CHAPTER THIRTEEN

Architectures of the Future: Building a Better World

We, as societies, are rapidly building the information architectures of the future. As we do so, much will turn on how interconnected these systems are and how their interoperability is managed. For instance, interoperability is at the core of the fast-growing social web, of Facebook and Google and those whole swaths of life mediated by networked mobile devices. The principle of interoperability is proving essential as we move our health records into electronic formats in search of better care at lower costs. It makes possible the air traffic control system that keeps travelers far safer than they would be without it. Interoperability is central to the development of sustainable global marketplaces of ideas, goods, and services. Interoperability is a key to the long-term preservation of the world's knowledge and heritage.

Three new architectures on the horizon—cloud computing, the smart grid, and the Internet of Things—show us why it is vital and urgent that we get interop right at the levels of both theory and practice. These systems also serve as targets to aim for; each of these emerging systems is designed to be part of the solution of some of the most pressing problems humankind faces. For each, interop is a key design consideration. The degree to which they are interoperable will determine, in no small part, how effective the systems are. Without enough interop, these systems will not come into being. If they are made to be too highly interoperable, new problems could arise. Each of these three emergent architectures will enable us to create a better world if we strike the right balance between seamless interconnectedness on the one hand and extensive amounts of friction on the other.

Interoperability is an essential factor in the development of these three developing complex systems. The cloud provides an essential part of the computing infrastructure of the future; individual users, entrepreneurs, companies small and large, and governments around the world will rely on it. The smart grid has the potential to solve the energy crisis that we face from San Francisco to Beijing and in every city in between. The Internet of Things might well establish a highly interconnected, data-generating universe that opens up new spaces for creativity, innovation, and experimentation.

All of these architectures of the future demonstrate how important it is to get interoperability right at the theoretical level. They also show how hard it can be to make the right design choices when it comes to implementation at the four layers of the interoperability model. These future-oriented examples demonstrate the importance of breaking down a series of barriers, from the technical to the cultural, to the establishment of meaningful interconnectedness. Finally, these examples make plain the importance of collaboration, over extended periods of time, between and among private-sector and government actors for getting interop right.

Cloud computing is the primary new infrastructure for computing and the Internet. In a cloud-computing environment, the basic functions

of computing remain the same. We still use computers to share information with one another socially and professionally, to book travel, to purchase new books, and so forth. In most cases, we as consumers do not notice much of a difference when our experiences online are powered by cloud computing. But a new kind of magic is at work behind the scenes. The way in which these tasks are carried out is changing rapidly in a cloud-based environment. These technological differences may seem subtle, but they are in fact profound. They involve and in turn lead to much higher levels of interoperability in many respects.

Cloud computing is the delivery of computing as a service rather than as a product. The difference between the previous computing paradigm and cloud computing is the idea that shared resources (such as data) and software are provided to computers and other devices as a metered service over a network. Typically, that network is the Internet. In a world of cloud computing, one would never need to get a disk in the mail to put into a personal computer. In a cloud-based infrastructure, all the information, computer code, and processing power is located on the network. The personal computer—or smartphone, for that matter—functions mostly as a simple device to access what is happening online.[1]

Cloud computing is a familiar concept, at one level. Many people already rely upon Google's Gmail service for their e-mail, use LinkedIn and Facebook to manage their contacts, and share files with one another through Dropbox and YouSendIt. These services demonstrate the basic idea of cloud computing. Any user who has a web browser and a network connection can access information he or she has stored online and can process it using technology that is hosted in the cloud. Very little needs to be kept locally on a computer or a smartphone. The functions are similar, but the architecture that supports them is different.

Although these changes for ordinary computer users are important in their own right, cloud computing brings with it more substantial changes for businesses and governments and for the way they handle information. In the context of businesses and governments, a more formal definition of cloud computing helps illuminate the changes underway. According to the

US National Institute of Standards and Technology (NIST), cloud computing is a "model for enabling convenient, on demand network access to a shared pool of configurable computing systems (networks, servers, storage, and so forth) that can be rapidly provisioned and released with minimal management effort or service provider interaction."[2] This formal definition is useful because it focuses attention on the way professional technologists handle their systems. Just as consumers might choose to keep all their e-mail on Google's servers and rely upon Gmail as an e-mail program, IT managers can provide more extensive services to their users while avoiding costly, time-consuming processes associated with managing hardware locally and downloading software to every workstation.

Cloud computing does not entirely change what businesses and governments can do using computers, but it makes things much cheaper, faster, and more efficient. For example, the city of Miami relied upon cloud computing to improve its 311 nonemergency phone line, which citizens use to report issues such as potholes or missed garbage collections directly to the local authorities. Using a cloud computing infrastructure, Miami developed a web application that enables users to track service requests online and to see other requests that have been made in the area.[3] The difference that cloud computing has made in a case like Miami's is that the city did not have to develop all the functionality on its own, nor did it have to invest in expensive infrastructure to make this new service run. The city of Miami only had to build an application that established an interface to cloud-based services on behalf of its citizens.

Cloud computing is also making a big difference for small companies, which can now launch ambitious services that they never could before. A Canadian entrepreneur, for instance, had the idea of creating a new type of online bookstore, one that mimics the shelves of a brick-and-mortar store. In this new online store, book covers are organized in the traditional way, where one can get lost for hours browsing books as in a real bookstore. After a few months of work, the entrepreneur launched the new online store.[4] In the cases of this new online bookstore and Miami's pothole-reporting application, cloud computing's networked infrastructure made the deployment of ambitious projects possible on a time line, and at a cost, that was

previously unthinkable. It was possible to create such systems before, but at costs that were prohibitive for many organizations.

