CHAPTER 2

Foundational Concepts and Frameworks

From within the secret court of men's hearts, Tom was a dead man the minute Mayella Ewell opened her mouth and screamed.

—Harper Lee, *To Kill a Mockingbird*, 1960

We cannot escape the secret courts within the hearts of men. Opinions, impressions, judgments and prejudices are formed, often instantly and subconsciously, based upon available data, context, and experience. The availability of greater and greater quantities of multimedia-enriched data makes more acute the imperative to manage and respect the power of information to impact individual lives as well as those of entire races and nation-states.

There's a terror in knowing what the world is about.

—David Bowie

This chapter addresses key definitions and concepts of privacy that anyone involved in engineering writ large (i.e., architecting, designing, developing, managing, and implementing components, products, services, processes, systems, or applications that process personal information) must understand to be successful as we enter a new stage in the Information Age—that of intelligence and data science. We also will define what privacy engineering is, what a privacy engineer does, and the goals of privacy engineering. In subsequent chapters, we will discuss how to apply these definitions and concepts to a privacy engineer's work, broadly defined as designing, creating, inventing, imagining, and building things that process personal information.

What Is Privacy?

A great majority of the complexity this book addresses arises, in fact, from the imperfections and difficulty of defining this multifaceted thing called privacy. There are different forms of privacy. *Data privacy* (also known as *data protection* in Europe), which is the kind of privacy this book addresses, can be discussed at great length, but finding one, global, consistent definition can be elusive. This chapter will propose an operational definition of data privacy as it is most often conceived by organizations that consume and process data about people and the governments and institutions who wish to regulate its many aspects and uses. This is not a book about public policy, philosophy, religion, or advocacy other than for privacy engineering.

Data privacy is one form of privacy that is derived from substantive privacy. *Substantive privacy* describes the right and ability of an individual to define and live his or her life in a self-determined fashion. Other forms of privacy attempt to describe and define this basic human fact. Data privacy is a derivative of the substantive right to privacy in that it is about data that has been created about an individual (1) by him- or herself, (2) by others through observations and analysis, or (3) by the consumption or processing (i.e., use) of that data about an individual by others.

Some of the other forms of privacy, or ways in which substantive privacy may be broken down, are behavioral privacy, decisional privacy, and physical privacy. They all interrelate and overlap in various ways. For simplicity sake, throughout this book, whenever we refer to privacy or data privacy we intend them as one and the same (i.e., data privacy) and if another form of privacy is intended, it will be identified.

THE DIFFERENT FORMS OF PRIVACY

There are different forms of privacy such as behavioral privacy, decisional privacy, and physical privacy.

Decisional privacy is really about being able to make decisions and choices without third-party inspection or intrusion. This may be thought of as self-determination within one's own private life. Not having to explain or justify one's behavior or share personal opinions or thoughts is an example of decisional privacy.

Behavioral privacy is about being able to act as one wants, free from unwanted third-party intrusion or observation (assuming no harm to others is incurred or laws broken). In this realm, people may dance in their living rooms or whistle in their cars or don various forms of dress or undress upon their own discretion.

Physical privacy is privacy about one's body or person. Modesty is another word for it. Some people are more sensitive to physical privacy than others.

Two things about the different forms of privacy should be noted. First, in many instances the examples overlap. Rarely is an example of one kind of privacy exclusive of another. Second, data privacy runs through all types of privacy because as soon as something about you or someone is observed or articulated (even just by you), you cantilever into the data privacy space. Data privacy is literally the language of substantive privacy forms whenever an action or behavior or even a stillness occurs. As such, as soon as any third party becomes involved in data that describe another person, data privacy becomes a fiduciary activity where access, sharing, or exchange of personal information is the corpus of the fiduciary trust.

THE SUBSTANTIVE NATURE OF PRIVACY

By Stewart Room, Partner, Field Fisher Waterhouse LLP

The right to privacy has been described in many different ways. US lawyers often talk about the Fourth Amendment prohibition against unreasonable search and seizures as protecting private spaces. European Human Rights law says that the right to privacy protects our home life, family life, and correspondence from unreasonable interference by the state. Legislation that is commonly grouped together as privacy laws has focused on the topics of health, financial services, children, electronic communications, and data security breaches. Famous court cases have protected the image rights of celebrities, the chassis of cars,[1] and office computers[2] all in the name of privacy. Statutory regulators use consumer laws to prevent the misselling of home closed-circuit television systems and smartphones as being privacy enhancing.[3]

Two golden threads run through this diverse list of interests, creating a common and uniting bond among them: the concepts of substantive and informational privacy. Within a civilized society, it is the desire to protect substantive and informational privacy that unites the celebrity, the child, the consumer, the smartphone, the camera, the home, the workplace, and the car. All theories of privacy and all privacy laws will pay service to one or both of these concepts.

The idea at the heart of the concept of substantive privacy is that people should be free to make decisions about how they lead their lives, free from interference by others. The idea at the heart of the concept of informational privacy is that people should be able to control the use of information about themselves. Within a state of privacy, these concepts reinforce and support each other; substantive privacy needs and relies upon informational privacy, and vice versa.

In this day and age it is readily appreciated that the threats to a person's privacy do not flow only from the state—the Identity Theft bogeyman is as much an icon for privacy interference as Big Brother—yet the example of the malevolent state provides the easiest way to demonstrate the relationship between and the concepts of substantive and informational privacy and their interdependencies. And among the many sickening examples of state-level evil that have plagued mankind and shamed our history, Hitler's Nazi regime in Germany stands among the very worst.

[1] *US v. Jones*, 565 US __, 132 S. Ct. 945 (2012).

[2] See, for example, *Copland v. United Kingdom*, 62617/00 [2007] ECHR 253 (3 April 2007). See also the UK Information Commissioner's "Employment Statutory Code of Practice" (2008).

[3] See, for example, *US Federal Trade Commission v. HTC*, File No. 122 3049 (2013).

The Jew in Hitler's Germany was required to wear a yellow star. This badge said publicly "I am a Jew." The information it conveyed restricted the Jew to the ghetto and, later, it destined him to the gas chamber. The evil Nazi state controlled the information, and the substantive effects will never be forgotten. Shortly after the end of the war, Europe adopted the Convention on Human Rights, ensuring the right to privacy for all persons, so that these horrors could not be repeated. Yet even in the modern world, states still interfere with informational privacy to substantively maligning effects. The Internet is intentionally tapped in North Korea and China to gain information about dissidents, which creates a general appreciation of the presence of surveillance and creates fear, which causes modifications to substantive actions, decisions, and the way people live their lives.

