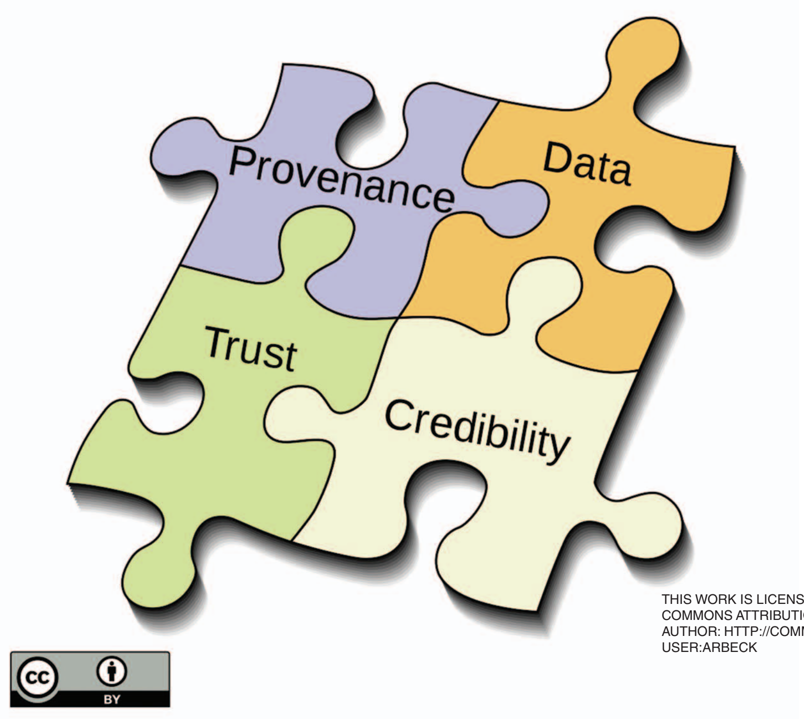

Social Use Cases for the ID3 Open Mustard Seed Platform

THOMAS HARDJONO, PATRICK DEEGAN, AND JOHN HENRY CLIPPINGER

Digital Object Identifier 10.1109/MTS.2014.2345197
Date of publication: 17 September 2014

Open Mustard Seed (OMS) is a project of the Institute for Data-Driven Design (ID3) and the M.I.T. Media Lab that seeks to develop new social ecosystems consisting of trusted self-healing digital institutions operating on open networks. The cornerstone of OMS is an open data platform that enables people to share all their personal data within a legally constituted "trust framework." This framework allows people to initiate their own "personal data store" (PDS) that can securely store and process static and dynamic data about themselves. All elements of the trust framework – open authentication, storage, discovery, payment, auditing, market making, and monetized "app store" services – are based on "privacy by design" principles. That is, privacy, security, and trusted exchange are built into the very design of the system itself.

It is important to make these principles a functional reality in digital networks if we are going to unlock the great stores of latent value that open networks hold. As postulated by Reed's Law, 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 [1].

However, the latent value of "Group Forming Networks," or GFNs, as David Reed calls them, cannot be accessed unless there is an appropriate network architecture and associated platforms and tools. We need a network architecture and software systems that can facilitate the formation of trust and social capital in user-centric and scalable ways. This is particularly important as more sectors of commerce, governance, and social life are shaped by large databases of personal information whose opaque uses are causing legitimate concerns about data security, personal privacy, and social trust.

> Privacy, security, and trusted exchange are built into the very design of the system itself.

OMS seeks to let individuals negotiate their own social contracts regarding the uses of their personal information. By providing a consent-based platform to manage data directly and responsively, OMS enables, by design, the emergence of new sorts of effective, quasi-autonomous governance and self-provisioning. And it achieves these goals without necessarily or directly requiring "government" (as opposed to "governance"). Online communities working in well-designed software environments can act more rapidly, and with greater legitimacy, than conventional government institutions. In this article, we give examples of such collective action in three *social use cases*: data sharing in local neighborhoods, exchange of digital assets (in particular crypto-currencies), and massive (even planetary) scale data sharing.

Design of Open Mustard Seed

There are two key building blocks in the architecture of OMS: the Trusted Compute Cell (TCC) and the Trusted Compute Framework (TCF).

