one cloud provider to another with a higher degree of assurance [18].

3 OMS Communities

One of the key aims of the OMS project is to make available new infrastructure on the Internet to allow people to create their own highly distributed social ecosystems for governing shared resources, including their personal data. The OMS uses the notion of *manifests* to express modes of operations of a given TCF as well as the rules of behavior for a community that has been established using a TCF.

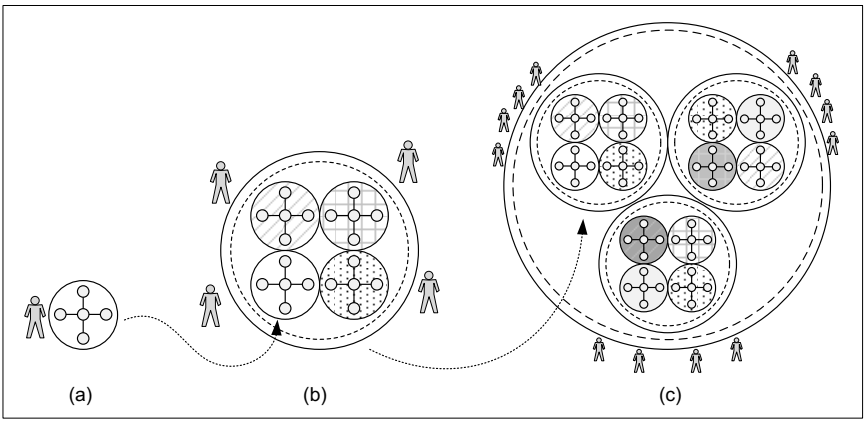

Fig. 4. (a) Private TCCs, (b) Community TCCs and (c) Institution TCCs

3.1 Creating Communities

When one or more users seek to establish a self-organizing community, they must define the purpose of the community and a number of "operating rules" for the community which are expressed internally within the TCF as manifests. Some of these operating and behavioral rules can be complex. Some examples are as follows:

- How the group is to be formed, governed, managed and evolved.
- How users interact and share information based on individual consent.
- What data is collected, and how they are accessed, stored and logged/audited.
- Access policies and access-control mechanisms by which the data is protected.
- How a user may join, suspend or withdraw from the community or institution, and how their personal data can be extracted upon departure.
- What data is retained regarding a departed user and the fact of his/her participation in the community or institution.

It is worth emphasizing here that a human person may participate in several digital communities, own and operate multiple TCFs, and thereby have "slices" of their personal data spread across several digital communities. In all these instances the common requirements include individual consent, control over personal data, and data sharing as an opt-in choice. These personal data stores should be heterogeneous distributed repositories to protect the individual against unauthorized collection of data, inference and linking of data that violates the privacy of the individual [8,4].

3.2 Private and Portal TCCs

The design of the TCC is intended to allow TCCs to be recombinable and embeddable units of logic, computation and storage. An individual person at minimum can represent himself or herself as a solitary unit by creating a lone or private TCC cell contained within a TCF (see Figure 4(a)).

However, life becomes more interesting for that person if he or she participates in a digital community through the use of one or more TCCs that he or she owns and controls. Using the same cell paradigm, the person can launch another distinct TCC that he or she can then use to establish a community-shared TCC. We refer to this as a *Portal TCC* because it represents an entry-point or portal to a shared TCC running shared applications. This is abstractly shown in Figure 4(b).

A portal TCC allows its creator to pre-define the purpose of the TCC, the applications allowed to operate in the TCC and the rules-of-operation (manifests) that govern the TCC. A complete and functioning portal TCC is thus referred to as a *Community TCC*. In order to be accepted into and participate within a Community-TCC (Figure 4(b)), an individual newcomer must agree (opt-in) to the terms of participation of the community as expressed in that TCC's manifest. Such manifests are accessible through public-APIs as a means for "discovery" of resources in that Community-TCC.

Figure 4(c) attempts to illustrate the situation where the community shown in Figure 4(b) participates in a larger community or what we refer to as an *Institution TCC*. Such an Institution-TCC also has its manifests that must be accepted by Community-TCCs and individual TCCs before they can join the Institution-TCC.

4 Future Work

There are a number of future challenges that we want to address, using the OMS as a platform for research:

- *A new Internet stack for Digital Institutions*: There is a need to broaden the notion of "layers" of the (future) Internet by introducing a new "stack". Such a stack should identify distinct layers pertaining to the personal data ecosystem, the open data commons, and digital institutions (see Figure 5). Just as in the Internet stack of today, in the *Digital Institutions Stack* each layer makes use the of "services" of the layer below it, while exposing new services and APIs to the layer above it. We envision that new Internet services will appear in each of the layers, and that each layer will evolve to become an ecosystem in itself.
- *Computational law*: The notion of self-governance is core to the value proposition of communities operating using the TCF and TCC constructs. As such, there needs to be a new perspective regarding "law as algorithm" where rules could be automatically enforced by the TCCs. In other words, law could be self-enforcing in a community that operated the TCFs and TCCs. The rule