But why is any of this important to the privacy engineer? Simply put, remembering the very real connections between information and substantive actions and decisions creates a mental knot in the handkerchief of the mind (not to be glib about the use of information and the design of information processing systems). Often the substantive effects of information mishandling are hard to see, fathom, or articulate. The connection between a yellow star and a gas chamber is nonobvious. The harms or distress that may result from a security breach can also be nonobvious, likewise those resulting from data profiling, data aggregation, or data monetization. The privacy engineer will understand, however, that adherence to the principles and disciplines of engineering will provide the best prospects of understanding the substantive risks that can flow from the processing of personal information, and that engineering gives the best prospects for risk mitigation.

A captain of the industry has famously stated that the boundary between lawful data processing and unlawful interference with privacy is a "creepy line," a statement that for good or bad will sustain along with "the right to be let alone" within the lexicon of privacy. If the boundary between lawfulness and illegality is to creep and shift, the risk of unwelcome substantive effects becomes embedded within the organization. A risky business may accept this, but the privacy engineer who understands the connections between information and substantive privacy will understand the truth of this fascinating area; the boundary cannot creep and change, but should be fixed. This can only be achieved by coding the boundary into the architecture of the processing system.

Privacy Engineering

Too often the necessary controls and measures to protect personal information required by a process, application, or system are either ignored or bolted on at the 11th hour of development. When this happens, it usually results in poor user experience, with subpar protections, unnecessary overhead, and customer dissatisfaction.

This is not a wishful or hopeful book about the management of data centers or leadership. This is a practical and pragmatic book that charts out an approach allowing for innovation from many workbenches—legal, technical, political, artistic, or logical. We can call these disciplines, when they come together to create something that promotes the best of data privacy, the innovative and beneficial uses of personal information or those that chase out uncertainty and risk to data wherever possible: *privacy engineering*.

"Engineering" has been defined by the Engineers Council for Professional Development as the creative application of "scientific principles to design or develop structures, machines, apparatus, or manufacturing processes, or works utilizing them singly or in combination; or to construct or operate the same with full cognizance of their design; or to forecast their behavior under specific operating conditions; all as respects an intended function, economics of operation, and safety to life and property.[4]"

Privacy engineering as a discrete discipline or field of inquiry and innovation may be defined as using engineering principles and processes to build controls and measures into processes, systems, components, and products that enable the authorized, fair, and legitimate processing of personal information.

Privacy engineering may also be applied to the creative innovation process to manage increasingly more complex data streams and datasets that describe individual humans. Privacy engineering can be considered the gathering and application of privacy requirements with the same primacy as other traditional feature or process requirements and then incorporating, prioritizing, and addressing them at each stage of the development lifecycle, whether its for a process, project, product, system, application, or other.

The intent of privacy engineering is to close the gap between privacy policy and the reality of systems or technologies or processes. The greater the mismatch between the two, the greater the opportunity for needless inefficiencies, risk, or both.

The risk of failure to follow a privacy engineering approach will be discussed in greater detail in later chapters. In short, poor system design, poor policy requirement gathering, or poor communication (which are the hallmarks of design without privacy engineering techniques) may cause risk or harm to the inventors of such systems, the owners of them, and the individuals described or implicated by the data, or all of the above. Further, the monetary, reputational, organizational, or even criminal risks or harms will only increase for those who fail to recognize a privacy engineering approach as systems become more complex and personal data more valued.

Privacy engineering is not merely a call for mindful engineering where personal information is involved. The call for privacy engineering use and study is a call for leadership, innovation, and even a good measure of courage to change the status quo for design and information management.

Once every system owner, designer, and user expects and understands privacy engineering principles, we expect that privacy engineering will become so integrated into standard innovation cycles that there will be no need for reference to a discrete practice. Rather, the principles of privacy engineering will be an obvious and necessary part of engineering of any kind when personal information is involved or potentially involved.

[4] www.britannica.com/EBchecked/topic/187549/engineering

When privacy engineering becomes ubiquitous, individuals will not be treated as "inventory," and data about them will be viewed as a special asset, important, sometimes profitable, and always one with a fundamental ethical value. When this happens, systems that use personal information will be designed, implemented, and decommissioned accordingly.

However, to accelerate the arrival of this day and the ability to safely unlock the rewards of the Internet and the personal information service economy, there is an urgent need for leadership and for stakeholders to act expeditiously in adopting and extending the vision of privacy engineering as articulated throughout this book. Getting to privacy engineering ubiquity will require many acts of courage and cunning. But, as clearly articulated by Ford Prefect in Douglas Adams's *A Hitchhiker's Guide to the Galaxy*, "Don't Panic" and always carry a towel. Please consider this book your towel.

WHAT ARE THE "REAL" PRIVACY RISKS?

So far, most of the individuals who have gone to jail for data privacy violations have been hackers, spammers, identity thieves, and peeping toms. Unless related to large or multimillion dollar operations, most of the convictions do not receive wide-scale coverage in the mainstream media attention. It is the same with data breaches, which, unfortunately, are increasingly commonplace and thus less newsworthy.

But jail isn't the only possible repercussion for misbehaving in the privacy space and getting caught. Increasingly, corporations and organizations are being cited for privacy violations and are being fined, given sanctions, being placed under regulatory supervision, or pilloried in the public square of opinion. Some of these fines have been in the multimillion dollar range, required recoding of software and data deletions, resulted in multiyear sanctions requiring biannual privacy audits being submitted to regulatory authorities for review, or caused a decline in shareholder value.

We propose that privacy engineers take responsibility for:

- Designing and constructing processes, products, and systems with privacy in mind that appropriately collect or use personal information
- Supporting the development, implementation, and measurement of privacy policies, standards, guidelines, and rules
- Analyzing software and hardware designs and implementation from a privacy and user experience perspective
- Supporting privacy audits
- Working with other stakeholders to ensure privacy requirements are met outside as well as inside the engineering space

We propose that privacy engineers, in addition to better protecting and ensuring the proper use of personal information in the things they design, build, and implement, will provide the following benefits to individuals, as well as government and business enterprises:

- Protection for customers, users, or citizens
- A more objective basis for a trusted data platform
- A foundation to drive more thoughtful and higher-quality personal information services, sharing, and engagement

These benefits can lead to better and more information from users, which in turn helps to build and inspire better user experiences, better applications, better services, better products, and greater innovation.

Before we get into the toolbox for privacy engineering or the implications privacy engineering has for organizational design, let's explore some key privacy concepts and frameworks.

Personal Information

It is critical for privacy engineers to thoroughly understand how personal information is defined and how its definition evolves and shifts over time. Personal information is the asset protected by privacy rules, processes, and technologies. Traditionally, personal information has been defined as information that directly identifies or, in combination with other data, allows for the identification of an individual (i.e., basic examples are an individual's name, address, phone number, or national or tax identification number) or any otherwise-anonymous information that when combined can only be a single person. An example of this would be "the CPO of Sun Microsystems in 2005," because there is only one person who fits this description. An example of anonymous information would be "three of the thousand engineers carry laptops," because the characterization fits more than one person and, therefore, does not identify anyone in particular.