As with most things in computing, matters get complicated very quickly. Cloud computing services come in many flavors, with acronyms such as SaaS (cloud software as a service), PaaS (cloud platform as a service), or IaaS (cloud infrastructure as a service). It is also possible to build both public and private clouds, as well as hybrids. The basic idea behind the fundamental shift in IT, however, is straightforward: in the same way that many of us are using web-based e-mail services such as Gmail or Yahoo! Mail, businesses can use cloud services to store and process their data and to develop and run applications. Businesses can also create entirely new services for their customers that connect to Gmail or Yahoo! Mail. For instance, software-as-a-service models allow for contacts in one program, such as Gmail, to be integrated with an online service like Salesforce, which salespeople use to track leads and monitor progress with customers.

The benefits of cloud computing are many—and they depend upon interoperability in order to be realized. For individuals, cloud-based services—ranging from Facebook (for social information) to Dropbox (for basic text files) and SoundCloud (for music)—are convenient, and they are often free to consumers. These services also provide a level of data security that would be hard, perhaps even cost-prohibitive, to achieve on a single personal laptop. For companies and governments, cloud computing is typically a matter of cost savings and time-to-market. Cloud computing offers the promise of low-cost, flexible access to scalable computing resources and enables institutions to outsource noncore activities to someone else's staff. The US federal government, for example, launched a landmark cloud computing initiative in 2010 that is expected to reduce its data center infrastructure expenditure by approximately 30 percent over the coming years.[5] If the systems involved do not have a high degree of interoperability at multiple levels, the government will not accomplish its cost-reduction goal.

Cloud computing is by its very nature all about interoperability. The entire point of cloud computing is that it allows for new levels of interconnectedness—at the network level, instead of at the level of personal computers. Systems that run in the cloud need to be able to interact seamlessly at the

data and the technology layers. For instance, in order for a contacts system in Gmail to be able to connect to Salesforce's software in the cloud, these systems need to be made highly interoperable with one another, either directly or through an intermediary. In the previous computing environment, the systems involved were vastly less interoperable. The computers on which users stored their contacts would not be connected to the computers that stored the sales data, processing power, and software to run the programs that supported the sales operation. In the cloud-based environment, the level of interconnection across the board is substantially higher.

As a consequence of these heightened levels of interconnection, this new computing architecture also brings with it new challenges. The list of concerns is long. In a world of cloud computing, will we know in which country the data we are using are hosted and stored? Which government has jurisdiction over data in the cloud? Will our existing safeguards prove strong enough to protect confidential data in the cloud? Who owns data in the cloud? Who can access those data? What happens after data in the cloud are deleted? The list of issues related to data location, privacy, security, and ownership goes on and on. Accountability and liability are major questions: who will pay, or how will costs be apportioned among multiple players, if a highly interconnected system breaks down? The answers to these questions will be especially important when various companies work together in the cloud in the course of providing services to end users. Transparency is also a major consideration: how will computer users know who is handling their data, and where they are located in the world in geographic terms, at any given moment? In this new, complex environment, policy makers worry about tracking information, not to mention liability, among various different clouds, which may be held in a mix of hands and across multiple jurisdictions.

To delve into the specifics of how to create and manage interoperability in the cloud, let's start with cloud-based social networking sites like Facebook, Orkut, LinkedIn, Ning, or Xing (a social network of businesspeople in Europe, with over 10 million members). These services are familiar ways that consumers use cloud computing to access personal and social information. Technical interoperability in this type of communication service typically has two dimensions: vertical interoperability (interoperability within a single platform) and horizontal interoperability (interoperability across different platforms). Despite being hosted in the cloud, many social networking sites are constructed to be deliberately noninteroperable along the horizontal dimension. In other words, they are designed not to work together with competing social networking sites.

There is very little horizontal interoperability across social network sites, which are holding important personal information in a series of private clouds. There is no simple way to move your account data—a feature called "data portability"—from one social networking site to another. However much you might want to, you simply cannot move your Facebook friends to Twitter or from Twitter to LinkedIn, and so forth. You would have to download each individual photo, link, and note and then upload them all over again. The logic behind these interoperability barriers goes back to the phenomenon of network effects. The greatest asset of companies such as Facebook and Twitter is the breadth of their user base and the quality of the "social graph" that this broad user base comprises. The more people use a particular service, the more likely others are to join. From a business perspective, it makes a lot of sense for these companies to lock in users by making it attractive to join but difficult to leave and start a new profile with a competitor.

The kind of interoperability we see in the cloud-based environments of Facebook, Orkut, Twitter, and LinkedIn is vertical, not horizontal. These services have built in a great deal of interoperability within the services they control but fairly little across competing services. For instance, they excel at implementing new ways for users to take content from other sources and rebroadcast it through their own services. Facebook recognizes when users include a website or video in status updates and pulls up thumbnails or even onsite portals so viewers can watch the video without leaving the page they are on. Twitter recognizes links and encourages shortening the links to fit its 140-character limit. Similarly, these companies want to maximize their user base and convenience by making themselves broadly

accessible, and so they develop apps that allow users to access their services on mobile phones, some even through SMS platforms. This interconnection is made much simpler and more extensive by the fact that these services reside entirely in the cloud.

Cloud-based services also support innovative ways to improve health care through better access to health information. Interoperability made possible by the cloud is at the heart of these improvements. For instance, Microsoft's HealthVault enables patients and their doctors to do a better job managing diseases like diabetes and heart disease. Patients at the Cleveland Clinic (a nonprofit medical center that performs care, research, and education functions) use at-home monitoring devices to measure data on glucose levels or heart rates; the data are uploaded to HealthVault and are then immediately available to the patients' health care providers, allowing them to follow the progress of their patients remotely and to draw on more comprehensive and accurate data when meeting with the patient at the clinic.[6] In a precloud environment, this exchange of information would be much slower and more expensive. The vertical interoperability within HealthVault and related services allows people in separate locations to work together on a common task, such caring for a patient's health. Neither the patient nor the doctor needs expensive computing equipment or software on their local machines; they just need a way to access the Internet. These cloud-based technologies have been well received by patients and hold much promise for improving medical treatment.[7]

These examples of health care and social networking show the effects of moving computing power and information to the cloud. In both cases, the cloud-based approach leads to a higher degree of interconnection among people and data. These examples illustrate how high levels of vertical interoperability can greatly benefit end users by linking together different devices, databases, and ultimately people. These examples also illustrate that interoperability in the cloud cannot be taken for granted. Interop in the cloud needs to be created by design and managed with care.

