CHAPTER 1

Technology Evolution, People, and Privacy

It isn't all over; everything has not been invented; the human adventure is just beginning.

—Gene Roddenberry

This chapter takes a look at the history of information, technology, beneficiaries of that technology, and their relationship to data governance development over time. Innovation in business models, technology capabilities, and the changing relationships in the ownership and accessibility to data has resulted in a fundamental shift in size and complexity of data governance systems. Additionally, the increasing trend where collective numbers of individual consumers actually drive information technology, also known as consumerization of information technology (IT), adds yet more complexity to business relationships, fiduciary duties toward data about people, and underlying system requirements.[1] In short, this chapter introduces the context of informational privacy evolution and its relationship to new, shiny, and complex things.

Complexity—in requirements, systems, and data uses—has led to increasingly sophisticated personal data management and ethical issues, the dawning of the personal information service economy, and privacy engineering as a business-critical and customer satisfaction imperative and necessity. This book will unpacked that complexity and then examine how technology and people have interacted and how this interaction has led to data privacy concerns and requirements.

[1]One of the first-known uses of the term consumerization to describe the trend of consumer to business technological advancement is in the early 2000s. See David Moschella, Doug Neal, John Taylor, and Piet Opperman, *Consumerization of Information Technology*. Leading Edge Forum, 2004. http://lef.csc.com/projects/70

The Relationship Between Information Technology Innovation and Privacy

Throughout history, one can correlate innovation and the use of information technologies to pivotal moments in the history of privacy. In fact, there are many examples where technology either directly or indirectly impacts the sharing of personal details.

Take, as an example, the Gutenberg press and the invention of movable type. The development of the printing press and movable type not only directly led to the emergence of inexpensive and easily transportable books but also contributed to the development of the notion of personal space, privacy, and individual rights, as noted in Karmaks "History of Print": "[Print] encouraged the pursuit of personal privacy. Less expensive and more portable books lent themselves to solitary and silent reading. This orientation to privacy was part of an emphasis on individual rights and freedoms that print helped to develop."[2]

Then in the19th century, technology took privacy in another direction. The book *The Devil in the White City*[3] describes another time where movement and communication, facilitated by rail travel, inexpensive paper and writing implements, and increasing literacy, also added to the mass documentation and sharing of everyday life—from grocery lists to documented invention notebooks to planning for grand world fairs. This documentation of personal life created additional rights and obligations to share that information in culturally acceptable ways. So much temporal information also helped to piece together the lives of those living in that period of explosive innovation and growth in a manner never before available to historians or anthropologists. One wonders, will we feel the same about our old MySpace postings throughout time?

Another example (also in the late 1800s) of innovation of information technology that resulted in a pivotal privacy moment was the invention of the camera—or more precisely, rolled film. In 1888, George Eastman invented film that could be put on a spool, preloaded in easy-to-handle cameras, and sold much like today's disposable cameras.[4] The technical innovation of this new film and packaging allowed for cameras to become more portable (or mobile) and thus allowed more people access to becoming "Kodakers" or photographers. These technical advances widened the range of subject matter available to the photographers to include people who did not necessarily desire their behavior to be captured on film.[5]

Two years later, prominently citing the example of photography as technology capable of intrusion upon individual space and publicity, Warren and Brandies wrote an article that first articulated the right to privacy as a matter of US jurisprudence.[6] Note, the Warren and Brandies article, "The Right to Privacy," was not the first articulation of privacy rights; in fact, one can go back to biblical times to find discussions of substantive privacy.

---

[2]"Printing: History and Development." http://karmak.org/archive/2002/08/history_of_print.html. Copyright © 1994-99 Jones International and Jones Digital Century. All rights reserved.

[3]Erik Larson, *The Devil in the White City*. New York: Vintage Books, 2003.

[4]http://inventors.about.com/od/estartinventors/ss/George_Eastman.htm

[5]As discussed in later chapters, placing value on data, reputation, and brand creates incentive for privacy preservation and assigns appropriate weight and value on technology that would escalate or diminish that value.

[6]Samuel Warren and Louis Brandeis, "The Right To Privacy," *Harvard Law Review*, 4, no. 193 (1890). www.english.illinois.edu/-people-/faculty/debaron/582/582%20readings/right%20to%20privacy.pdf

SACRED REFERENCES TO PRIVACY

By Jay Cline, President of privacy consulting firm MPC

The inventions of the camera, database, and Internet browser gave rise to modern Western ideas about privacy. But the seeds of privacy were planted in world cultures and religions long before these technological innovations.

Perhaps the first privacy-enhancing technology was the fig leaves of Adam and Eve, the first couple of the Jewish, Christian, and Islamic faiths. In Genesis 3:7, the pair implemented a bodily privacy control: "Then the eyes of both were opened, and they knew that they were naked. And they sewed fig leaves together and made themselves apron."

In Genesis 9:23, after several generations had passed, the value of bodily privacy had become a broader social objective people helped one another accomplish. This was apparent when Noah's sons discovered him drunk and unclothed in his tent: "Then Shem and Japheth took a garment, laid it on both their shoulders, and walked backward and covered the nakedness of their father. Their faces were turned backward, and they did not see their father's nakedness."[7]

This respect for bodily privacy expanded within Jewish culture to encompass all private activity, even in the public space. You could harm someone if you viewed their private affairs without their awareness. According to Rabbi David Golinkin, author of *The Right to Privacy in Judaism*,[8] the Talmud contains two teachings on this topic:

> The Mishnah in *Bava Batra* 3:7 states: "In a common courtyard, a person should not open a door opposite a door and a window opposite a window."
>
> The *Rema* adds in the *Shulhan Arukh* (*Hoshen Mishpat* 154:7) that it is forbidden to stand at your window and look into your neighbor's courtyard, "lest he harm him by looking."