Traditionally, the term for these data elements has been *personally identifiable information* (PII) or, alternatively it could be called *personal information* (PI). Using different nomenclature can create unnecessary confusion due to unnecessary distinctions. The real issue is does the data alone, or in combination with other data, identify a single individual? The term PII is useful, however, in terms of determining which elements make a collection of information personal or identifying which data elements need to be removed to depersonalize or deidentify it. We will use PI as our convention throughout the rest of the book.

Some forms of PI are additionally considered "sensitive," either culturally, under the law, or both (e.g., the type of information that can be used to embarrass, harm, or discriminate against someone). Different cultures consider different categories of PI as sensitive PI, but the following are fairly common:

- Information about an individual's medical or health conditions
- Financial information
- Racial or ethnic origin

- Political opinions
- Religious or philosophical beliefs
- Trade union membership
- Sexual orientation
- Information related to offenses or criminal convictions

Largely due to the explosion of the Internet, mobile computing, and telecommunications technology, the definition of PI is evolving to include unique device and network identifiers such as the universally unique identifier (UUID) and Internet protocol (IP) addresses. The Federal Trade Commission effectively redefined PI to include certain types of what used to be considered machine data such as device ID and IP addresses when it stated in its 2010 report, "Protecting Consumer Privacy in an Era of Rapid Change," that:

> *the proposed framework is not limited to those who collect personally identifiable information ("PII"). Rather, it applies to those commercial entities that collect data that can be reasonably linked to a specific consumer, computer, or other device.*[5]

It should be noted that not all device IDs or IP addresses should be considered PI de facto. Some devices, just as some IP addresses, are not associated with an identifiable person or personal system.

HOW TO THINK ABOUT DEIDENTIFYING OR ANONYMIZING DATA

One way to remove risk or potential harm in processing personal information is to only use what is needed. One strategy for this is to deidentify or anonymize the data before using it.

Anonymizing or deidentifying data begins when deciding what to collect or use.

If personal information is not needed, then it is better not to collect or use it.

Always ask (1) is the information needed to serve the purpose of the processing; and (2) what is the minimum amount of information that is needed?

Example: Birth date: Is the day and month of birth needed or the actual birth data (day, month, year)? If the purpose is to automate birthday salutation, then month and date of birth should be sufficient. If the requirement is to ascertain age as part of authorizing access to content on a web site, just ask month and year, or age, or better yet, ask the age in ranges of 5 years.

[5] Federal Trade Commission, "Protecting Consumer Privacy in an Era of Rapid Change: Recommendations For Businesses and Policymakers," p. 43. www.ftc.gov/os/2012/03/120326privacyreport.pdf

Example: Geographic location: If the requirement is geographic, is GPS needed or will street address, ZIP code, or just city and state meet the need?

The second part of the discussion has to do with uses of the data. Some of the uses of the data may require the elements that make it personal information; others may not. So it becomes important to think about how to anonymize or deidentify data.

Does PI – P = I? In other words, if one removes the personal, is what is left just information? Well, technically yes; but this is something you may not want to be right merely on a technicality.

Consider the number of people in the data pool. For instance, although the information may be anonymous (because the personal identifiers have been removed), the data is still very distinct and the pool of possibilities so small that it might effectively reflect only three or four people. So, although the information does not truly identify a single person, the group is so small that an educated guess can easily be made as to whom is in it. You could say there are different levels of anonymization. One in 10 is different from one in 10,000.

Another vector to be considered is the methodology. How was the data anonymized? Were the unique identifiers removed completely from the dataset or were they merely replaced with a pseudonym?

If it was replaced with a pseudonym, does the pseudonym pass a reidentification test? Or can the data still be used to take action or contact a person? If it doesn't pass the reidentification test or it still can be used to contact a person or reasonably linked to a system, then it cannot be truly called "anonymized," perhaps deidentified, but not anonymized.

A third vector to consider is whether specific data elements are needed or whether ranges or categories suffice. In other words, using an executive income report as an example, one can remove name and titles, but even in large organizations, the actual income may be unique enough that it identifies an individual even though all other descriptors have been removed or genericized.

Finally, if the decision is to aggregate data, make sure it is anonymized as well. Aggregate data about a single individual is not necessarily anonymized.

Privacy

Merriam-Webster's Dictionary defines privacy as:

1. a: the quality or state of being apart from company or observation: seclusion

   b: freedom from unauthorized intrusion one's right to privacy

2. archaic: a place of seclusion

3. a: secrecy

   b: a private matter: secret

According to Yael Onn et al. in *Privacy in the Digital Environment*:

> *The right to privacy is our right to keep a domain around us, which includes all those things that are part of us, such as our body, home, thoughts, feelings, secrets, and identity. The right to privacy gives us the ability to choose which parts in this domain can be accessed by others, and to control the extent, manner, and timing of the use of those parts we choose to disclose.*[6]

Privacy defined colloquially seems to be subjective rather than systematic or governed by objective or pragmatic requirements; privacy is certainly contextual, including cultural and time-sensitive contexts that introduce variability and complexity. What one person may feel is the appropriate level of privacy can change, based on the situation. One person's sense of what is the appropriate privacy level for a given situation may be different from another's. Further complicating this is the fact that across the world, cultural values and social norms vary widely. Finally, the same person's notions and sensitivities may change over time and context, which is to say, what one may want to share at one point in his or her life may change as life progresses, just as it changes based on the environment.

Consider, as an example, the act of wearing a bathing suit. An office worker would probably feel that his or her sense of privacy was being violated if a condition of employment was to wear a bathing suit to work; but this is not so for a swimming pool lifeguard. External social and cultural norms would also be violated in the former instance (contextual). However, even for a lifeguard, the type and cut of bathing suit is a factor to acceptability, social normative value, and sense of well-being (subjective).

The challenge of privacy engineering is to architect and design products, processes, or systems that are sufficiently configurable to allow this sort of control.

An Operational Definition of Privacy

Data privacy may be defined as the authorized, fair, and legitimate processing of personal information. Much of the activity resulting from this functional definition will appear to focus on organizations' and the management's philosophies and policies from that

[6] Yael Onn et al., *Privacy in the Digital Environment*. Haifa Center of Law & Technology, 2005.

perspective, but it must always be remembered that the individual data subject—literally the *subject* matter of the information (i.e., the individual to whom the data applies)—remains the ultimate requirement-setting entity. To the extent feasible, flexibility built into privacy-engineered solutions will always be critical to properly govern that very human variability. Note, too, that it is not always possible to make everyone happy.

Although this operational definition may seem deceptively simple, we can break it down into its components to start to see this definition as the beginnings of a pragmatic framework to not only define data privacy but also to begin to build it from these foundations (Figure 2-1).