Cloud computing is just emerging as an important computing paradigm. The emerging challenges associated with cloud computing are close variants of the challenges we have explored throughout this book, yet they are heightened by the extensive degree of interoperability. For instance, cloud-computing service providers sometimes have strong business incentives to limit horizontal interoperability across platforms or services, much as Apple initially limited horizontal interoperability with its music services. Complex technical issues also limit certain kinds of interop. Proprietary data formats used by different types of cloud service providers may limit the extent to which data can be rendered useful from one part of the cloud to another. The lack of open and standardized infrastructure formats for data puts limits on data portability even in cases where services might choose to work together. Noninteroperable contracts (known as service level agreements) among cloud providers can also work against portability by setting different parameters for how an end user may interact with a cloud service and how that user's data ownership rights are governed.

Many of the interoperability challenges listed here—especially the technical and contractual issues—are best resolved by industry players. Current industry-led cloud initiatives, such as those hosted by the World Economic Forum and the Aspen Institute, are designed to resolve these exact problems by bringing together the big players in the cloud-service industry. This collaborative approach may well lead to new interoperability standards and common protocols that will help take advantage of the best aspects of cloud computing while mitigating the problems associated with it.

As in other interop settings, there is also an important role for governments to play. States can do more than merely stand by to intervene if market forces fail to resolve some of the important interop issues. Governments can serve as conveners and facilitators of standards-setting initiatives. NIST's important work—on definitional challenges, research, development of use cases, and reference architecture development—nicely illustrates how this government role can be played to benefit various stakeholders, including consumers. The US federal cloud-computing strategy, an important statement of the power of government procurement (which allots $20 billion in federal IT spending to the cloud), will shape the cloud-computing landscape. And finally, there is one challenge that

only governments can resolve. An additional layer of particularly persistent interoperability problems in the cloud—where data cross borders—stems from divergent national laws and regulations. Governments in the digital age must grapple with the essential task of ensuring that these institutions work together better than they do today.

The electrical grid is remarkably similar to cloud computing in terms of how it works. Both are, by design, highly interoperable complex systems that enable ordinary people and businesses alike to draw upon common resources to carry out everyday tasks. On the electrical grid, consumers demand and receive power without needing to understand the devices or infrastructure that deliver what they need. Interoperability enables the electrical grid to function seamlessly for vast numbers of consumers, who do not have to call the power company every time they want more electricity to flow to their home or business. Interop is also at the core of the next generation of electrical grids, which are expected to be much "smarter" than the current systems.

The electrical grid is among the most significant engineering achievements of the twentieth century. This complex network—power plants, transmission lines, and other components—that enables the generation of electricity and the transmission, distribution, and control of electrical power is itself a marvel of interoperability. The next phase of development of the grid is probably the construction of the "smart grid," which would be a vastly more efficient way of allocating energy and a boon to the environment. The development of this new layer to the grid involves creating new ways to share information between and among parties about the flow of power.

The development of the smart grid is one of the most important interoperability problems on the horizon. In addition to linking virtually all homes and buildings into a power network, the smart grid would establish high levels of interconnectedness in the data about energy demand and consumption. These new forms of interoperability bring with them the promise that the smart grid will operate more efficiently than the existing

power grid, but they also bring new problems, such as concerns about privacy and security.

Large parts of today's power grid are based on design principles and implementation choices that grew out of the first electrical networks and the technology available in 1900. Since then, the world has changed dramatically. The population in need of electricity, on a global basis, has grown exponentially. New sources of demand—especially due to the increasingly widespread use of computing and other electronic devices, such as televisions and radios—have put additional pressure on the grid. As grids have become increasingly large, interconnected, and international, their vulnerability has increased as well. The massive blackouts affecting millions of people in the United States over the past decade are important reminders of the enormous pressure that modern life puts on our largely outdated electrical infrastructure, even in the most highly developed nations. Many additional challenges, including security threats, lurk just underneath the surface.

Power companies and their regulators have launched large-scale initiatives to modernize today's power grids and to make them more reliable, efficient, and safe. Government regulators, as well as those who provide power around the world, are thinking hard about the capabilities that a modern grid must have. Among the top requirements is the ability of the network to "heal itself." The idea is that the grid can be configured so that it can deal automatically with problems such as power outages or service disruptions. Information from grid usage can be used to nudge consumers toward better behavior, so as to save energy and to allow the grid to operate more efficiently overall.

The smart grid is in its infancy as a technology. It is not widely deployed in a way that is obvious to consumers, but a great deal of planning and development is underway. The smart grid promises to deliver electricity from suppliers to consumers, but it will also offer built-in digital technologies to facilitate communication between them. Today's digital communication network makes it possible for sensing, measurement, and control devices to be made to interoperate. These devices can collect and pass information

about the condition of the grid among themselves, allowing the grid to respond dynamically to events that occur anywhere in the power generation, distribution, and demand chain. The smart grid can adjust the power flow in response to changes in the environment—for instance, by throttling down what each home or office can demand in very hot weather to avoid brownouts. The smart grid can also alter demand in positive ways. Consumers can make better choices about what they really need at peak times. The smart grid allows for dynamic pricing during peak periods of usage, making consumers smarter about the real costs of energy consumption. The system can also act dynamically to prevent systemic failure by temporarily shutting down a distribution line at crucial moments.