The Book of Proverbs, a collection of wisdom of right living prevalent in the ancient Jewish culture, contains three verses praising the value of confidentiality:

> "Whoever goes about slandering reveals secrets, but he who is trustworthy in spirit keeps a thing covered." (11:13)
>
> "Whoever covers an offense seeks love, but he who repeats a matter separates close friends." (17:9)

[7] www.biblehub.com/genesis/9-23.htm

[8] "The Right to Privacy in Judaism," David Golinkin, Schechter Institute of Jewish Studies, http://www.myjewishlearning.com/practices/Ethics/Business_Ethics/Contemporary_Issues/Privacy/A_Responsum.shtml.

> "Argue your case with your neighbor himself, and do not reveal another's secret." (25:9)

The Christian scriptures didn't highlight the concept of privacy. But Mohammed, living 600 years after the time of Jesus, continued the Jewish respect for private affairs. Abdul Raman Saad, author of "Information Privacy and Data Protection: A Proposed Model for the Kingdom of Saudi Arabia," identified the following privacy-friendly verses in the Quran:

> "O ye who believe! enter not houses other than your own, until ye have asked permission and saluted those in them: that is best for you, in order that ye may heed (what is seemly). If ye find no one in the house, enter not until permission is given to you: if ye are asked to go back, go back: that makes for greater purity for yourselves: and God knows well all that ye do." (An-Nur: 27–28) (24:27–28)[9]

> "O ye who believe! Avoid suspicion as much (as possible): for suspicion in some cases is a sin: And spy not on each other behind their backs. Would any of you like to eat the flesh of his dead brother? Nay, ye would abhor it. . . . But fear God: For God is Oft-Returning, Most Merciful." (Al-Hujurat: 12) (49:12)[10]

As Christianity matured, its high regard for confidentiality—as an expression of obeying the biblical commandment to not bear false witness against a neighbor—became more evident. Chapter 2477 of the *Catechism of the Catholic Church*[11] states:

> *"Respect for the reputation* of persons forbids every attitude and word likely to cause them unjust injury. He becomes guilty:
>
> —of *rash judgment* who, even tacitly, assumes as true, without sufficient foundation, the moral fault of a neighbor;
>
> —of *detraction* who, without objectively valid reason, discloses another's faults and failings to persons who did not know them;
>
> —of *calumny* who, by remarks contrary to the truth, harms the reputation of others and gives occasion for false judgments concerning them."

---

[9]Information Privacy and Data Protection A Proposed Model for the Kingdom Of Saudi Arabia, Abdul Raman Saad, *Abdul Raman Saad & Associates, Malaysia,* 1981, page 3.
[10]"Information Privacy and Data Protection A Proposed Model for the Kingdom Of Saudi Arabia," Abdul Raman Saad, Abdul Raman Saad & Associates, Malaysia*,* 1981, page 29.
[11]*Catechism of the Catholic Church*, Libreria Editrice Vaticana, Citta del Vaticano, http://www.vatican.va/archive/ENG0015/_INDEX.HTM, 1993.

It could well be that it was these ancient cultural foundations, and not primarily the rise of technology, that led delegates to the United Nations in 1947 to embed a right to information privacy in section 12 of the Universal Declaration of Human Rights: "No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks."[12]

Interestingly, these seeds of privacy found in the monotheistic faiths did not grow in the same way in the East. The Mandarin word for privacy—*yin si*—generally translates as "shameful secret." According to Lu Yao-Huai, a professor at Central South University in Changa City, a person asserting a need to withhold personal information could easily be seen as selfish or antisocial. "Generally speaking, privacy perhaps remains a largely foreign concept for many Chinese people," she wrote in "Privacy and Data Privacy Issues in Contemporary China."[13]

Similarly, in their article "Privacy Protection in Japan: Cultural Influence on the Universal Value," Yohko Orito and Kiyoshi Murata, professors at Ehime and Meiji universities, respectively, explain that Japanese citizens may not share the European view that privacy is an intrinsic right. "[I]nsistence on the right to privacy as the 'right to be let alone' indicates a lack of cooperativeness as well as an inability to communicate with others," they wrote.[14]

In related research, Masahiko Mizutani, professor at Kyoto University, and Dartmouth professors James Dorsey and James Moor stated, "[T]here is no word for privacy in the traditional Japanese language; modern Japanese speakers have adopted the English word, which they pronounce *puraibashi*."[15]

In the late 1960s, there were many concerns that governments had access to massive stores of personal information in easily accessible formats. The US government's use of databases in what was then the Department of Health, Education, and Welfare, in particular, led to the first articulation of the Fair Information Practice Principles (FIPPs). The FIPPs, which will be discussed in more detail in later chapters, are widely considered the foundation of most data privacy laws and regulations.

We are at another pivotal privacy moment given the ongoing and ever accelerating pace of information technology innovation and consumerization. This acceleration is being driven by market demand—individuals who want new and different functionality

[12] www.ohchr.org/EN/UDHR/Documents/UDHR_Translations/eng.pdf

[13] Privacy and Data Privacy Issues in Contemporary China, Lü Yao-Huai, Kluwer Academic Publishers, 2005.