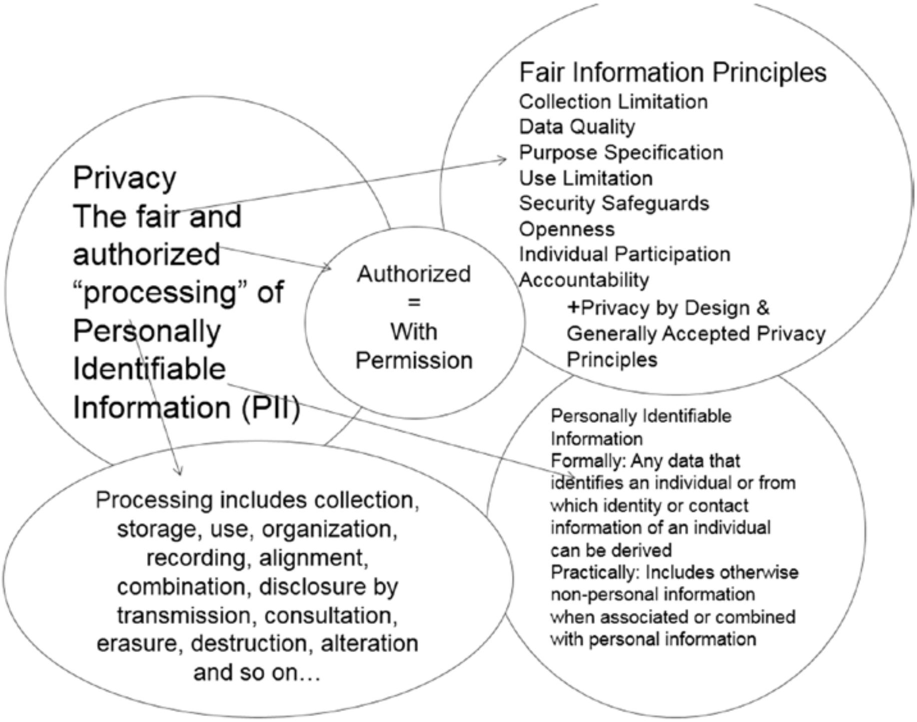
Privacy
The fair and authorized "processing" of Personally Identifiable Information (PII)
Authorized = With Permission
Fair Information Principles
Collection Limitation
Data Quality
Purpose Specification
Use Limitation
Security Safeguards
Openness
Individual Participation
Accountability
+Privacy by Design & Generally Accepted Privacy Principles
Processing includes collection, storage, use, organization, recording, alignment, combination, disclosure by transmission, consultation, erasure, destruction, alteration and so on...
Personally Identifiable Information
Formally: Any data that identifies an individual or from which identity or contact information of an individual can be derived
Practically: Includes otherwise non-personal information when associated or combined with personal information

Figure 2-1. *What is privacy?*

We have already discussed and defined personal information, so now let's turn to what is meant by processing, authorized, and fair and legitimate.

Processing of Personal Information

Data is processed upon any action or inaction that can be performed in relation to that data or dataset. Processing personal information includes, but is not limited to, collection, storage, use, sharing, organization, display, recording, alignment, combination, disclosure by transmission, copying, consultation, erasure, destruction, and alteration of personally identifiable information and any data related to it.

DOES THE USE OF DATA FIT WITHIN A CULTURAL CONTEXT

By Martin Abrams, Executive Director and Chief Strategist for the Information Accountability Foundation

The slogan “Keep Austin Weird” works really well for that swinging Texas city, but the culture in Hebron, Texas, would likely not be associated with “weird,” at least not in the same way. Local cultures are reflected in the way people interact with people. And privacy is one of those areas where culture is reflected.

Privacy scholars such as Alan Westin who established the basis for modern privacy management understood that privacy culture is a function of how a society balances the autonomy of the individual against the interests of the group, and then factors in the way a society defines a space reserved for the individual, free from observation from others. Although residents of both Hebron and Austin might have similar views on concepts of space, the balance between individual expression and community cohesiveness would be very different. Understanding cultural diversity and applying it to privacy is difficult enough when making decisions about what is an appropriate use in Texas, now think about looking at a global program that needs to work in Germany, Japan, weird Austin, and stern Hebron. How does an engineer begin building application requirements that fit the cultural context of diverse populations?

Let’s use an example. Millions of smartphones are sold each year in places as diverse as Galesburg, Illinois; Bangalore, India; and Frankfurt, Germany. Each smartphone has a unique signature, just like each of us have distinct finger prints. All smartphones are designed to run on Wi-Fi networks. This design factor saves consumers money on their monthly mobile bills. It is no surprise that most consumers want to save money, so they set their phone to look for available Wi-Fi networks.

An innovative engineer quickly figured out that one can track a device through a physical space like a store by equipping the space with Wi-Fi. Furthermore, the engineer can see how much time the individual spends within a physical quadrant and can then link that information to the activities that take place in that quadrant. If it is a store, the activity is most likely shopping. For example, if the mobile device is in a home improvement store, the engineer now knows how long the device spends in the paint department and when it moves from paint to window treatments. Maybe he or she can even link the shopping activity to the items purchased and track what the device buys over time. It’s not the device that buys the item, it is actually the individual holding the device; while the device might not have a cultural perspective, the individual does. It really doesn’t make any difference whether we know the name of the individual. The actions we take based on tracking the device are particular to that individual. So the privacy question becomes: Is it appropriate to take actions based on the predicted behavior of the individual holding the device?

The answer is: It depends.

In the United States we have many conflicting values. First and foremost, we believe that we are free to observe what we are free to see and hear within the public commons. In the physical world, we, as a society, have defined the public commons: Pretty much, it is anything outside one's home. It is the public street, the shopping mall, front yard, and the courtyard, if one is flying over in an airplane. Furthermore, we are free to express ourselves based on how we process what we have observed. Making a sales offer is a form of expression. This value is captured by the First Amendment to the US Constitution.

The American people also cherish seclusion. That means, in our private space, we are free to do what we will do and think what we will think without fear of others observing and using what they hear and see. Our home is our castle, and it is not part of the public commons. You may watch me in my front yard, but you may not look in my window and invade my seclusion.

In the United States, the Wi-Fi-enabled store is the public commons. The observation of a device in a public space is probably okay, even if some might consider it obnoxious. Furthermore, we are free to think about what we have learned and apply that knowledge for practical ends such as increasing sales.

The preeminent nature of observation based on free expression doesn't have the same deference in other cultures. In those cultures, the sense that privacy as a fundamental right trumps the recording of what we observe and making use of that information. This is particularly so for most other Western cultures. In Germany or France, the collection of the device signature, if it is easily linkable to an identifiable individual, is probably subject to data protection law. Such a collection would be a processing of personal information that requires either permission from the individual or the law. Furthermore, any additional processing of that information, even storage, would also require permission from the law or the individual. We are talking about the same activity in different locations and having two different takes on whether the use is appropriate.