The Trusted Compute Cell can be considered a cell unit (see Fig. 1) that individuals control in order to specify and implement their personal data preferences in networked computing. A TCC can be replicated, conjoined with other cells, and enhanced with capabilities that are context-specific. It helps to see the TCC 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 as an identity manager, personal data manager, registry of his or her connections (to other TCCs), and an applications execution manager, among 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,

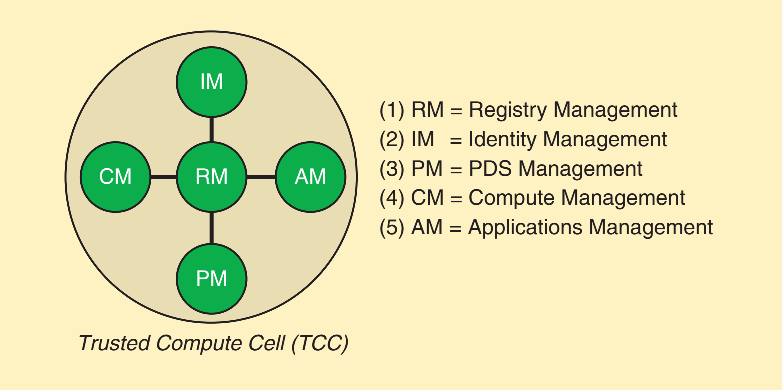
IM
CM
RM
AM
PM
Trusted Compute Cell (TCC)
(1) RM = Registry Management
(2) IM = Identity Management
(3) PM = PDS Management
(4) CM = Compute Management
(5) AM = Applications Management

Fig. 1. Components of the TCC.

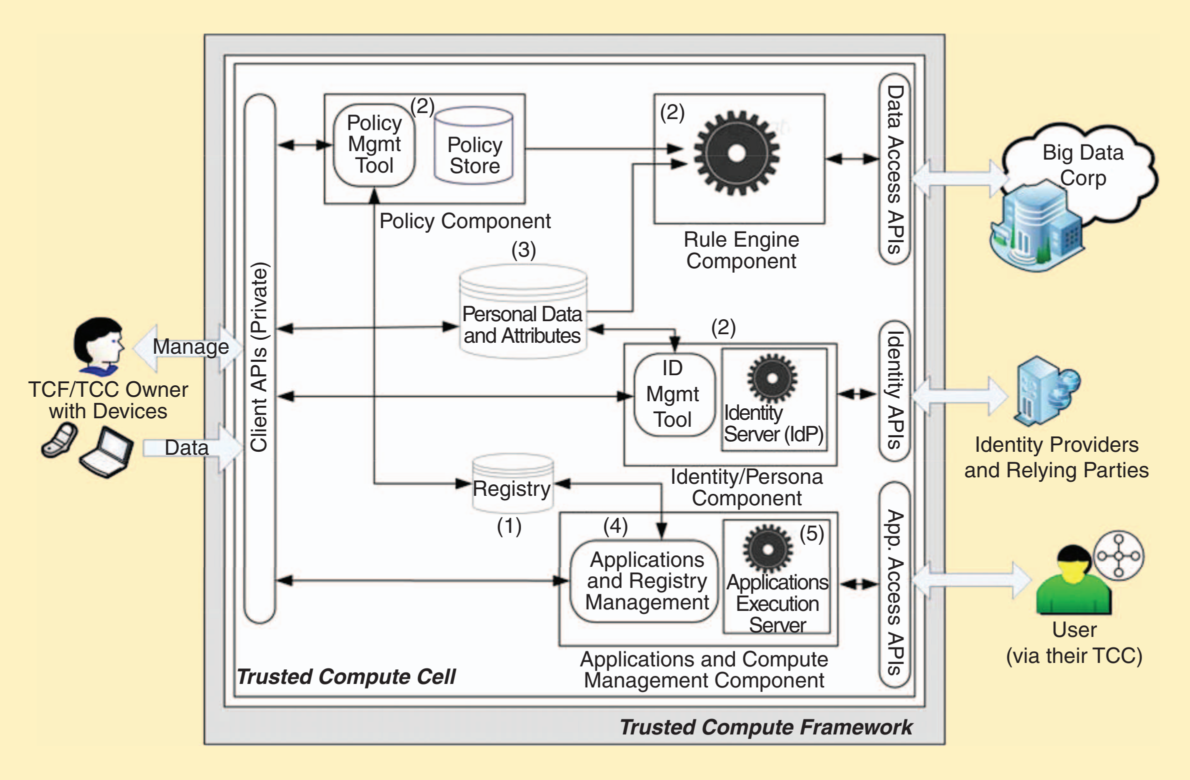

Fig. 2. Functions of the TCC.

shared applications management, and other group-supporting services.