[14] Privacy Protection in Japan: Cultural Influence on the Universal Value, Yohko Orito and Kiyoshi Murata, http://bibliotecavirtual.clacso.org.ar/ar/libros/raec/ethicomp5/docs/htm_papers/52Orito,%20Yohko.htm

[15] The internet and Japanese conception of privacy, Masahiko Mizutani, James Dorsey, James H. Moore, Journal Ethics and Information Technology, Kluwer Academic Publishers, Volume 6, Issue 2, 2004, pages 121-128.

from technology and uses of information—and market creation—enterprises and governments attempting to capitalize on new and expanded business models. The time for privacy engineering has arrived as a necessary component to constructing systems, products, processes, and applications that involve personal information. In today's world, systems' products, processes, and applications that involve personal information must be thought of as personal information or privacy "ecosystems" and like any ecosystems must be treated in a certain way to not only exist, but also to grown, thrive, and flourish.

To better understand this moment and the precipice we stand on, it is worth taking a few steps back and reviewing the history of information technology through a history of the network.

The Information Age

Technological support for the Information Age can be described as starting with the invention of the Gutenberg press and moveable type, where documentation, movement, and sharing of information left the realm of the elite few and entered into the popular culture. Suddenly, the possibilities for data transfer and influence expanded far beyond the social circle of the "author."

The introduction of the telegraph and telephone or the ENIAC (for Electronic Numerical Integrator and Computer, which went online in 1947 and which many IT historians call the "first electronic general-purpose computer") was a similarly remarkable leap in the ability to process and data.

For the sake of simplicity, this book will focus on the recent past to discuss various stages where information technology, norms, practices, and rules combined to allow for data gathering and sharing within an enterprise and with individuals. Framing and noting the various risks and opportunities within various stages in the Information Age creates a context for the ensuing discussion surrounding the mission and purpose of the privacy engineer and the call to action for the privacy engineer's manifesto, as presented later in this book.

Within the Information Age, this discussion will focus on five separate evolutionary stages, as shown in Figure 1-1.

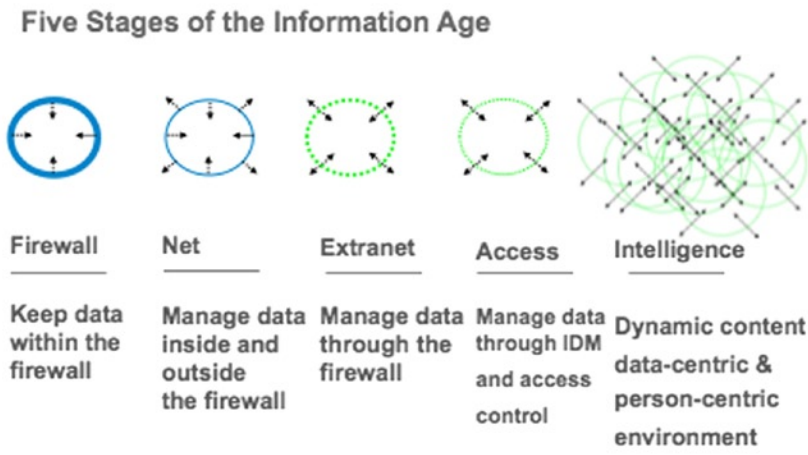

Figure 1-1. *Five stages of the age of information*

Each of these stages has evolved from one to the next in a cumulative fashion, not only because information technology became more consumer friendly and more easily accessed and implemented, but also because user, creator, and builder-driven innovation forced its evolution. Also this evolution was enabled in no uncertain terms by the realities of such things as Moore's law,[16] which correctly predicted that base technologies would become inexpensive, ubiquitous, and available for experimentation and growth.

The Firewall Stage

In the firewall stage, technology was limited[17] to discrete islands of compute capabilities (Figure 1-2). Where systems were connected to external systems, a fairly simple firewall was sufficient to maintain system integrity and exclude unauthorized users. This is that period of time before the Internet was leveraged widely as a commercial tool. Online activity, for example, was limited to networks such as Prodigy, CompuServe, and AOL. Bulletin board systems (BBS) and the Internet were the province of academics and researchers.

[16]Gordon Moore, one of the founders of Intel, observed in 1965 that the number of integrated transistors doubles approximately every 2 years with concomitant falling costs and rising efficiencies associated with production.

[17]In all of these discussions, technology limitations and capabilities are those that are widely deployed and accessible by enterprises or individuals. The first working mobile phone, for example, existed in the 1940s but did not have the innovative impact until decades later.

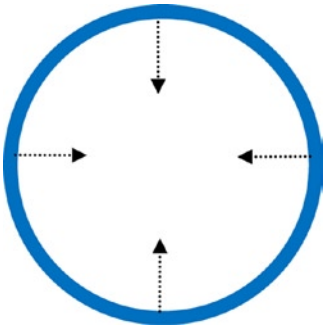

Figure 1-2. *Firewall stage*

MY LIFE WITHIN THE FIREWALL

By Michelle Finneran Dennedy

In the late 1980s I was, in fact but not title, one of the early chief information security officers for a conglomerate, multinational oil and gas company. My title, in reality, was temporary summer receptionist.

My retrospective title is based on one of the many duties required of me at the company. In addition to fetching coffee, screening visitors, and locking up packages when the addressee was unavailable, I was also in possession of "the Key." The Key opened the all-important broom closet that housed, in addition to brooms, the Wang computer that I unlocked to allow the monthly reconciliation work to happen within the accounting department, under the direction of a very distinguished white-haired gentleman named Mr. Gerold.[18]

I was never hacked. The spread sheeting capabilities were never compromised. The data was never leaked or misaddressed to the wrong party. I had a rare perfect security track record for confidentiality, integrity, and availability.