US culture puts a premium on free observation in the public commons, while societies with traditional data protection have no such deference for free observation. So, if an engineering team were to develop an observation model for a client that is dependent on observing devices in a physical space, the application would probably work in US stores but would be a violation of both societal norms and laws in stores in Western Europe. The analysis might be entirely different in Asia, where rights to seclusion are limited but where such observation might be seen as violating norms necessary for a crowded society where physical space is limited. The laws are different because the cultures are different.

These differences in privacy culture have impacted digital public policy for more than 30 years. Justice Michael Kirby, former chief justice of the Australia High Court, led the experts that developed the Organisation for Economic Co-operation and Development (OECD) Privacy Guidelines between 1978 and 1980. He said the most difficult issue he had to overcome in leading that group was the huge deference Americans give to free expression. Even though these differences are understood, we tend to default to what feels comfortable to each of us. Business concepts based on monetizing the fruits of observation have been developed in the United States, but when the same applications are applied outside the United States, we tend to see friction.

Ninety percent of the privacy issues that concern both individuals and regulators are the same no matter where the activity takes place. These include ensuring security, accommodating transparency, and not facilitating illegal behavior by others. If one deals with these issues, one can have a fairly high level of certainty that an application is okay. Moving beyond what is the same, one can anticipate key cultural markers. One such marker is at what age an individual reaches the age of maturity. This influences the consent children and adults are able to grant.

Lastly, one needs to be truly sensitive to cultural differences related to observation. You know when the technology tracks behavior, so tracking is an indicator that a cultural review is necessary when a technology is taken from one geographic market to another. Such applications probably require a privacy impact assessment (discussed in Chapter 10) with experts who understand the cultural frame. Lastly, there are cultural aspects to automated decision making. If applications make decisions without human intervention that impact the ability of an individual to gain employment, get credit or insurance, or travel, one should check cultural norms related to such decision making.

Just be sensitive to the fact that what is appropriate where you are doesn't mean it will be appropriate somewhere else; if you keep this in mind, you should be successful in your data-use initiatives.

Authorized

Authorized processing of personal information only happens where the person or organization processing it has appropriate privilege for that processing. Additionally, there is a chain of custody and a sense of fiduciary responsibility that must follow the PI throughout the lifecycle of its processing. For example, those who can access a system containing PI must be authenticated to be the person he or she claims to be and that individual must also be acting within a role that would allow him or her to process the data within a system.

The type of data, the nature of the processing, as well as local laws and regulations will determine the nature and level of permission that may be required. The four primary protocols for permission gathering are:

- Opt out/Opt in
- Implied consent
- Informed consent
- Express consent

Opt out allows processing of PI unless or until an individual rejects data processing according to the context at hand. *Opt in* (the logical twin to opt out) is where no processing is allowed unless and until permission is granted. These concepts are relatively new in the comparative areas under the law, as discussed below, particularly in Common Law jurisdictions.

Context, narrowness of purpose, and transparency practices can make opt out or opt in relatively effective mechanisms.

Implied consent is a relatively straightforward concept where the context of collection and other processing is deemed so routine, obvious, and expected that permission for processing within this context may be implied by an individual's participation in the contextual relationship at all. An example of implied consent would be when PI is used for necessary processes (business or otherwise). When you give your name and telephone number for a reservation, the permission to use it to hold your table and for the maître'd to use it to call you is implied because it is necessary and within the scope of the function for which it is being used. However, if the maître'd chose to send text messages to the reservation number to solicit charitable donations to his favorite charity, he would be violating the implied consent to use contact information.

Informed consent relates to a very well-established and understood area of contract and tort law where a data subject has all relevant and timely facts to enable a reasonable choice of whether, how, how much, and for what purpose data will be processed. A good example of well-informed consent in a nondata context is the difference between giving consent or accepting the risks of skiing vs. receiving medical treatment from a trained doctor. In the former example, an individual is physically aware of his condition, standing on a snowy mountain, on two small skis. Yet there may be unexpected risks, and thus a disclaimer may be written on his ticket, but that disclaimer may be in smaller type and with no individualized explanations. In the latter example, however, the doctor and patient have very different levels of expertise, the procedures and risks may be unfathomable to the reasonable layperson, and the side effects may be unknowable without specific clarity. The type and depth of disclaimer and expository of risks and rewards are much different and far more extensive in this case.

Informed consent requires some responsibility and action on the part of the data subject and so may never become universally accepted as the standard for gaining or maintaining authorization, but its longevity in other fields of risk management and conflict resolution and the various aspects that allow breaking informed consent into measurable components make this form of consent particularly attractive to the budding privacy engineer.

Express consent is simply where a person takes a specific observable action to indicate and confirm that they give permission for their information to be processed. An example of this is checking a box that says, “yes” on an online form.

So that it does not go unrecognized, express consent and informed consent are both subspecies of the opt in.

The strength and validity of any of these permission forms and types depend on the clarity, conspicuousness, and proximity of data processing intended to be governed by authorization. It must be clear that the user knew what was being accepted to make the permission valid when permission was granted. Similarly, permission must be freely given and not under duress for data processing to be authorized to the appropriate degree.

The other key ingredient is, for all these different forms of permission, they must be presented before personal information is collected and before it is processed. For example, there has been much debate about the ability for web site operators to use cookies on the first page of a web site where notice is presented about the possibility of data collection through electronic means. In fact, the difficulty in ensuring that data subjects know and understand the potential and actuality of data privacy in a clear, conspicuous, and proximate fashion is one of the many reasons that those processing the data, governing bodies, and users are skeptical that a governance and enforcement regime focused on “Notice and Consent” is effective in today’s data-enriched environment.

Permission is only one component of ensuring that PI is processed with authorization. In addition to ensuring that one has permission to use the data, one also has to be able to manage and prevent unauthorized use or access to the data. This requires using controls and measures to ensure PI and related data is processed in an authorized and legitimate manner. These controls and measures can take the form of administrative, logical, technical, and physical routines or a combination of all of these, which will be discussed later in this chapter and in Chapter 6.

THE EVOLUTION OF CONSENT

By Eduardo Ustaran, Data Protection Lawyer and author of *The Future of Privacy*

Is individual choice still the essence of data privacy law? In the early days of data protection as a regulated activity, putting people in control of their information was thought to be what mattered the most. From the 1980 OECD Guidelines to the latest version of the EU e-privacy directive, consent has been a cornerstone across legal regimes and jurisdictions. European data protection law is based on the principle that an individual’s consent is the most legitimate of all legitimate grounds to use information about people. But does this approach still hold true? Can we—as individuals—attempt to have a meaningful degree of control over the vast amount of information we generate as we go about our lives?