The OMS project designed the TCC as a cell unit from which larger digital "organisms" and social constructs can be created in network spaces. To perform these functions, the TCC must fulfil five distinct technological functions, as illustrated in Fig. 2:

We need a network architecture and software systems that can facilitate the formation of trust and social capital in user-centric and scalable ways.

Identity Management: The function of identity management includes authentication, authorization, audit and log, core-identity and persona management, group identity management, assertions and claims management, single-sign-on (SSO) establishment, and more [2].

Personal Data Store (PDS) Management: The PDS system is a component inside the TCC that 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 the de-personalized data available to the market. 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.

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.

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.

Registry and Cell Management: The registry in the TCC is the

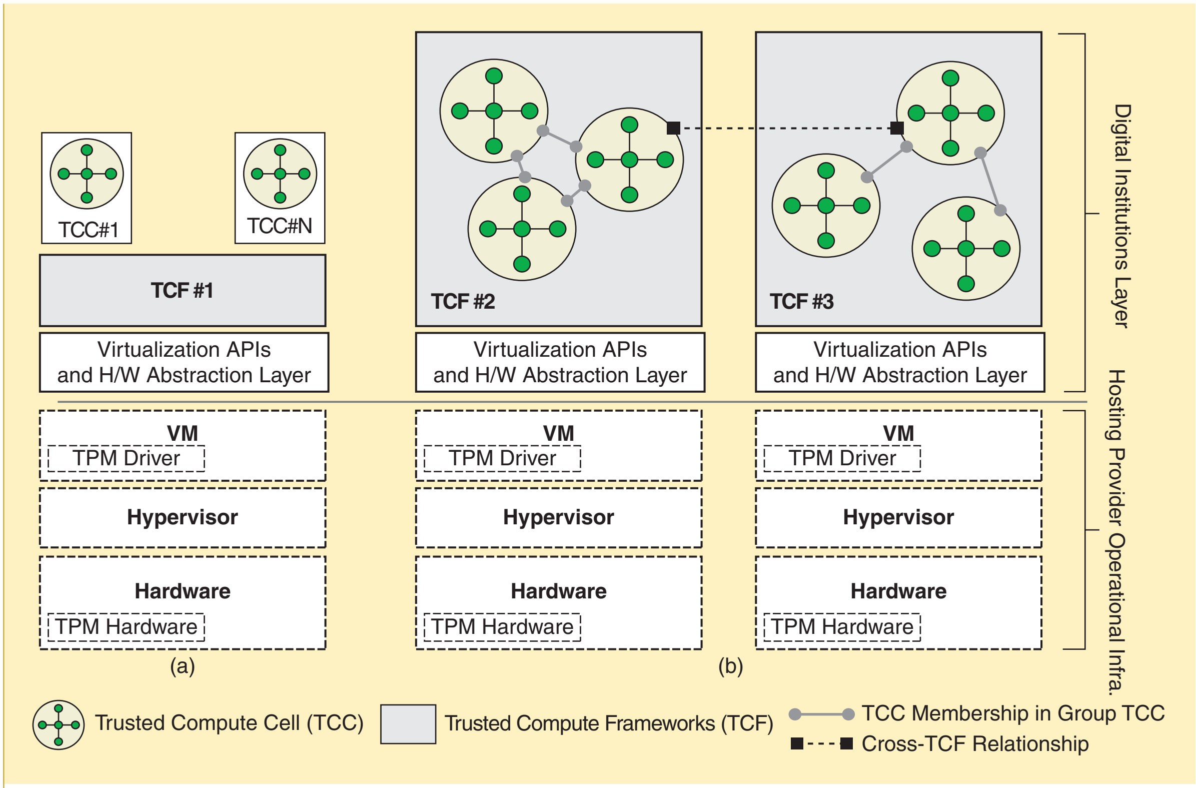
TCC#1
TCC#N
TCF #1
TCF #2
TCF #3
Virtualization APIs and H/W Abstraction Layer
Virtualization APIs and H/W Abstraction Layer
Virtualization APIs and H/W Abstraction Layer
VM
TPM Driver
Hypervisor
Hardware
TPM Hardware
VM
TPM Driver
Hypervisor
Hardware
TPM Hardware
VM
TPM Driver
Hypervisor
Hardware
TPM Hardware
(a)
(b)
Digital Institutions Layer
Hosting Provider Operational Infra.
Trusted Compute Cell (TCC)
Trusted Compute Frameworks (TCF)
TCC Membership in Group TCC
Cross-TCF Relationship

Fig. 3. Relationship between the TCC and TCF.