Now, the Wang computer was not linked electronically to other systems; nor did it do very much more than help the basic computations of a limited number of authorized people during the 9-to-5 workday. Limited functionality helps security and prevents privacy and confidentiality intrusion but it is also, well, not very functional or exciting.

That said, I dare any current CISO to claim that they have a perfect security track record.

[18]Not his real name, but he was truly a lovely man.

The network was still a highly controlled and governed environment where connectivity was limited by the features of the operating systems, hardware, compatibility with telephone networks, and by the expectations and practices of information technology users. An enterprise would often operate using a local area network (LAN) set of networking protocols, but its functionality and capacity were limited. Typically, data from outside sources were brought into the enterprise by means of batches or created internally and converted from analog to digital. In a like fashion, data would be moved from the enterprise in batches. People still communicated using letters created on a once ubiquitous, now museum quality, IBM Selectric typewriter. During the firewall stage, enterprise data was maintained within the protection of a digital firewall[19] as well as a physical firewall: brick, mortar, and locked filing cabinet.

Because data was contained inside physical organizational boundaries, security and privacy issues were limited and were essentially defined by the perimeters of the secure environment.

It was during the firewall stage when forward-thinking policymakers documented the FIPPs and they were adopted by the Organisation for Economic Co-operation and Development (OECD).[20] These principles became an internationally accepted set of guidelines for processing personal information. And, although the FIPPs clearly indicate the firewall stage was not without privacy concerns or the potential for greater harms, the primary focus at the time was the fear for government misuse of private information rather than commercial enterprise abuse. In addition, policymakers recognized the increasing pressure to establish a standard for handling data across jurisdictions.

Although the cost of memory, bandwidth, throughput, and compute and processing power were all still at a premium compared to today's capabilities, the increasing mobility of people and the pressure to create new, global communities foretold of an innovation bubble

Market dynamics and innovation brought compute power and network capabilities within reach of individuals and not solely the province of business and government with the availability of the affordable personal computer and Mosaic, the first Internet "browser" for the World Wide Web.

The Net Stage

The combination of the Mosaic browser, HTML (HyperText Markup Language), and customer-ready hardware and software (i.e., hardware and software that did not require an advanced engineering degree) may have been the mixture of combustibles that ignited and accelerated market dynamics and led to the consumerization of information technology that we take for granted today because it allowed nontechnical users to access and share information in a convenient fashion. It also accelerated and set in motion the dynamics that have led to the widespread consumerization of data (including personal

[19]A firewall is a system designed to prevent unauthorized access to or from a private network.

[20]Organisation for Economic Cooperation and Development (OECD), "OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data" (September 23, 1980). www.oecd.org/internet/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm

information) and the need for privacy engineering to reap this opportunity because individuals became the focus of observation, processing, and preference mining, which became one of the most powerful business models in modernity.

The net stage was a golden time for perceived anonymity (Figure 1-3). The belief was that with the net, no one knew who you were unless you announced yourself. The *New Yorker* ran a now famous cartoon showing a dog at the keyboard of a PC with the caption of "On the Internet, nobody knows you're a dog."[21] No one thought of him- or herself in a public space online unless they announced themselves (i.e., published content or by participating in an online forum).

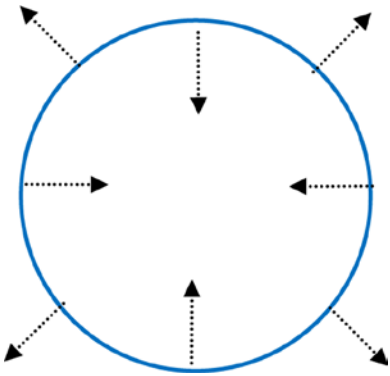

Figure 1-3. *Net stage*

The two primary privacy conversations during this time were e-marketing (i.e., spam) and identity theft. Data was increasingly transported and shared through the net, but this sharing was somewhat unidirectional. The Internet pushed data out to the public; the intranet pushed data into the enterprise. Targeted advertising and profiling were in their infancy. The net was a means of publishing and marketing. PDAs (personal digital assistants) were not connected devices for the most part. E-mail and job listings were the killer apps of the Web.

[21] http://en.wikipedia.org/wiki/File:Internet_dog.jpg

The Extranet Stage

With the introduction of the extranet,[22] the network moved into another major phase. The extranet stage[23] can be described as the age of the portal (Figure 1-4). If during the net stage the network was largely a push medium primarily used for publishing (business and governments) and reading information (consumers and citizens), extranets signaled the net as an interactive medium—an environment where one was invited behind the velvet rope into the enterprise but still not necessarily included as a fiduciary, contractor, or employee. Extranets were controlled spaces where authorized users could access information and tools and take care of limited things themselves. So-called self-service services were available to customers and other interested parties for everything from tech support to banking to benefits administration and more. Extranets allowed systems and functionality that used to exist only behind the firewall to be surfaced and exposed to "authorized" individuals.

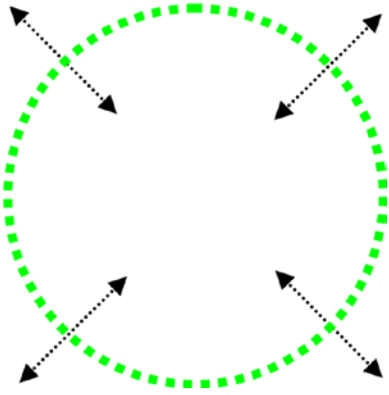

Figure 1-4. *Extranet stage*

These developments meant two things. First, an enterprise was no longer monolithic with a distinct "inside" and "outside" the firewall. The firewall became more porous as more and more ports had to be opened to allow users, functionality, and external applications in. Second, though the notion of user IDs and passwords existed before the extranet stage, the rapid growth of extranets as an enterprise facilitating and expediting medium resulted in the rapid growth of identity management solutions. The use of the extranet is significant for more reasons than the thinning of the firewall.