Information about who we are, what we do, what we are like, and how we behave is captured every single second of the day. From the moment we turn on the light (or the Blackberry or our smartphone) in the morning to the moment we turn it off in the evening, every action that involves using technology is recorded somewhere.

The Internet has maximized this in such an unprecedented way that the value of the information we generate by simply using it makes other more traditional identifying factors look trivial. From a legal perspective, this phenomenon has entirely distorted the meaning and scope of personal data, but the point is that information about us is constantly flowing around the world without our knowledge, let alone our consent.

Let's face it, attempting to put people in control of their own information by giving them the power to consent to the uses made by others is simply unachievable. The concept of consent should not be underestimated. The ability to make choices is what makes people free. However, pretending that we can take a view in any meaningful way as to how information about us is gathered, shared, and used by others is wishful thinking. We cannot even attempt to recognize what personal information is being made available by us in our daily comings and goings, so how could we possibly decide whether to consent or not to every possible use of that information? Consent might have been a valid mechanism to control data handling activities in the past, but not anymore.

So what now? Is data privacy dead? I hope not. But in the same way that our ability to control our own information is moving away from us, our responsibility to decide what others can know about us is also receding. Our privacy is less than ever in our own hands because the decision-making power is not really ours. Any legal regime that puts the onus on individuals (who are meant to be protected by that regime) is bound to be wrong. The onus should not be on us to decide whether a cookie may reside in our computer when hardly anyone in the real world knows what a cookie does. What the law should really do is put the onus on those who want to exploit our information by assigning different conditions to different degrees of usage, leaving consent to the very few situations where it can be truly meaningful.

The law should regulate data users, not data subjects. Like it or not, individuals have a limited role in the data-handling decision-making process. That is a fact, and regulation should face up to that fact. Technology is more and more complex, while our human ability to decide remains static. Feeding us with more detailed and complex privacy policies does not change that. In the crucial task of protecting our personal information and our privacy, consent can only have a residual role. Continuing to give consent a central role in the protection of our privacy is not only unrealistic, but also dangerous because it becomes an unhelpful distraction for individuals, organizations, and regulators. The emphasis must simply be put elsewhere.

Fair and Legitimate

Of all the concepts that underpin the notion of data privacy, the ability to provide information handling that is "fair and legitimate" is probably the most complex and difficult to reduce to a scientific rule or even an approximate measurable metric.

The concept of *fair and legitimate* processing is not limited to the organizational view of fair as "necessary" (or, more often, "desired") processing. However, a series of principles called the Fair Information Practice Principles (FIPPs)as embraced by the OECD in the OECD Privacy Guidelines, is a useful prism through which to look at the notion of fairness and legitimacy.

Fair Information Processing Principles and the OECD Guidelines

The original FIPPs were developed by the Department of Health, Education, and Welfare in the 1960s in reaction to and concerns over implementation of large government databases containing information on US citizens. As mentioned earlier, the principles were then extended by the OECD in 1980 in a document titled "The OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data."[7] These principles, commonly know as the OECD Principles, have since become the foundation for much of the existing privacy legislation and thinking throughout the world. More important, they continue to be a cornerstone in grounding governments, businesses, and consumer advocates in their approach and dialogues on privacy and the use of personal information. In other words, they form the common vocabulary in which privacy is discussed. As we detail later in this chapter and elsewhere in Part 2, most privacy laws and regulations (and thus privacy policies and the privacy rules) are derived from the FIPPs and the OECD Guidelines.

Collection Limitation Principle

The OECD Guidelines, published in 1980, state that "There should be limits to the collection of personal data and any such data should be obtained by lawful and fair means and, where appropriate, with the knowledge or consent of the data subject."[8] This means before PI is collected or processed in another fashion, the processor must obtain permission to process the data. There are rare exceptions to this requirement, including certain types of law enforcement practices and for "national security" purposes.[9]

Given the increasing reality of law enforcement requests and requirements from around the world, it is imperative that privacy engineers contemplate such uses and their potential conflict with the "Collection Limitation" principle for their processing.

[7] An outgrowth of the Organisation for European Economic Cooperation (OEEC), which was formed in 1948 and chartered to run the Marshall Plan, the OECD, established in 1961, consists of 34 countries who work collaboratively to "to help governments foster prosperity and fight poverty through economic growth and financial stability."

[8] The OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data. http://www.oecd.org/internet/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm#part2. All quotes from the OECD Guidelines come from this source.

[9] Even those cases are not consistent from jurisdiction to jurisdiction and, in those cases, there must be other control processes in place to ensure that individual rights are not being violated and that the data is collected in a manner that allows law enforcement to use them for policing or security.

Data Quality Principle

From the OECD Guidelines: “Personal data should be relevant to the purposes for which they are to be used, and, to the extent necessary for those purposes, should be accurate, complete, and kept up-to-date.”

There are two key ideas in this principle. The first is “relevancy” (i.e., the data collected/used must be genuinely pertinent to the purpose and proportional, that is, only the appropriate amount and type of data to suit the purpose for its collection or processing). The second idea is accuracy. This is important because it creates obligations on behalf of the entity that controls the data to ensure data integrity. This requirement has evolved to also require giving data owners the ability to access their data and correct or update any errors.

It should be noted that data “integrity” is one of the core principles and goals for the security practitioner as well. For security, confidentiality, integrity, and availability are key markers for success and planning security requirements. Throughout this book we will note where synergies and common goals exist such as the case of *data integrity*. In doing so, the building and maintenance requirements for privacy engineers should be viewed as additive to other requirements rather than competing or negating “compliance” post facto requirements.

Purpose Specification Principle

From the OECD Guidelines: “The purposes for which personal data are collected should be specified not later than at the time of data collection and the subsequent use limited to the fulfilment of those purposes or such others as are not incompatible with those purposes and as are specified on each occasion of change of purpose.”

This principle provides guidance regarding the type and quality of transparency or notice. From an innovator’s perspective, creators of systems or services should carefully consider how PI will be used throughout the lifecycle of the current situation and should plan ahead as carefully and fully as possible to ensure that enough flexibility for data processing is introduced into the system and any contextual cues, including notice leading to transparency and understanding of data use.

Use Limitation Principle

From the OECD Guidelines: “Personal data should not be disclosed, made available or otherwise used for purposes other than those specified in accordance with Paragraph 9 [Purpose Specification Principle] except:

a) With the consent of the data subject; or

b) By the authority of law.”

This principle qualifies both the limits for data processing and the expectations of the data subject and also suggests conditions for potentially adding to the type, kind, and timing of data processing when that processing was not included in the initial authorization. As discussed previously, some legal enforcement should be contemplated and presented in the original “Purpose Specification of the Notice.”

Security Safeguards Principle

From the OECD Guidelines: "Personal data should be protected by reasonable security safeguards against such risks as loss or unauthorised access, destruction, use, modification or disclosure of data."