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.

The Trusted Compute Framework (TCF)

The TCF is a larger unit of computational capability that is designed to operate in the virtual environment on top of a virtual machines layer. One useful way to view the TCF is as a virtual resource container within which one or more TCC operate. The primary purposes of the TCF are: 1) to support the secure and uninterrupted operations of the TCCs; and 2) to ensure the TCF as a compute unit can operate atop the virtualization stack (e.g., hypervisor layer, security monitor layer, hardware abstraction layer, etc.) operated by the cloud provider.

Fig. 3 illustrates a generic virtualization stack with a TCF environment containing the TCCs. Fig. 3(a) illustrates a TCF with multiple TCCs, where the TCF and the TCCs are viewed as a portable constructs that are moveable from one virtualization stack to another. Fig. 3(b) shows two different TCFs (#2 and #3) running multiple TCC cells with relationships or links among them (within the same TCF and across TCFs).

> The vision of a data-driven society is not likely to progress unless we can develop credible systems of law and governance to protect the security and privacy of personal data.

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.

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

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 secure boot-up and shutdown under the owner's control, migration and the secure archiving of one or more TCC inside a TCF.

VM provisioning and 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.

Policy and 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.

Security and 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.).

At a minimum, an individual person can represent himself or herself as a solitary unit by creating a lone or private TCC cell contained within a TCF. 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.

Data Commons and Digital Law

The OMS architecture and functionality is inspired not just by Reed's analysis of how to reap value from networks, but also by the extensive scholarship of Elinor Ostrom, the Nobel Laureate in economics in 2009. Ostrom's pioneering work identified key principles by which self-organized groups can manage common-pool resources in fair and sustainable ways [4]. If data were to be regarded as a common-pool resource, Ostrom's research suggests that it would be possible for online groups to devise their own data commons to manage their personal data in their own interests.

These insights open the possibility for the data commons to be the basis for self-organizing digital institutions in which law would have a very different character from the kinds of 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. New forms of law based on computable code could provide powerful new platforms for governance and new checks against corruption and insider collusion [4]. Law could become more dynamic, evolvable and outcome-oriented, and the art of governance could be subject to the iterative innovations of Moore's Law. Designs could be experimentally tested, evaluated by actual outcomes, and made into better iterations.

The vision of a data-driven society [5]–[7] is not likely to progress, however, unless we can develop credible systems of law and governance to protect the security and private of personal data. Open Mustard Seed seeks to provide just such a platform. The remainder of this chapter is a semi-technical discussion of the design of the OMS infrastructure. The basic goal is to let people build their own highly distributed social ecosystems for reliably governing shared resources, including access to personal data. The OMS can be viewed as a new kind of "social stack" of protocols consisting of software and legal trust frameworks for self-organized digital institutions.

Security and Privacy Considerations

The OMS system is also designed to be modular in that it can be installed by individuals within their own computer system, or be hosted and operated by a third party (such as a cloud provider). In each deployment scenario, there are a number of security and privacy issues that emerge.

Regardless of the mode of deployment, there are a number of challenges that are common across deployment situations. These translate to security and privacy requirements for a TCF/TCC design and implementation. These features protect the user's personal data in the Personal Data Store inside the TCC, and assure that the TCF operates as a virtualized resource container in the manner for which it was designed, regardless of the cloud provider platform on which it is running. Some key security and privacy requirements include unambiguous identification of each TCC instance, unhindered operations of a TCC instance and its enveloping TCF, and truthful attestations reported by a TCC instance regarding its internal status.