---

[22]An extranet is a private network that uses Internet technology and the public telecommunication system to securely share part of a business's information or operations with suppliers, vendors, partners, customers, or other businesses. It will typically have an inner firewall that protects crucial enterprise databases. There is usually an outer firewall that screens incoming data so that only invited source data is allowed in. Between the two firewalls, there may be databases that share data between external enterprises and the enterprise itself.

[23]During this stage, data were managed through a sophisticated firewall environment, but the corporate network was essentially extended to enable remote access by trusted parties.

Functionality, which heretofore was only possible in proprietary online environments, was now within reach of the many (not quite the masses yet). Users began to use the net in a fundamentally different way. It became a "private" space of interaction between designated teams, circles, and groups. Whereas before, the Web had been a publishing medium, it was now a sharing and collaboration medium.

Without a doubt, the ability to join groups changed the nature and kind of information that was now traveling the information highway. This also meant a change in "business intelligence." Whether it was the data shared, the interactions, or just the metadata[24] (data about the data and data about the interactions), business intelligence had a new resource to draw from.

Access Stage

As technology has continued to advance, more open and ubiquitous access tools and functionality information began to change the ways that people used technology, how they communicated, and, most important, what they shared. Participants were not just acquiring information, but they were also contributing, refining, sharing, and broadcasting it—sometimes to closed, selected groups and sometimes to all (i.e., the public). The key difference between the extranet stage and the access stage was the magnitude of sharing and the ease of access to enabling technology (identity management [IDM], social networks, blogs, and smartphones) (Figure 1-5). More and more, people used technology to connect with one another, to participate, and to share their lives—work and personal. Just as people had once used meetings, the water cooler, or parties as places to meet and chat and access one another, now they used the net.

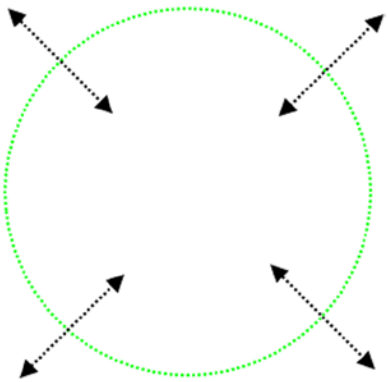

Figure 1-5. *Access stage*

[24]We will discuss metadata in detail throughout Part 2 of this book.

As the nature and ability to share grew during the access stage, so too did privacy concerns. Some of these concerns relate to the type and nature of information that individuals were willing to share in public and quasi-public settings as well as questions surrounding the general public's understanding of the power and potentially lasting impact of tools and technologies. This is a fundamental awareness or behavioral cognizance asymmetry that we still suffer from today.

Additional concerns were raised by the growing desire for governments and other enterprises to use and exploit larger and larger datasets about individual and aggregate users of technologies in the name of providing "service" or "creating community" or just plain "marketing."

Struggling legislators have grappled with these consumer and governmental interference issues by attempting to add increasing legal penalties to the miscollection and use of data. California's now watershed SB1386[25] data breach notification law is one such example, where collectors and keepers of information about people were forced to reveal data loss or theft to individual data subjects[26] in the hope of helping individuals to prepare against identity theft or other misuse.

Although this law did not come into effect until 2003—far after other comprehensive data protection laws and frameworks—this California State statute was arguably one of the first laws to create rapid, expensive, and inevitable change in corporate and governmental planning and prevention. Breach notification requirements continue to be adopted across the globe as more territories seek to protect their citizens and create requirements for tangible and measurable data protection protocols, tools, organizations and education.[27]

The Intelligence Stage

The intelligence stage is the new, now and future frontier (Figure 1-6). This stage in computing and communicating and creating is about people, devices, and systems seamlessly making handshakes, connecting, processing information, and providing services that are designed to improve the quality of life and are tailored to our needs. It is driven by increased bandwidth, throughput, processing power, analytic skills, data-reading abilities, and the desire to provide value. Here, at last, consumerization—where individuals alone or collectively—is able to drive the changes of the feature sets of computing as much as the former stages of technology forced conformity to the technology.

[25] www.leginfo.ca.gov/pub/01-02/bill/sen/sb_1351-1400/sb_1386_bill_20020926_chaptered.pdf

[26] A data subject is simply the individual who is described by data elements either alone or in combination with other data elements.

[27] The advent (or development) of the chief privacy officer (CPO) role, in particular, as well as the need for the professionalization of privacy as a distinct profession, in general, were other key developments during this stage of the Information Age.