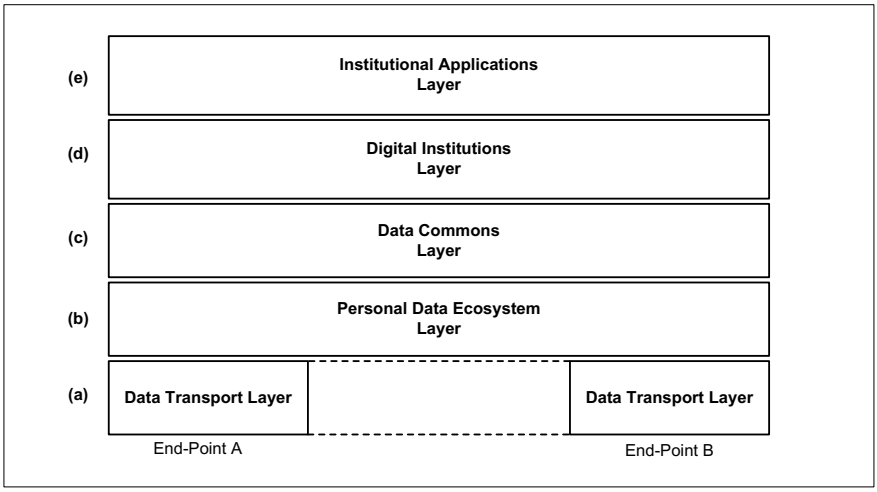

Fig. 5. A New Internet Stack for Digital Institutions

engine inside the TCC could be developed into a digital "law enforcement engine".

– *Protocols for personal data exchange*: A new generation of "protocols" need to emerge that view personal data stores (contained within TCCs) as legitimate end-points. Such a protocol would be key to making personal data a true digital asset [4]. These new protocols would not only exchange data but also observe, negotiate and enforce the legal trust frameworks governing the usage of personal data.

Acknowledgments. We thank Professor Alex (Sandy) Pentland from the MIT Media Lab for his support in this work. We also thank Stephen Buckley from the MIT Kerberos and Internet Trust (MIT-KIT) consortium for his ongoing support.

References

1. Reed, D.P.: That Sneaky Exponential – Beyond Metcalfe's Law to the Power of Community Building (1999), `http://www.reed.com/dpr/locus/gfn/reedslaw.html`
2. Ostrom, E.: Beyond Markets and States: Polycentric Governance of Complex Economic Systems. Nobel Prize Lecture (December 8, 2009), `http://www.nobelprize.org`
3. Pentland, A.: Reality Mining of Mobile Communications: Toward a New Deal on Data. In: Dutta, S., Mia, I. (eds.) The Global Information Technology Report 2008-2009: Mobility in a Networked World, World Economic Forum 2009, pp. 75–80 (2009), `http://hd.media.mit.edu/wef_globalit.pdf`

4. World Economic Forum, Personal Data: The Emergence of a New Asset Class (2011), http://www.weforum.org/reports/personal-data-emergence-new-asset-class
5. ID3, Institute for Data Driven Design (2013), http://www.idcubed.org
6. Clippinger, J.: The Next Great Internet Disruption: Authority and Governance. ID3 (2013), http://idcubed.org
7. Clippinger, J.: A Crowd of One: The Future of Individual Identity. Public Affairs/Perseus Group (2007)
8. Pentland, A.: Data Driven Societies (2012), http://media.mit.edu/pentland
9. Mani, A., Ozdaglar, A., Pentland, A.: Stable, Fair Societies as the Natural Product of Local Exchange Networks. In: Proceedings of the 2010 Workshop on Information in Networks (2010), http://media.mit.edu/pentland
10. Trusted Computing Group, TPM 1.2 Specifications (2011), http://www.trustedcomputinggroup.org
11. The Jericho Forum, "Identity Commandments," The Open Group (2011), http://www.opengroup.org
12. Hardjono, T., Greenwood, D., Pentland, A.: Towards a trustworthy digital infrastructure for core identities and personal data stores. In: Proceedings of the ID360 Conference on Identity, University of Texas (April 2013)
13. OASIS, "Assertions and Protocols for the OASIS Security Assertion Markup Language (SAML) V2.0 (March 2005), http://docs.oasis-open.org/security/saml/v2.0/saml-core-2.0-os.pdf
14. de Montjoye, Y.A., Shmueli, E., Wang, S., Pentland, A.: openPDS: Regaining ownership and privacy of personal data (2013) (Submitted for publication)
15. Trusted Computing Group, Trusted Computing Group Web Site, http://www.trustedcomputinggroup.org
16. Zic, J., Hardjono, T.: Towards a cloud-based integrity measurement service. In: Journal of Cloud Computing: Advances, Systems and Applications (February 2013)
17. Trusted Computing Group, TCG Interoperability Specifications for Backup and Migration Services (v1.0), Trusted Computing Group, TCG Issued Specification resources (June 2005), http://www.trustedcomputinggroup.org/
18. Berger, S., Caceres, R., Goldman, K.A., Perez, R., Sailer, R., van Doorn, L.: vTPM: Virtualizing the Trusted Platform Module. In: Security 2006: 15th USENIX Security Symposium, Vancouver, Canada (July-August 2006), http://www.usenix.org