New, interoperable appliances can help consumers act in ways that protect the environment through energy conservation. The deployment and integration of smart consumer appliances and devices such as smart meters and smart thermostats, along with automated control of equipment, will help empower consumers to respond to changes in the grid and adjust their behavior accordingly. Before the smart grid came along, this option was only available to very large energy consumers—such as providers of cloud-computing services who need extraordinary amounts of power to run their massive systems. Under the new paradigm, users can allow the smart grid to turn off certain appliances, such as dishwashers or washing machines, during peak times to reduce demand and cut costs. The smart grid also enables decentralized sources of power—for instance, energy generated via solar panels on a house—to be fed back into the system. This feature becomes particularly useful for emerging building prototypes for structures that generate more energy than they use.[8]

The smart grid is not a substitute for the traditional power grid. It is an overlay, built on top of the ordinary electrical grid, made up of highly complex communication equipment and sophisticated metering system. Think of the smart grid as an "energy Internet" that integrates a number of different technologies and functions into one network of networks. Each of the technologies that is part of the smart grid (ranging from smart sensors to improved grid-level storage technologies) has positive effects on its own.

But when these components can all work together in a coordinated way, using all layers of interoperability, they create significant efficiency gains and will become a pillar of any future solution to the energy and climate crises we face. In some sense, interoperability on the smart grid *is* the smart grid.

The interop problems associated with the emerging smart grid are many and complex, and for the most part they are far from resolved. Thus far, only pieces of what will eventually be the smart grid exist. Many of the smart appliances that will support the grid are still just on engineers' drawing boards. Those that are in production are not in widespread use. There are still major capital costs to be borne by energy companies to support these appliances and the related information networks. In addition, all the necessary technologies and the ways in which they are supposed to interact across the different layers of the smart grid are today insufficiently standardized.

The first problem associated with interoperability and the smart grid is the basic fact that the smart grid does not yet exist. It is a fruitful example of an interop problem because it is on the drawing board, calls for high degrees of interconnectedness, but has not yet been built out in full. The levels of interoperability still need to be set, and many different people would like to have a say about them—certainly, the companies building the smart grid and their regulators have a role, but so too do consumers and those in civil society who look out for privacy concerns and issues related to public security. The whole set of problems associated with high levels of interoperability will arise. How can we ensure that data about our private activities in the home are not shared with the wrong people? How can we keep the smart grid from being hacked by terrorists? How can we ensure that technology that works in 2015 will still work in 2025, or is at least flexible enough to adapt, avoiding the problem of lock-in? How can companies and governments build smart grids that allow for—better yet, encourage—diversity and innovation?

The enormous scale of the smart grid poses special challenges for interoperability. The number of players involved is vast: customers (which means virtually everyone in a given society who is on the grid); utilities;

equipment designers and manufacturers; local, state, and national governments; and environmental groups have a stake in the outcome of the standardization process. Add in the international dimension, and the challenges multiply. In addition, a version of network effects works against interoperability. Efficient peak pricing is such an example, where the benefits only emerge when a certain number of people buy into the system and purchase smart measurement devices that report back accurate data. The delayed return on investment at the outset makes people reluctant to invest in this type of technology, which slows down the adoption of interoperable smart grid technology.

Once established, there will be drawbacks to interoperability on the smart grid. One major risk of broad interoperability based on standards-setting is technological lock-in. Standards-setters need to ensure that the rules established at the beginning of the smart grid's existence are stable enough to support development of the smart grid but malleable enough to support innovation over time. Security is a second potential drawback. A highly interoperable smart grid is more complex than the less-networked grid of the past, offering more points of vulnerability. Standards need to be set that will address the cybersecurity risks presented by full-scale implementation of the smart grid. And privacy concerns associated with the smart grid are paramount. In the extreme case, any switch of the light from on to off and vice versa would be tracked, reported, and analyzed. These data, if not properly safeguarded, could be misused by marketing firms, burglars, stalkers, and others who would do harm to users of the smart grid.

Whether in the United States, Europe, or Asia, governments are the major drivers of interoperability on the smart grid. They acknowledge the need and desire to create the smart grid. Everyone knows that we cannot enjoy the benefits promised by the smart grid without getting multiple aspects of interoperability right. But optimal levels of interoperability, across so many dimensions and involving so many actors, are not easy to achieve. These government actors also see the range of possible drawbacks associated with increasing the degree of interconnection on the grid.

The development of the smart grid in the United States is guided by a public-private effort with real promise. NIST, as in the case of cloud computing, is helping guide industrial development of this next-generation system. NIST is leading an industry-wide, collaborative, and, so far, constructive standards-setting process for the smart grid. A federal law—the Energy Independence and Security Act of 2007—tasked NIST with developing a framework for interoperability on the smart grid, and Congress has funded NIST to carry out this important work. In cooperation with the US Department of Energy, NIST has identified the key issues associated with interoperability on the smart grid. The primary goal is to establish the right standards to facilitate optimal levels of interoperability on the emerging smart grid. Smaller initiatives target key issues, such as cybersecurity and privacy, that must be addressed before the smart grid connects all our homes and businesses in new ways. The group has identified over a dozen standards as priorities, including standards for smart meter upgrades, common specifications for price and product definition, energy use information, and precision time synchronization.[9] The group has come to early agreement on the way information will be used to communicate between the utilities and the consumer and the way information is to be organized.[10] This collaborative, design-oriented model holds a great deal of promise for getting interoperability right as the smart grid comes online at scale in the United States.

This development of the smart grid illustrates why interop by design is so crucial. The drawbacks of smart grid interoperability have to be taken seriously and considered carefully at the outset of the process. These potential or actual costs have to be weighed against the benefits of what interoperability enables. In the case of the smart grid, interoperability is a constitutive principle; it is essential for the system to work. The smart grid is a pure interoperability case: the enormous benefits that the smart grid offers are the benefits of interoperability itself. Higher levels of interoperability can improve the reliability and efficiency of the electrical grid, reduce the price of electricity, create a platform on which new products and services can be developed, and promote environmental quality and renewable energies. On balance, the question is not whether or not we should have smart-grid interoperability but, rather, how we can work together to overcome the remaining barriers and deal proactively and responsibly with the potential drawbacks.