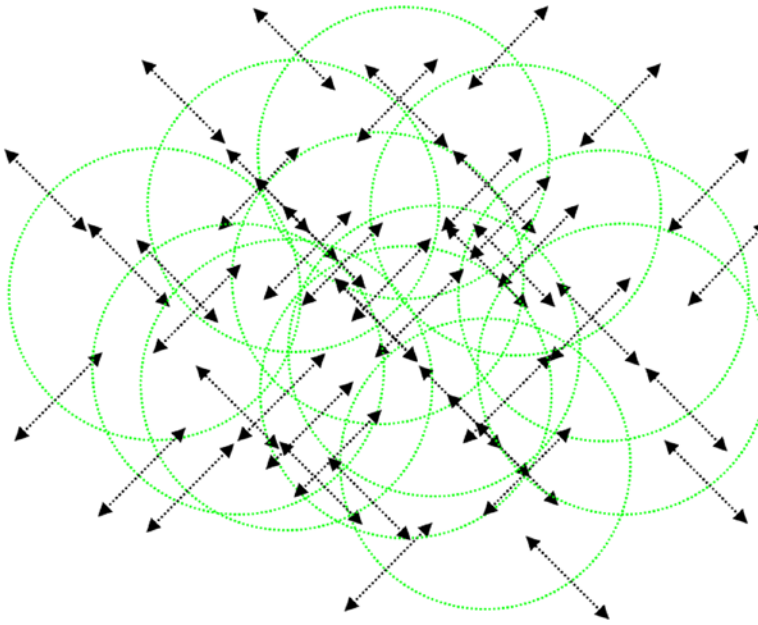

Figure 1-6. *Intelligence stage*

Some early examples of the computing in the intelligence stage are:

- Smart grid technologies recording and optimizing energy use on homes within communities
- Mapping apps that provide real-time traffic updates and suggest course corrections
- Connected appliances such as mini-bar refrigerators that automatically inventory themselves
- Augmented reality and gaming as a tool as well as recreation
- Localized shopping applications that give real-time pricing comparisons

These apps take in user-provided information, observed information or behavior, and output results that can be life improving, labor saving, and time efficient.

Whereas the hallmark of the access stage was the sharing of information, the intelligence stage may be considered as far more person and data centric rather than tool centric. In this stage, the use of information provided or collected and behavior and information observed can drive technology, social, cultural, and ethical change.

One of the implications of the dawning intelligence stage is the implication that power may be derived from being a creative, flexible thinker who can effectively gather, distill, and communicate information from a variety of sources.

THE INTERNET OF THINGS AND PRIVACY

By Tyson Macaulay, Vice President, Global Telecommunications Strategy, McAfee

Pity the fool who insists on a definition of the Internet of Things (IoT). There are literally dozens (50+ at last count by the IOS Special Working on the IoT in March 2013), originating from august and well-regarded institutions. So let's put one out there for the purposes of this discussion and leave it as a stake in the ground and reference point.

Here we go: the IoT includes devices that are manipulated by people (smartphones, desktops, tablets), devices that support very limited interfaces with people or animals (point of sale devices, medical devices), and devices that observe or manage the physical world (remote sensors, location trackers, meters, industrial controls, smart anything) in automated or semiautomated manners. And it all sits on a common network technology like Internet protocol (IP) or behind a gateway sitting on an IP network. One way or another, most of these networks are connected.

When Is Personally Identifiable Information in the IoT Actually Personally Identifiable Information?

Pity the fool who looks for consensus related to what equals personally identifiable information (PII) in the IoT.

The massive amount of data present in the IoT means there is no question that the IoT, en masse, is personal. It simply is. If you can access, correlate, and associate identity and activity in the IoT, you will pretty much be able to write a biography that will shock mothers and end marriages. Every time.

For instance, *if* you could capture the data flows from a given device (say a power meter), and *if* you could sift out the extraneous signaling and network handshakes from the service payloads, and *if* you could get the mapping of the device IP address to a subscriber ID held in a usage database, and *if* you could map the ID to a subscriber's real name held in a customer management database, *then*, maybe you might have personal information for a bachelor in a bachelor apartment and have breeched a law. Maybe.

That is a lot of *ifs*. But more important, it assumes that all this information—already segregated for business reasons unrelated to privacy—can be brought together without obstacle.

A Proposed PII Code of Conduct

What would be useful for risk mangers in the IoT are some basic rules or a code of conduct for dealing with privacy in the IoT.[28] For instance, start with a truth we can agree on, hopefully:

IoT Privacy Maxim: Information is personal if identity can be correlated with activity.

If information is about activity or events that are not about people, then it is not intrinsically personal. For instance, information about the temperature of the nickel smelter is probably not personal even in the wildest dreams of the most partisan privacy advocate.

However, if the identity of a person getting on a bus is recorded in their transit pass, and the time, date, and GPS coordinates of the bus are also logged somewhere, then PII could certainly, but not necessarily, have been generated. It is about correlation between identity and activity.

If IoT data flows contained information that could be correlated for later use or disclosure about a identifiable person, it might be PII. But not so fast!

Move to rule number 1:

IoT Privacy Rule #1: PII exists if correlation of identity to activity is viable and probable.

A frequently cited tenant of the audit profession is "would a well-informed and reasonable person agree?" When it comes to privacy and the IoT, the same tenant should apply. Is the assertion of both viability and probably of correlation rendered by reasonable and well-informed people? Is it reasonably possible to affect the correlation? Are the sources accessible such that a reasonable and well-informed person believes that it would come to pass given the time, skills, resources, and motivations of putative threat agents? Without the security jargon: is this a serious risk?

The IoT is personal to the extent that data containing both identity and activity can be correlated. Correlating an identity to the data generated by everything else a person comes into contact with physically and logically and you have the whole picture. But getting access to that identity is all too often assumed to be simple or even viable, when in fact it is not. This is where the delta between technically competent and incompetent advocates will become apparent and a danger that

[28]We do recognize this as a fraught proposition of addressing complex questions with simple answers, as Isaac Asimov so famously illustrated over 50 years ago.

swallows IT project whole, like Charybdis from Homer's *Odyssey*, sucking in ships and crew.