In the case of a hosted deployment of OMS, additional legal and technical challenges also exist. In a hosted multi-tenant environment using virtualization stacks, there is the need for non-interference across system processes as well as clear identification of components and process belonging to each OMS instance. Although a number of these challenges still exist today, the industry has begun providing technological building blocks for trustworthy computing [8]–[10] – many of which can be used for the TCC and TCF implementation.

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) [8] on both client and server hardware can be used as a starting point to address the security needs of the TCF and TCC. Features such as "trusted boot" of a TCF could be deployed more widely if this trustworthy computing hardware were deployed by cloud providers. Similarly, a secure "TCF migration" protocol could be envisaged based on the migration protocol designed for the TPM hardware. 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 [11].

How TCC and TCF Enable Users to Self-Organize OMS Communities

Given these considerations concerning data commons, digital law, security and privacy, the OMS uses the notion of manifests to express modes of operation for a given TCF as well as the rules of behavior for a community that has been established using a TCF.

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" that are expressed internally within the TCF as manifests. For example, the manifest must be able to represent and implement:

- 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; and
- 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 an individual may participate in several digital communities, own and operate multiple TCFs, and thereby have "slices" of their personal data spread across several digital communities (without any sharing of information among those communities).

OMS for Collective Action: Social Use-Cases

There are a number of features of the OMS whose merits are best illustrated through use-cases or scenarios, and which therefore point to the potential impact of the OMS aspects to future digital communities as presented in the previous section.

Neighborhood Digital Communities

One of the motivating use-cases early in the development of the OMS is the need for people living in neighborhoods to share data without being tied to or dependent upon current social networks. An example is a group of parents wishing to establish a car pool schedule for children living in the same area. Each family would establish a family TCC that would collect and retain data about the family's schedules, the GPS locations of all the family members, and other family-related behavioral data. Families who wish to participate in car pool scheme would collectively establish a community TCC into which they would contribute data (e.g., daily schedule) from their private family TCC. The community TCC would have its own "social contract" with its members, enabled and enforced by the TCC.

In this use-case the OMS facilitates group actions [1] while preserving privacy and resilience of personal data through the distributed model [12] of the family TCCs. Although small scale, this use-case illustrates the potential for the OMS as a technical mechanism to pool resources (in this case the vehicles in the neighborhood) towards a common goal [3].

Migrateable Platform for Digital Assets Exchange

One key aspect of the OMS design is the ability for applications to be executed within the TCC, thereby allowing communities to choose not only the specific goals of their shared community TCC but also the specific applications that realize the shared goals. This aspect is attractive to communities with dispersed members wishing to share or exchange digital assets or currencies, such as BitCoin [13] or Ven [14]. Thus, for example, a community TCC could run a trading application that essentially makes the TCC into a community-operated exchange for assets, one that was potentially independent of any sovereign. The migratable feature of the TCC/TCF construct means that it is not tied to any computer system or hosted service on the Internet.

> An individual may participate in several digital communities, and have "slices" of their personal data spread across several digital communities.

Quantified-Self Massive Scale Data Sharing

The recent explosion in public interest in the quantified self is easily gauged by the increasing numbers of personal health related gadgets and electronic devices. Other similar movements are also emerging seeking to collectively use personal data to better society and the environment (e.g., the Billion Person Project [15]). One common requirement in these efforts is the need to share data across communities, states, and even nationalities in order to obtain the desired global social impact. Although electronic devices on the user end somewhat satisfies the challenges related to data collection at the person level, there remains open some challenges with regards to the storage of this data and the sharing of personal data elements at a massive scale (e.g., billions of people) while preserving individual privacy.

The OMS as a platform allows the gradual scaling up of digital communities while preserving each community as distributed autonomous organization (DAO). For example, individuals who seek to share or trade their carbon consumption data [15] (as captured and measured by their electronic devices) could store this data in their individual TCC. Larger units of communities of people (e.g., neighborhoods, towns, or cities) could use a community TCC to collect certain data from participating individual TCCs, possibly adding a layer of abstraction to this data to prevent re-identification of individuals. This organic pattern of the DAO can be repeated as we scale up, using the TCC as a common building block for individual representation as well as community representation.

Looking Further Afield

In the long term we hope OMS can enable the emergence of new sorts of effective, distributed autonomous organizations that are self-provisioning and which operate on the basis of social contracts that are negotiated by its members. The hope is also that the OMS project can provide input into the technology industry, especially sectors that develop products and services in the data privacy and in the virtualization spaces (software and hardware).