A third architecture of the future that depends heavily on interoperability is the Internet of Things. The IoT, as it is called, is an emerging information network that, counterintuitively, has to do with physical objects. The basic idea behind the IoT is that virtually every physical item—a razor blade, a bottle of water, a radiator, a chair, a car—can be turned into a type of tiny computer (a "smart object") and be connected to the Internet. Of the three emerging examples that we describe in this chapter, the IoT is the most speculative and the least certain to develop in a predictable fashion. The future of the IoT depends to a large extent on the question of whether we will be able to overcome the various interoperability barriers at many layers. The IoT is also a development that is more controversial than either cloud computing or the smart grid: it is not clear to many people whether in fact it should be built at all.

The benefits of an IoT could be substantial and widespread, but they are speculative enough to be hard to see. The world is awash in data. A networked universe of things, each connected to the Internet, would enable us to collect, aggregate, analyze, and use these data in unprecedented ways. An IoT could help improve our lives as patients; could help companies, markets, and governments work more efficiently; and, most important, could let us begin to address some of the most pressing societal challenges we face, including the efficient allocation of natural resources such as energy and water.

Interoperability is the DNA of the IoT. Interoperability issues are written all over it, ranging from the purely technical, as in the case of RFID standards, to the institutional, such as adequate legal safeguards for privacy. The approaches used to overcome these interoperability problems has to take into account drawbacks such as information overload, threats to privacy, and security concerns.

The IoT is made up of a universe of smart objects. The idea of smart things is not new, but it has only recently become possible to produce extremely small and low-cost networked computers that can be merged with physical things. Although the grand vision of the IoT, in which billions of things are connected with each other, is still a dream (or, for many people, a nightmare), the first instances of the IoT are beginning to come into view.

The tensions inherent in the IoT as a concept—its promise as well as the fears to which it gives rise—are revealed through an examination of a series of experiments. One such experiment is an unusual building in California. A team of visionary researchers at the Mobile and Environmental Media Lab at the University of Southern California asked, If a building could talk, what would it say? They wondered how a building might "feel" about the comings and goings of people, whether it could be affected by their moods and desires, and what kind of relationship it could have with its occupants if it could communicate with them.[11] The lab team framed the questions as an experimental design project, created a vision of a building, and identified a set of innovative technologies that could give answers to these questions. This is the birth of what is known as the Million Story Building project.

Using an actual building at the School of Cinematic Arts as a test environment, the lab team designed a series of location-specific interactions in the built environment and created an interface to the building by using mobile phones, sensor networks, and software applications. Through these technologies, the students, faculty, and staff of the school can, in effect, *interact* with the building on a daily basis and, in some sense, develop a relationship with it. The idea is to learn about the people in the building and what they are doing there. The building creates user profiles by aggregating data about its inhabitants, learning names, locations, and activities; in turn, the building can offer back to its visitors tailored information according to their perceived interests. Using movie clips, photos of different areas of the building, and other digital materials, the building introduces itself to its visitors via digital technologies—think of interactive applications on a smartphone that guide a user through the smart building. The effect is that visitors encounter a gamelike environment as they use the building. People are asked to complete more difficult tasks within the building over time, much as gamers perform harder and harder quests in video games. As inhabitants interact with the building and provide information

about themselves and what they are doing, the building records these activities as a digital archive—a living history of the new building.[12]

The concerns associated with the Million Story Building project are as easy to see as its attractions. Smart buildings, enabled by the IoT architecture, can help visitors find their way and move around efficiently, can perhaps even have personalized cups of coffee waiting for them at a kiosk near their workstations or classrooms. Perhaps smart buildings will interact with increasingly precise fitness applications, such as Fitbit (think of a networked pedometer for the social web), to help visitors stay fit as they go about their daily routines. The benefits for social and architectural historians are also plain: wouldn't we love to know how people in antiquity made their way through their built environments?

The same technologies that make these benefits possible—primarily, localized sensor-based networks—also give rise to Panopticonesque fears. We have already lost much of our privacy by recording our social lives online; the IoT might well lead to a similar deterioration of individual privacy in physical space as well.

Jails are experimenting with early implementations of the IoT.[13] A county jail in the United States has started using extensive RFID technology to track detainees and guards, enabling jail officials to understand better the interactions among them. On visiting a cell, a guard first scans a tag on the doorframe, which records his presence at the cell. The guard then scans an RFID wristband on the prisoner, which records the prisoner's identity along with the date and time. Additional information, including the reason for the visit, is recorded as well. The benefits associated with this experiment might extend beyond the obvious security implications. If theory holds, the effect of recording these real-space interactions may be to encourage more appropriate interactions on the part of both the guard and the detainee. But the well-known problems raised with Jeremy Bentham's Panopticon could be raised about this IoT-powered prison experiment, too.

Schools, too, may soon turn into highly networked computing systems. A research project is offering an East Coast school funding to experiment with RFID technology.[14] Early ideas for this project include simple measures, such as tracking school property—laptops and books in the library, for instance. More extreme versions include the tracking of students, for example, by giving them RFID-equipped backpacks. This measure could enable parents, school officials, or police to locate students quickly in case of an emergency. Some parents like the idea of their child's backpacks being trackable, much as they like the GPS functionality in the smartphones they give their kids at ever-younger ages. But these experiments, too, trigger serious privacy concerns and well-placed worries that children are losing the ability to grow and thrive while not being tracked at every moment.