To people who contend an IP address is PII, we say "show us." Show us how to (legally or illegally—you choose) get logs from the devices that issue the temporary IP addresses (carrier DHCP) to gateway devices (home modems or business routers), then get account IDs assigned by different systems (RADIUS), and then the logs from the account ID system that relate the (yet again) separate billing systems, which ultimately identifies people. Then show us how you get event logs from the gateway devices (which rarely do any logging at all) and match those to the temporarily assigned internal IP addresses (home/business DHCP) within the home or business. And then make sure the person using the internal device is the same as the person paying the bills. Seriously.

This could bring us to a second rule, about viability and probability if the nature of the information is still uncertain:

IoT Privacy Rule #2: PII exists if identity and activity information exists in the same repository.

So what might "viable and probable" look like as far as identity correlation in the IoT is concerned? Identity data stored in the same repository (information source managed and accessible by the same applications, users, and administrators) as the activity artifacts associated with that identity (logs, transactions, media recordings, etc.) would viably be PII. Even if the identity data were obscured in some manner, it would still be possible through this single repository to correlate activity and an obscured identity. Meaningless but unique identifiers, over time, will usually yield identity if they can be readily compared to IoT activity.

As a counterbalance to Rule 2 is Rule 3:

IoT Privacy Guideline #3: PII is not intrinsic when identity and activity artifacts are in separate repositories.

If the identity information and IoT activity artifacts are logically or physically separated into two or more repositories, correlation should be assumed nonviable in the face of legitimate controls. In other words, if multiple repositories must be correlated and there are auditable security programs in place to prevent unapproved usage and disclosure of the data, there is no assumption that PII exists. Especially if security has been controlled among repositories, and the custodian of the information is of good character.

What About the Network?

Information and data exist in three primary states: (1) at rest (in storage of some sort), (2) in use (in active memory and being processed), or (3) in motion (within the network, moving among processing or storage).

Our earlier discussion about PII rules was centered around an assumption that data are most often accessed while at rest, in a repository. Information in use is also accessible but is far more complicated to gain access to, and a "viability" argument will rapidly come into play: accessing volatile memory used for processing requires highly specialized tools, skills, and privileges—and sometimes physical access to the guts of the system. But what about data in motion?

If you really want to know everything about someone, you tap their network connections. The ability to tap network connections is essentially the ability to watch everything. So does this mean that networks are the ultimate form of PII, being some form of "dynamic repository" subject to all the regulation and controls of PII? The answer is "no" and here are just two reasons why networks are not the ultimate vessels of PII.

First, within any given network, many of the data streams are specifically encrypted from source to destination. So understanding what is in the data stream is frequently not possible, although traffic pattern analysis remains possible even with encrypted data streams. So the PII is limited to the fact that a given network address (not "identity") communicated with a place on the Internet at a certain time and in a certain volume.

Second, most devices that originate substantial amounts of potential PII these days are mobile devices, like smartphones. Mobile devices tend to traverse many networks throughout the day. Mobile devices might start on the home Wi-Fi network, move to the 4G cellular network on the way to work, offload to the employer's office network, offload to the local café network at 10 a.m. and then again at lunch and then again at 2:30 p.m., back to the employer network, and then the 4G network, and finally the home network. All these networks are frequently, independently controlled. Also, the same device will be assigned unique, recycled IP addresses each time it jumps from network to network. Trying to track such devices and collect their traffic falls in the "nonviable" category for the National Security Agency, Superman, and probably God. Network-based correlation by default usually fails IoT Privacy Rule #1—"not viable," although exceptions will exist but must be proven rather than assumed.

The Dawning of the Personal Information Service Economy

The Information Age, the service economy, and the ability to provide and derive value from personal information are combining like never before and, accordingly, a new class of services is unfolding; these new services are classified as "personal information services." There are currently at least two classes of personal information services. The first types of services are those that aspire to help individuals manage and protect personal information. Security tools, single sign-on/identity management services, "do not track" technologies and policies, and compliance solutions for managing web-based cookies are all examples of this kind of personal information service.

The second type of personal information service are services that use personal information to provide value—sometimes to the individual and often to the enterprise. Examples of such services are personalization tailored to individual wants and desires, device recovery, or data retrieval and cloaking services. Clearly, there is overlap between data management and value-based services and a near infinite possibility for combining value propositions for personal data in emerging business, cultural, and individual value scenarios.

As individuals contribute more about what they want to do and what they want their communities to do (either socially or economically), the combination of all these actions will impact the whole economy. Personal information services may become a pivotal economics resource that can drive or measure an economy.

Data-Centric and Person-Centric Processing

There is a powerful movement toward data-centric and person-centric computing. Data centric implies that data and information processed from it are primary design drivers. Person centric implies that a person is also a primary design driver. Taken together data-centric and person-centric processing involves the processing of personal information (PI) and thus potential privacy concerns. Privacy engineering is a crucial competency when designing and implementing data-centric and person-centric systems. Data-centric and person-centric design and execution require a proactively engineered system architecture because:

- It takes data to protect data. We need to collect data from customers and those with whom customers may interact to determine whether privacy rules based on statutory or enterprise privacy policies apply.
- The scope of PI is expanding. What was once considered just "machine" data (i.e., not personal) is being recognized as something else.
- DV > DR = Success. A well-designed system ensures that data value (DV) exceeds data risk (DR)
- Privacy engineering is about user experience, brand definition and augmentation, and meeting customer satisfaction.
- Privacy engineering also translates into repeatable engineering principles rather than handcrafted one-off design and execution.