Any entity controlling PI must protect it from unauthorized access or processing. This principle clearly invokes the wide and complicated discipline of security for all types of data but focuses the requirement to specifically protect personal data. This is one of the overlaps between privacy and security that will be discussed later in this chapter.

Openness Principle

From the OECD Guidelines: "There should be a general policy of openness about developments, practices and policies with respect to personal data. Means should be readily available of establishing the existence and nature of personal data, and the main purposes of their use, as well as the identity and usual residence of the data controller."

Publication of privacy policies and statements is one means to achieve a level of openness in and about an organization.

Individual Participation Principle

From the OECD Guidelines: "An individual should have the right:

> a) to obtain from a data controller,[10] or otherwise, confirmation of whether or not the data controller has data relating to him;
>
> b) to have communicated to him, data relating to him within a reasonable time; at a charge, if any, that is not excessive; in a reasonable manner; and in a form that is readily intelligible to him;
>
> c) to be given reasons if a request made under subparagraphs(a) and (b) is denied, and to be able to challenge such denial; and
>
> d) to challenge data relating to him and, if the challenge is successful to have the data erased, rectified, completed, or amended."

This principle describes an individual's right to update, correct, and know which data has been collected about them from a given entity. It is closely related to the accuracy principle. Much innovation is required for this principle, in particular in a world of vastly dispersed and complex data sharing and processing even to achieve relatively simple goals.

[10] Author note: A data controller is the entity that is responsible for determining how data is processed. The data controller gives direction to the data processor. Sometimes the data controller and data processor are one in the same; sometimes not, such as in outsourcing. In such as situation, the service provider is the data processor.

An example of some of this complexity may be the fulfillment of an online contact lens service where an individual may be described by a common carrier, an ophthalmologist, a fulfillment center, a manufacturer, and more. For any one individual to possibly glean where and when his data changes hands among all of these specialized and related steps is a daunting task indeed.

A PRACTICAL APPROACH TO DETERMINING IF DATA COLLECTION AND USE IS "FAIR AND LEGITIMATE"

Here is a two-tiered process to determine if data is needed.

The first tier is to ask the question. Is this data needed? Not wanted, needed.

If they answer is yes and all other design and architectural reviews and options (such as not collecting at all, truncating or de-identifying the data) have been exhausted, then run each data element through the following set of formulas:

I need X to do Y

Without X I cannot do Y.

If the answer to the first two equations is true, proceed to the third:

Y is a subset of uses for the data for which Z has given permission (Y < ?).

If the answer to this equation is true, then ask, does it pass the smell test (fit the spirit of the permission, as well as the letter). If the answer to this is yes, then proceed.

If the answer is no, then based on the data and the use (i.e., the risk), explore what level and type of notice and consent are required and consider who best to expand the existing permission to cover the contemplated use.

If there is reluctance to go back to an individual for permission, then someone has to ask what is the locus of that discomfort. It usually is because the benefit is not so much for the person but for the organization or because there is a lack of proportionality between the risk to the privacy of the individual vs. the benefit to him or her. Knowledge of this will help the real goals and purpose of the processing to surface, which will then lead to a more productive discussion of how to address and manage the risks.

Accountability Principle

From the OECD Guidelines: "A data controller should be accountable for complying with measures which give effect to the principles stated above."

This principle means whomever is controlling the data, that is, in charge of determining how they are going to be used and processed, is the party who will be held responsible for ensuring the data is processed in an authorized and fair and legitimate manner and will bear the consequences if they are not.

THE INTERSECTION OF PRIVACY, UNIQUE IDENTIFIERS, AND COLLECTING TELEMETRY

Telemetry is the collection of information about machines and systems. It is often collected remotely to monitor how a system is functioning so that issues can be detected and resolved in advance or in order to provide services. Sometimes it contains unique identifiers. The most obvious of these were IP address, but there were also things like machine name, media access control (MAC) address, and so on.

Although collection of telemetry was not considered in the past the same as collecting personal information now, there have always been privacy concerns with it. These concerns were mainly whether the collection of it was authorized or not and thus whether it was a form of spyware or not (think industrial espionage).

However, with the widespread adoption of smartphones, PDAs, and other devices, the quantum leap in the ability to collect, parse, and understand patterns (i.e., Big Data or Data Science) and the ability to act on those patterns and push communications to devices (or take other actions) based on what was once just considered machine data has all changed.

Now unique identifiers such as those collected as part of collecting telemetry need to be examined and considered. The important thing to remember in evaluating whether a unique identifier falls under the definition of PI is that not all unique identifiers are equal. Below is a list of characteristics to consider when evaluating unique identifiers to see if any one of them is something that can reasonably be linked to a person or a person's device (vs. a system that front ends a network):

Uniqueness

Reidentification (correlating an identifier with other data that leads to the ability to identify the user)

Using as an "anchor" to aggregate and analyze information from one or more sources

Permanence

Frequency of change

Ease of change

Reachability (can it be used to contact or track)

Other Governance Standards of which to be aware

In addition to the OECD Guidelines, there are other frameworks such as the Generally Accepted Privacy Principles (GAPP), the 1995 EU Data Directive (also known as Directive EU 95/46/EC), the Federal Trade Commission's version of the FIPPs, the Asia-Pacific Economic Cooperation (APEC) Privacy Principles, and International Organization for Standardization (ISO) Standards that will inform how personal information and privacy issues are managed and governed. In the previous section, the OECD Guidelines have been highlighted to explain the notion of fair and legitimate processing of personal information. These other frameworks help one get to a more granular and comprehensive view of *data governance*, which will be discussed in Chapter 3.

Privacy Is Not Confidentiality and Security Is Not Privacy

Confidentiality is about protecting designated nonpublic information (often information that is either a trade secret or proprietary) (Figure 2-2).