The benefits of highly networked tracking equipment are easier to see in the case of health care than in the case of schools. Hospitals, for instance, are increasingly turning to IoT-powered technology to improve patient monitoring, automated medication, and the preservation of patient data. In hospitals in some developing countries, RFID tags for infants are used to prevent baby thefts. These examples also show the connection between two major interop stories: the IoT and the electronic health records example we explored in depth in Chapter 11.

IoT applications can serve trivial purposes as well as profound. The IoT has recently made its way into the coffee shop, for instance. The SMUG is a smart mug, outfitted with an RFID chip that enables personalized ordering and fast purchasing. SMUGs allow customers to communicate their coffee preferences and payment information to their favorite coffee shops—a step further than the Starbucks smartphone payment initiative that we mentioned in Chapter 3.

If you have made it this far into the book, you can no doubt see the extent to which the development of an Internet of Things depends on interoperability. And, in turn, as the IoT grows, the degree of interconnectedness and interoperability will increase. Unlike the Internet, however, the IoT brings the effects of high degrees of interoperability out of the digital realm and into the physical.[15]

It is not obvious how to build the IoT at scale, even if the fears associated with it were set aside. The physical world has constraints not present in the digital. In the context of the IoT, we once again have to contend seriously

with the geographic distances among trillions of potentially relevant physical objects. The characteristics of materials once again become relevant. And the topography of the surrounding environment again serves as a major constraint. This enormous diversity is a real challenge for establishing the IoT with high levels of interoperability, which is invariably context-specific.

These physical constraints are essential to the puzzle because the IoT requires object-to-object communication. Physical objects will need to transmit information, and many of them need to be able to "listen" and understand transmissions. In the IoT context, technical interoperability means that a signal can get from physical object A to physical object B. Semantic interoperability means that A (alarm clock) and B (coffee machine) can understand each other, that these objects "speak the same language." Except for the underlying wireless network systems that carry these signals, no single global standard has emerged that regulates interoperability—either technical or semantic—comprehensively.

Interop is hard to accomplish for the IoT at the technology and data layers, but it is even harder at the human and institutional layers of our model. Imagine the problems that occur if two hospitals, within a reasonable distance of each other, run different versions of the IoT, from different vendors, to track aspects of their patients' care. Even if each internal hospital system works seamlessly from a technical perspective, an interop gap will persist between the two organizations that will prevent them from working together in the most effective ways to help patients who move from one to the other. Even if the technological systems, as well as the organizational structures and processes, could be patched together in some ways, different cultural norms, legal requirements, and other higher-level interoperability barriers may stand between the two hospitals, their employees, and their patients.

These many challenges can be overcome, especially where a profit motive helps provide a driving force. The IoT is coming into being most quickly in the context of large-scale businesses that stand to benefit from tracking physical objects through networked technologies. The IoT is slowly but steadily coming into being in some industries, such as retail

and consumer goods. Here, certain emerging de facto standards—for instance, the EPCglobal standards, which are helping standardize item-level tagging—have been established for key components of the IoT system, especially with respect to RFID technology. The use of these standards will probably expand to related industries, including the textile and pharmaceutical industries.

The development of the IoT links back to the discussion of systemic efficiencies in Chapter 7. There, we discussed how bar codes have been used to improve inventory management, logistics, and accounting processes. Large retail companies, such as Walmart, have pushed their top suppliers to use RFID tags on all cases and pallets of consumer goods shipped to its distribution centers and stores. RFID technology has increased efficiencies throughout its supply chain. For instance, Walmart has been able to restock RFID-tagged items three times as fast as nontagged items. Walmart's use of RFIDs is an early example of the IoT serving a constructive, demonstrative purpose. Similar uses of basic IoT technologies across organizations and units have been reported in other industries, including the automotive industry, where RFID technology speeds up vehicle pickup and improves customer service.

And yet despite these early successes, the vision of the ubiquitous Internet of Things—a system that connects large parts of the physical world with the digital—remains primarily on drawing boards and in computer science laboratories. Some people may prefer it to remain there, merely a concept for experimentation in universities and corporate R & D facilities. The success or failure of the IoT in the long run, as well as its desirability, will depend largely on how interop is established and maintained.

Interoperability plays a crucial role as an enabler of these three emerging architectures of the future—cloud computing, the smart grid, and the Internet of Things. Interoperability in the cloud is essential from the user's perspective. The degree and the nature of interop are among the key factors determining whether cloud-based technologies will be adopted or distrusted in the long run. In the case of the smart grid, interoperability is absolutely

essential; in some sense, interoperability is the DNA of the new grid. And the vision and practice of the Internet of Things cannot even be contemplated without positing high degrees of interconnectedness among things and between physical and digital space.

Interoperability issues arise at all four levels of our layer model in these three emergent examples. The interoperability challenges at the technical level are significant across all three architectures; they range from issues of data formats to intricate aspects of semantic interoperability. But technical interoperability is not the only serious challenge. In all three cases, the way people and institutions develop and use the interoperability in these systems is just as important as how the data are designed to flow within and across them. Each of these cases also poses legal and policy issues, especially related to privacy and security, as more and more data flow across the boundaries of states in ways that are hard to track and manage.

Although each of these three emerging architectures raises problems, as a matter of substance, the potential benefits of increased interoperability in each case should ultimately outweigh the drawbacks. That is certainly true in the case of cloud computing and the smart grid; it is less obvious in the case of the IoT, which is both harder to envision and more controversial on its face. In the near future, societies will need to focus on how to manage the costs and benefits of the interop that is inevitably part of each of these three systems.

No matter what complex system we decide to build next to make this world a better place, it will require a shared commitment to increasing the interoperability of our systems, our institutions, and ourselves in productive ways. Whether at the technology, data, human, or institutional level, the optimal degree of interop will emerge in these three cases, and in others we can hardly imagine, only as a result of a massive collaborative effort. A broad group of stakeholders, from both the private and public sectors, will need to work together strategically, in good faith, over a long period of time to get interop right. As a matter of process, the principal drawback of this type of collaboration is that these processes take more time than ad hoc, private-sector innovation ordinarily does on its own. And these processes

are hard to pull off: they require deep trust, ongoing commitment to active engagement and openness, and a willingness by participants to set aside short-term gains in favor of shared long-term systemic improvements. The benefits of such large-scale collaboration, though, far outweigh these costs: stability as new systems come on line; efficiency and other immediate benefits for consumers and businesses alike, with well-managed downside risks; and sustained innovation in new emerging systems.