PRIVACY CAN'T BE FIXED

By John Berard, Founder, Credible Context

Got your attention? Privacy is certainly a problem that can be solved. But first our mindset needs to change. Privacy—encompassing the transparent collection, secure storage, meaningful use and scheduled deletion of personally identifiable data—has no single right answer.

That's because privacy is not a single, static goal. Unlike the strength of a bridge, yaw of a yacht or recurring field in a relational database, privacy is a lock that opens with a combination, not a key. In designing systems to deliver on a commitment to privacy, the variables of time, place, platform and intended use are only a few of the constantly changing inputs that can overwhelm a more traditional, linear engineering approach.

Whereas a bridge needs to accommodate the weight of cars and trucks and a yacht must navigate the hull pressure of tide and wind and a database seeks to create order out of business chaos: privacy hopes to deliver on something even more challenging—the expectations of people. Worse, the need is to meet the expectations of people not in a group but as individuals. There is no engineering table for privacy.

This is guidance that can be drawn from former IBM executive Irving Wladawsky-Berger[29] who made a career of applying technology solutions to new classes of highly complex problems, "many based on disruptive innovations which we have not encountered before." Sound familiar? No development has been quite as disruptive as the Internet and the digital data stream it creates.

This is why many find "privacy by design" such a compelling concept. As described by the course catalog at Carnegie Mellon, which offers a privacy engineering degree, the emphasis of "privacy by design" is on "safeguards that can be incorporated into the design of systems and products from the very beginning of the development process."[30]

Rather than retrofit systems with data protection and privacy attributes, the notion is, "Wouldn't it be great to build *them* in at the start?" But is that the answer? The difficulty is in defining what "them" are.

To engineer a solution to meet the demands of consumers, business, and government for more transparent, informed, and value-driven use of data, we need to think holistically. Data protection and privacy engineering cannot only be about structured collection, hardened storage, authenticated access, and clear use, but must also accommodate the kind of variability that is human behavior.

[29] http://blog.irvingwb.com/
[30] http://www.cmu.edu/news/stories/archives/2012/october/oct15_privacymasters.html

We know that most existing systems running on data are designed as *point solutions*. The Internal Revenue Service (IRS) needs data to ensure the size of our refund, supermarkets need data to stock their shelves and colleges see social security numbers as a way to manage applicant and student files.

But what happens when the IRS can see new enforcement value in our data? What happens when the supermarkets see more value in selling our shopping preferences to advertisers? And what happens when college entrance decisions become based not on scores but on social context?

In each case the very best engineered solution to what had been the present problem did not have the ability to anticipate what insights might arise or the flexibility to cover whatever might come next—and something always comes next.

Data can now be connected, collated and queried in ways previously unimagined. The results can be of great benefit. But the dark cloud in all this is our inability to predict what stories our data will tell. This increases anxiety over who gets to tell them. Systems able to manage this uncertainty and unease are a tall order whose solution requires that we flip the engineering model on its head. To be blunt, we need to begin at the end rather than the beginning.

In many respects, the model for effective privacy engineering cannot be public works like the Hoover Dam, concerned with resistance, rebar strength and the heat of hydration of concrete, or a software program like Microsoft Word, built, in part, to correct spelling and grammar.

If the goal is to deliver on privacy, especially at the edge of the network as represented by the smartphone in the hands of its owner, a quite different model must emerge, one as fluid in its approach and design as it is hard and fast in its results. The one that comes most to mind was devised more than 60 years ago to solve the problem of the delivery of supplies to a constantly moving army. The result was the birth of Operations Research and an end to World War II.

As studied at Cornell University, Operations Research "deals with decisions involved in planning the efficient allocation of scarce resources to achieve stated goals and objectives under conditions of *uncertainty and over a span of time*."[31] That says pretty well what may be the best approach to delivering on a promise of privacy—contingent upon shifting variables of time, place, platform and purpose.

What is telling is that at the start Operations Research was not a single discipline but rather a matrix of many. The necessity of working together—manufacturing, transportation, topographic, finance and communications, to name a few—to solve a new problem on the battlefield may be the perfect metaphor for managing our privacy relationships today.

[31] www.orie.cornell.edu/news/spotlights.cfm?s_id=158

Although data are far from scarce, the usable insights they generate are exceptional; so much so that their pursuit has spawned whole new industries (e.g., Big Data). And if privacy, by its definition, is personal, then each transaction we make must be tagged at a different level of care and concern.

The implications for privacy engineering are as clear as they are counterintuitive. By focusing on the outcome of data use—less expensive health care, quicker oil and gas exploration, the most suitable advertising for an individual consumer—we can begin to design systems to be both focused *and* flexible.

Conclusion

Privacy engineering in the intelligence stage is crucial because information provided by or gathered about individuals often determines:

- What we build
- How we build it
- How it works
- How our customers use it
- How well it protects our customer or other persons involved
- The risks it may pose to our business and to future markets

Privacy engineering uses engineering principles and processes to build privacy controls and measures throughout system and data lifecycles (development, production, and retirement). Privacy is important to people impacted by the systems; privacy protection encourages trustworthiness and other factors that people expect when working with an enterprise or with an enterprise's systems. Privacy engineering will further assist in:

- Protection of customers and other people impacted by our systems and their data
- Improving trust by the people impacted by enterprises and their systems
- Developing secure and respectful computing that may drive more data sharing and engagement
- Gathering better information that will help create better tools
- Greater innovation and opportunity in the marketplace

All of these areas will be examined in this book. We begin our journey in Chapter 2 with a look at the foundation concepts and frameworks of privacy engineering.