Confidentiality ≠ Privacy

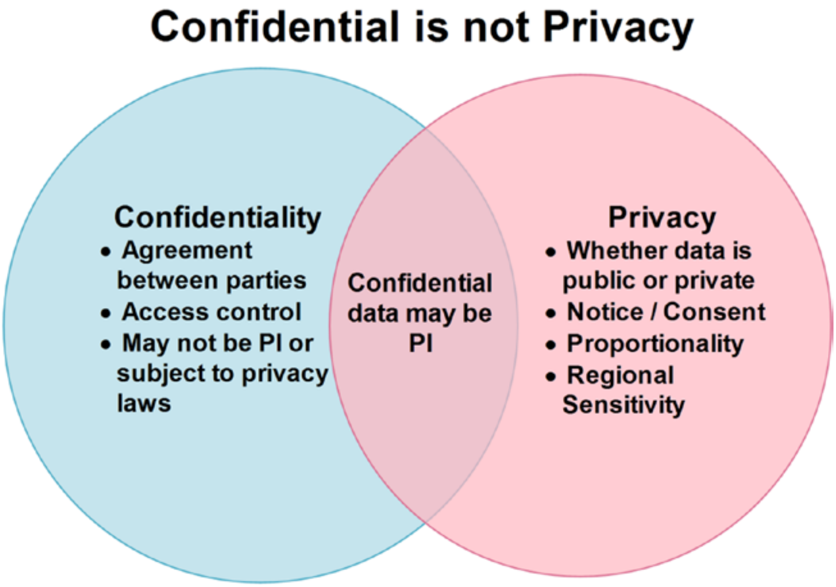
Confidential is not Privacy
Confidentiality
• Agreement between parties
• Access control
• May not be PI or subject to privacy laws
Confidential data may be PI
Privacy
• Whether data is public or private
• Notice / Consent
• Proportionality
• Regional Sensitivity

Figure 2-2. *Confidentiality is not privacy*

Confidentiality rules only apply to what is designated by agreement as confidential.

Sometimes confidential information is also personal information. For example, some information relating to the private lives of individuals may be confidential, such as medical records or family secrets. Sometimes, actually often, confidential information contains no PI.

This is the first difference between confidentiality and privacy. Confidential is an imposed label that signifies access control. PI is an organic label; it speaks to the substance of the information. Just as with that famous line in Shakespeare's immortal play *Romeo and Juliet* "A rose by any other name would smell as sweet," so it goes with PI. PI is always going to be personal information when it identifies an individual.

Another difference is that rules that govern or protect the PI apply whether the personal information is public or not. Just because PI is public does not mean it can be used or "processed" for one's own purposes. One example of this is e-marketing lists. Many of our e-mail address are publically available, but that does not mean they can be wantonly maintained on e-marketing lists without our permission.

A third difference, and perhaps the most important, is that when the PI is nonpublic personal information, keeping it "confidential" only addresses the access requirement and not the use or any of the other requirements of the OECD Guidelines.

So, although there is overlap between the safeguards used to protect personal information and the safeguards used to protect confidential information—most of the overlap is in terms of access control—protecting one is not the same as protecting the other.

Just as privacy and confidentiality overlap but are not the same, privacy and security overlap in that each is about data protection, but they are not the same (Figure 2-3).

Security ≠ Privacy

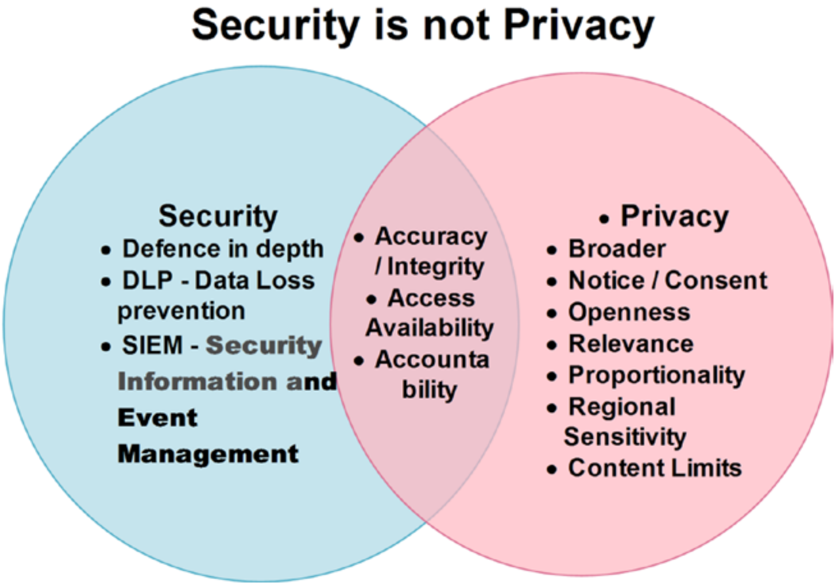

Figure 2-3. *Security does not equal privacy*

Information security has three areas of focus, known as CIA:

Confidentiality (i.e., preventing unauthorized access)

Integrity (i.e., ensuring the data is not altered without approval)

Availability (i.e., ensuring the data is accessible)

It uses logical, administrative, physical safeguards to ensure the CIA of the data is maintained. Aspects of security that do not overlap privacy include:

- *Defense in depth:* A sophisticated firewall structure can protect personal information.
- *Data loss prevention (DLP):* Discovering and monitoring the location and flow of sensitive data such as customer credit card data, employee PI, or corporate intellectual property.
- *Security information and event management* (SIEM)

The Overlaps

The safeguards enable the "authorized" in the "authorized access and use" element that is a cornerstone the operational definition of privacy. This is the first overlap between privacy and information security.

In addition to the fact that both "information security" and "privacy" are data protection regimens, other areas of overlap are:

Integrity (information security) and accuracy (privacy)

Availability (information security) and access (privacy)

Accountability (both)

Confidentiality (when the data is both personal information and nonpublic)

Information security's focus on data integrity overlaps with privacy's accuracy requirement in that both target ensuring the data is not altered with authorization.

Information security's availability requirement supports privacy's access requirement because if the data is not available, they cannot be accessed.

Both information security and privacy doctrines require data owners and custodians to be responsible for protecting the data in accordance with the respective protection regimen, which is a form of accountability.

And when the information is both nonpublic and personal information, confidentiality supports privacy because nonpublic data need to be kept nonpublic.

The Disconnects

The reason there is not a complete overlap between privacy and information security is threefold.

First, privacy has a wider set of obligations and responsibilities than information security does, such as:

- Collection limitation
- Openness
- Relevancy
- Use limitation

This means there are things privacy addresses that information security does not.

The second disconnect is confidentiality. Because PI is not always nonpublic (consider the phonebook), the notion of confidentiality does not apply. Also, in a resource-constrained world, if the data is not considered confidential, they are not always "valued" and the necessary measures to ensure authorized access and use will be overlooked.

Third, and perhaps most important, while information security techniques can be privacy-enabling technologies (PETs) (which means they are tools that enable privacy) and are often necessary, these PETs can also become "feral" if applied incorrectly (i.e., in an invasive manner). This is why you can have security without privacy, but you cannot have privacy without security. This will be discussed further in Part 2.

Conclusion

The purpose of this chapter is to enable you to understand the nature of privacy and privacy engineering.

This is the foundation and context for the guidance—the explanation of tools and techniques—that makes up the remainder of this book.

If you follow the guidance in this book, you will be poised for success and you will have a set of tools you can use and configure to enable privacy, but the actual success will ultimately depend on how you tailor the guidance that follows to specific situations (i.e., the data, the processing, whose data, and specific jurisdiction, regulations, or best practices that apply) and how you configure the tools we are providing. Chapter 3 will discuss privacy and data governance concepts.