CONCLUSION

The Payoff of Interop as Theory

How are we to manage the unprecedented degree of interconnectivity that has been created between and among people and systems in the digital age? This is one of the most significant questions of our age. Much depends on our ability to maximize the benefits of this unparalleled and growing level of connection and information flow while minimizing its potential risks. We need to get interop right as a matter of public policy, as we address big issues like sustainability and climate change. Interop is also important in the private sector as a matter of strategy, in terms of helping businesses thrive and innovate. The theory developed in this book is designed to help consumers, business leaders, policy makers, and the public at large to make more informed—and ultimately, better—decisions about the ideal level of interconnectivity among complex systems and their components, about what we want to get out of interoperability,

and about the breakwaters that should be put in place to make sure it stays at the optimal level.

The theory of interoperability outlined here can be used in four ways: first, as a framing device and an organizing principle—in essence, as high-level theory; second, as a description, to guide us in our understanding of certain phenomena, mostly to do with information and technology, in the age in which we live; third, as an effort to predict what the future holds and what debates will surround the subject of interoperability in years to come; and finally, as a normative device, one that should drive and inform the kinds of decisions policy makers ought to make in order to lead to the kind of good societies in which we all wish to live.

INTEROP AS HIGH-LEVEL THEORY

The theory of interop that we develop and test throughout this book draws together a series of seemingly unrelated events, innovations, and themes in such a way as to establish unexpected and revealing patterns. What, for instance, do the global economic crisis that started in 2008, health care reform, global climate change, and the emergence of the social web and cloud computing have in common? All have interoperability at or near their core, what makes them possible and what can make them dangerous. The study of interop helps us see the promise and the perils of highly interconnected systems in our increasingly globalized economy through the similarities and differences among these widely ranging examples.

As a theoretical framework, the study of interop sheds light on what tends to go right and what can go wrong with complex systems that rely upon a constant exchange of information, most commonly mediated by digital and networked technologies. Although some of the interop stories included in this book—such as the evolution of emergency systems, shipping containers, and bar codes—predate today's digital era, they have important relevance for interop in the current age. The implications of this theory of interop are highly relevant for the next generation of complex systems. After all, it was not possible for information to flow as quickly or as consistently across organizational and national boundaries even a few

decades ago. Nor have people and materials been nearly as mobile and interconnected as they are today.

One of the key insights offered by interop theory is the degree to which the proper functioning of systems that seem to be predominantly technical in nature—such as air traffic control systems, cloud computing, or the smart grid—depends on how well human beings and institutions can work together. Over the past decade, much thought and money have been spent making information and communication technologies more robust and improving the systems that rely on them. It is crucial that we advance our technological know-how and practices to ensure that our data are safe and our privacy protected. But the theory of interop also highlights that we have to think equally hard about the appropriate design of the fragile interfaces where technology, data, human, and institutional layers intersect if we want to harness the benefits of the unprecedented interconnectivity in the future. Examples such as emergency communications and health care information teach important lessons about what has worked and what has not.

INTEROP AS DESCRIPTION

Interoperability research does not only lead to an abstract theory; it also helps at a precise, descriptive level. The careful study of interoperability helps explain specific phenomena in a complex world. An understanding of how interoperability functions in the context of case studies reveals much about what makes complex systems work well and what leads to their failure. Our methodology has been to explore case studies where we imagine interoperability might be part of the magic behind a system's functioning, for good or for ill. These case studies have taken us from the worlds of information technology, commerce, and trade to health care, emergency response, and the related fields of energy and environment. These case studies are posted freely on the web, at http://cyber.law.harvard.edu/interop, for anyone to read. These are the raw data and collected stories that we have worked from in the pages of this book; we have woven these narratives into the frame of our argument. They also stand alone as rich

descriptions of how complex systems function and of where they can break down.

These case studies describe connections that are hard to see on the surface but that are essential to the functioning of our complex world. A look beyond the surface of everyday phenomena—such as digital music, bar codes on products, instant messaging, and shipment containers (the boxes in which goods tend to flow around the world, on large ships and on trains)—encounters the hidden links and information channels among systems, components, and applications. It also discovers how much their capacity to work together depends on a complex set of choices, made over a long period of time, by a large number of players. These players have typically included technologists, consumers, companies, legislators, courts, and others. To make things more complicated, many of these decisions have been made in an ad hoc, decentralized fashion—certainly without any grand interop plan to guide the way. Given this decision-making process, it is surprising how well many of today's systems work together and how interoperable our world has become. At the same time, many of the case studies also illustrate how hard it is to undo bad decisions of the past. The legacy problem and the problem of path dependency (which we observed especially in the library and e-health contexts) are reminders of how important it is to think about interoperability in a proactive, strategic fashion.

Interop helps us understand issues related to globalization and how our cultures differ from one another. A global perspective, as we look forward, can help expose culturally specific approaches to interoperability. China, for instance, with its enormous market size, has a particular set of strategic interests with regard to interop. Chinese government and private companies are developing independent standards for certain information and communication technologies outside the realm of the international standardization organizations described in this book. Chinese officials have seen the development of their own standards as a matter of potential competitive advantage, both in security and in the marketplace. Officials in the United States are beginning to see standardization and interoperability issues in a similar light.