Looking ahead, we think that the current notion of "layers" of the Internet will need to be expanded by introducing a new data-driven stack that recognizes the role of data at various granularities, provenances and functions. Such a data-stack should identify distinct layers pertaining to the personal data ecosystem, the open data commons, and digital institutions. The TCC and TCF components (or layers) within the OMS have begun to point to the possibility of these distinct new layers by architecturally calling-out the functions intended for each.

Related to the data-driven stack is the need for new data exchange protocols that can deliver coalesced data in the raw (or abstracted) with the contextual semantics of its creation and intended use, together with its terms of use as devised by the data's source. In this manner data can retain value at its inception – independent of its later consumers – and can be stored and transported across the Internet without losing any of its original entropy.

Finally, another motivation for the OMS project has been equitable access to data. Not only is access to accurate data important for a thriving digital economy, but also fair access by all legitimate parties is required to grow the digital economy on a global scale. As such, personal data must be truly recognized and practically treated as a new class of digital asset [7]. Enabling a person's access to data generated by his or her daily social behaviors and interactions (by virtue of their devices or online activities) goes a long way towards this end.

Author Information

Thomas Hardjono is the Executive Director and Technical Lead at the M.I.T. Kerberos & Internet Trust (MIT-KIT) Consortium.

Patrick Deegan (patrick@idcubed.org) is the CTO and Lead Architect for the Open Mustard Seed Project within the Institute for Data Driven Design (ID3).

John Henry Clippinger (john@idcubed.org) is the Executive Director and CEO of the Institute for Data Driven Design (ID3). He is also Research Scientist at the M.I.T. Media Lab Human Dynamics Group.

References

[1] D. P. Reed, "That sneaky exponential – Beyond Metcalfe's Law to the power of community building," *Reed.com*, 1999; http://www.reed.com/dpr/locus/gfn/reedslaw.html.

[2] T. Hardjono, D. Greenwood, and A. Pentland, "Towards a trustworthy digital infrastructure for core identities and personal data stores," in *Proc. ID360 Conf. on Identity* (Univ. of Texas), Apr. 2013.

[3] E. Ostrom, "Beyond markets and states: Polycentric governance of complex economic systems," 2009 Nobel Prize Lecture, *Nobelprice.org*, Dec. 8, 2009; http://www.nobelprize.org.

[4] D. Bollier and J. Clippinger, "The next great Internet disruption: Authority and governance," presented at ID3, 2013; http://idcubed.org.

[5] A. Pentland, "Reality mining of mobile communications: Toward a new deal on data," in *The Global Information Technology Report 2008–2009: Mobility in a Net-worked World*, S. Dutta and I. Mia, Eds. World Economic Forum, 2009.

[6] A. Pentland, *Social Physics: How Good Ideas Spread*. Penguin, Jan. 2014.

[7] World Economic Forum, "Personal data: The emergence of a new asset class," 2011; http://www.weforum.org/reports/personal-data-emergence-new-asset-class.

[8] Trusted Computing Group, "TPM 1.2 specifications," 2011; http://www.trustedcomputinggroup.org

[9] Trusted Computing Group, "Virtualized trusted platform architecture specification (v1.0)," Trusted Computing Group, TCG Issued Specifications, Sept. 2011; http://www.trustedcomputinggroup.org/resources.

[10] S. Berger, R. Caceres, K. A. Goldman, R. Perez, R. Sailer, and L. van Doorn, "vTPM: Virtualizing the trusted platform module," in *Security 06: 15th USENIX Security Symp.*, Vancouver, Canada, July-Aug. 2006; www.usenix.org.

[11] J. Zic and T. Hardjono, "Towards a cloud-based integrity measurement service," *J. Cloud Computing: Advances, Systems and Applications*, vol.2, no. 4, Feb. 2013.

[12] A. Pentland, "Check and balances for Big Government," *Scientific Amer.*, vol. 309, no. 4, 2013.

[13] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," *Bitcoin.org*, 2011; http://bitcoin.org/bitcoin.pdf.

[14] Ven, "Ven is a global digital currency that's easy to use and great for the environment," *ven.vc*; http://ven.vc, accessed July 15, 2014.

[15] BPP, "Billion People Project (BPP)," *billionpeoplemovement.org*, 2014; http://billionpeoplemovement.org.