Such diverging regional interop approaches are also visible in a comparison of everyday experiences. Consider, for instance, the dissimilar ways in which we in Western countries and our friends in Asia deal with different electrical plugs. In China, the solution to this annoying interop problem is not an adapter but, rather, a pragmatic, multiplug design of the power outlet itself, built into the wall. Or take the example of a contactless, interoperable smart card, called Suica (the Super Urban Intelligent Card), that is used to pay the fare on trains in Japan. This nicely designed card works outside of trains, too; the Suica is increasingly accepted as a form of e-money for purchases in stores, at kiosks, and in taxis. Meanwhile, in the United States, we carry around wallets stuffed with different credit cards, swipe cards to allow us various forms of access, and separate customer loyalty cards from our drug store, our grocery store, and the place where we get our coffee in the morning.

These examples from Asia suggest another lesson from our research: interoperability, in virtually every context we have studied, is in constant flux and is occurring at differing rates around the globe. Rapid technological progress combined with highly dynamic market forces will continue to create new interoperability challenges and at the same time change the character of old problems. But the problem side of the equation is not the only thing in flux. The ways in which we address interoperability challenges may change over time as well, because we will learn from our own successes or failures and will also be inspired by different approaches from other parts of the world.

This theory and these case studies may be most immediately relevant to those who work in the industries and areas examined in the specific cases, such as computing and the web, libraries, and health care information systems. The implications are easiest to see in the context of information technology companies. The importance of an interoperability strategy is obvious to those who work at Apple, IBM, Microsoft, or Oracle, in the high-tech world. Increasingly, the next generation of big information technology companies are betting even more on strategies of interoperability: Facebook, Google, and Twitter are all building enormous businesses by

developing, and sharing wide access to, highly interoperable platforms. The same is true of companies all around the world, many in Europe and others in the fast-growing markets of East Asia.

But these issues are highly relevant to policy makers and consumers, too. The job of setting policy in the digital era increasingly calls for a deep understanding of interoperability and how it affects a broad range of legal and policy outcomes. It is an issue of competitiveness and of national security. And for consumers, the level of interop that people demand has a powerful effect on the decisions companies make as they design their products and services. Higher levels of interoperability can be great for consumers in terms of convenience, but it can also pose risks for security and privacy, as we have seen in the cases of Google's Buzz and Facebook's Beacon products.

INTEROP AS PREDICTION

Interoperability theory helps company executives and government policy makers by enabling them to make better predictions. The study of interop helps decision makers look ahead as they try to anticipate the results of their actions today. For instance, a large technology company may want to know whether it makes more sense to allow free access to and connection with core systems, opening them up to other developers (as Twitter and Facebook have done on the social web), or whether traditional strategies of exclusion are a better way to go. In the online world at least, the increasingly common answer seems to be that high levels of interoperability lead to better results for individual companies, for the industry at large, and for consumers.

But a well-designed interop strategy, as we have seen time and again, must also get the degree of interoperability right. It is essential to realize that high levels of interoperability can lead to further problems, often related to security and privacy, homogeneity, and lock-in. It is important to craft interop strategies that take advantage of what we know to be the major advantages of highly interconnected systems while working hard to design systems that mitigate its several potential downsides. Interop theory can help guide this design process.

Smart interop strategies adopted by tech companies, as well as sound interop choices made by users and regulators, will help harness the benefits of digital interconnectivity while avoiding its risks. But the most challenging interop problems often stem from the sheer complexity of the systems we want to make work together. For instance, it is very hard to envision what a successful interoperability strategy for the next generation of air traffic control systems will or should look like, because there are so many stakeholders around the world and so many different technologies involved. The same is true of international financial markets, where it is very hard to model the effects of the most highly interconnected systems and the most complex financial instruments. Viewed from this angle, our studies highlight the urgency and importance of sound interop strategies design to handle complexity at a global scale. Our theory demonstrates how users, companies, and governments should expect to come up against limits of how effectively we can predict outcomes in the most highly interoperable, complex environments—a major trade-off that we must realize we are making as we continue the process of deep interconnection.

INTEROP AS A NORMATIVE MATTER

Finally, the close study of interop helps determine what we, as societies, *ought* to do in certain circumstances. The study of interop can inform decision making about what the most promising approach might be to any given new interop problem. Interop theory helps us consider how we might solve the problems that we expect to face in the near future. The health care debate and the need to preserve human knowledge in a digital era, for instance, are two pressing issues that will require governments, companies, and consumers to have a firm understanding of interop issues. The emerging architectures of cloud computing, the smart grid, and the Internet of Things also present intricate interop problems that we, as societies, will need to address.

At a granular level, this emerging theory of interoperability provides a framework for sound interop policy making and puts forth a process-oriented model for policy makers who are seeking to address interoperability

problems that have arisen or are likely to arise. Most of the cases we have examined here are not straightforward instances of clear lawmaking; they tend to involve cultural and societal factors that shape the responses by governments, and vice versa. These factors may influence, for example, the instruments a government may use in addressing a given interop problem. To generalize, European lawmakers have appeared to be more inclined to regulate interop ex ante than their US counterparts, whereas US lawmakers have tended to rely on market forces up front and to turn eventually to corrective ex post mechanisms as needed. Several of the most recent examples that we have studied, including e-health and the smart grid, suggest a possible trend toward convergence between the US and European approaches. Increasingly, blended approaches, where public and private actors work together to establish optimal levels of interop, play an important role on both sides of the Atlantic. And, as demonstrated by our examples from Japan, China, and beyond, such approaches, in addition to innovative strategies, are emerging around the world.

The greatest payoff from the close study of interop ought to be the manner in which it guides our decision making on some of the biggest questions of our increasingly global, interconnected, digital world. It should push us, as individuals and as societies, to acknowledge and address the costs and benefits of deep interconnection among technologies, data, humans, and institutions. We need to understand, too, the implications of the failure of complex systems to work together in optimal fashion. Fundamentally, a deep understanding of interop will help us as we work together, across our many roles and functions in society, to fashion the kind of world in which we wish to live.