The changing meaning of "personal data"

March 2011

By William B. Baker and Anthony Matyjaszewski

When FTC Commissioner Julie Brill last year described her vision of privacy in the future, which she dubbed Privacy 3.0, she opined that the distinction between "personally identifiable information" (PII) and "non-PII" is "blurring (http://blogs.berkeley.edu/2010/07/21/commissioner-brill-and-privacy-3-0-at-the-cwag-privacy-panel-2/)." This remark led the IAPP to start an inquiry into exactly what kind of information is "personal information" or "personal data" and how statutory definitions are subject to reinterpretation as technology evolves. The IAPP hopes that this initial effort will lead to further discussions about what *should* be protected by privacy law.

What is PII: Statutory Definitions

A starting point is a consideration of what constitutes PII under current statutory law. Is PII all information about a person? Does the information need to directly identify a person? Is it only recorded information? Does the information need to be true? Is a "person" only a natural person, or can they be legal persons such as corporations and organizations? If they are natural persons, does it matter if they are dead or alive?

These and other queries can be answered by examining the definitions of "personal information" or "personal data" in various countries. To begin a conversation about the nature of "data," the IAPP surveyed the definitions of personal data across 36 data protection laws in 30 countries. A summary of that research is attached.

Those 36 laws have taken many approaches. Some of these definitions, such as those in the United States, are relatively narrow and often specify particular items, while others, especially those in European Union countries and other laws modeled on their approach, tend to be broader.

Despite these differences, the statutes generally share a prototypical definition along the lines of "data or information relating to an identifiable person." All countries employ some variation of the phrasing, data "that allow the identification of a person directly or indirectly." For example, the European Union Directive on Data Privacy defines PII as data

> "relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity."

Despite this general similarity, laws differ in what actually qualifies for data protection. Some countries list specific examples of what can constitute personal data; others are satisfied with a more flexible—or ambiguous—definition. Although

specific definitions may offer the benefit of greater certainty, they are subject to criticism as rigid and incapable of responding to new developments. Conversely, the flexible definitions do allow for future adaptability but can lead to uncertainty.

Under these laws, data that do not constitute PII are regarded as "non-PII," subject to far less, if any, regulation. This concept has often applied to aggregated data and more recently has been extended to "de-identified" data from which identifying information purportedly has been removed.

Applying the Statutory Definitions

Looking abroad, the European Union Article 29 Working Party's Opinion 4/2007 (http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2007/wp136_en.pdf) offers further guidance on the meaning of personal data. The Working Party analyzed first, the type of data or information; second, the relation between the data and its subject; third, the concept of identifiability, and fourth, the person or individual to be protected.

Types of Information

The first step in the Article 29 Working Group's analysis looks at what *types* of information are protected. Consider first whether data must be in a recorded form to be protected or whether spoken words could come within the protection of the laws. Curiously, only Australia's Privacy Act 1988 and the United States' Health Insurance Portability and Accountability Act (HIPAA) expressly include protection for data that is *not* recorded either digitally or on paper in their respective definitions of personal data. The IAPP's research showed that most countries do not specify whether the data must be recorded—or if it can also be spoken words or opinions—leaving such matters open to interpretation or, perhaps, resolution, in cases involving difficult facts. This leaves open the possibility that not only recorded data but also information of a more ephemeral nature can fall under those nations' privacy protections. One conjures intriguing possibilities, as privacy laws in those countries

that do not address this matter could potentially protect one from having their names or other identifying characteristics spoken out loud. What a way to stop nasty gossip!

On the other hand, some privacy laws, such as those of Hong Kong and the United Kingdom, mention recorded data only. Singapore's Model Data Protection Code and the United States' Children's Online Privacy Protection Act (COPPA) further limit their reach to digital data or data collected online.

Must data about a person be true to be protected? Interestingly, only two nations—Australia and Singapore—explicitly state in their definitions of personal information that protection extends to both true and false data. The remaining surveyed laws do not address this matter in their definitions of personal information. Does this mean that these other countries will only provide privacy protection for true data? Most likely not, as may be inferred from other provisions in their privacy laws requiring data controllers to allow a person the opportunity to access and correct any false data pertaining to oneself, especially in the financial or credit sectors. Even though the definition of personal data in the European Union Data Protection Directive (http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML) also does not deal with the veracity of data, the Article 29 Working Party's Opinion 4/2007 states that both true and false data, as well as viewpoints, are covered by the directive. As such, the mere omission of a topic from the definition of personal information does not necessarily remove that matter from the scope of privacy protection.

Relationship to a Person

Moving to the Working Party's second analytical step—the *relationship* between the data and its subject—certain patterns emerge from the terms used in the various definitions of personal information in the laws researched by the IAPP. Privacy laws include terms such as "referring to," "relating to," "about" or "concerning" a person or individual. There is little substantive variance among the definitions, as they all

establish a link between the data and the person. After all, there presumably is little need to protect data that have no reference to someone whose privacy is being safeguarded.

Few problems exist where the relationship is quite straightforward, such as that of a name to a particular person. And this is especially true when the name is relatively distinctive, such as, say, Barack H. Obama. That does not mean, however, that a person necessarily has rights in her name. Ordinarily, the connections between the data and the subject are far more nebulous, and these can present difficult questions in privacy law.

The Article 29 Working Party's Opinion 4/2007 provides what is perhaps the most in-depth scrutiny of this factor by reducing it to three elements—content, purpose or result. The *content* element is perhaps the clearest of the three, as it addresses information about a person, such as their medical history, their contact information or their service record. Such information inherently refers to a particular individual.

The *purpose* element stipulates that even data that may otherwise not be considered personal information, such as a corporate call log of a company's telephones, may become personal information when used to monitor an employee's telephone activity. In such a case, the Article 29 Working Party would consider the data to be personal data relating to both the employee making the calls *and* the recipient of the call on the other end of the connection.

The *result* element posits that even data not about a particular person—thus lacking the content element—and not used to gain information about a person, lacking the purpose element—may still be considered personal information where a person's rights or interests are affected. For example, a satellite positioning system being used solely for the purpose of ensuring efficient dispatching of taxis or delivery vehicles would still provide personal information because the location data could potentially result in the monitoring of drivers' whereabouts and behavior.

The Article 29 Working Group's approach seems to leave room for the range of matter deemed to be "personal information" to expand—or shrink—over time in response to developments. The remainder of this article explores some of the issues that may arise in understanding what data and information will fall within the scope of "personal" under these laws.

Types of Persons

Although the meaning of *persons* who are entitled to privacy protection under these laws is listed fourth in the European Union analysis, we consider it third here. Predominantly, the definitions of personal information apply only to natural persons, or human beings. However, Argentina, Austria, Colombia, Italy and Switzerland also extend (http://www.austlii.edu.au/au/journals/PLPR/2001/58.html) privacy protection to legal persons such as corporations, partnerships or other organizations. The potential scope of this presents fascinating questions. Does it mean that corporations cannot be made identifiable, and any information that makes it possible to identify a specific company should be treated as personal information? Or perhaps the protection is meant to protect corporate secrecy or the privacy of the individuals within the corporate structure.

On the other hand, the Australian Law Reform Commission considered extending privacy rights to corporations but ultimately rejected the idea, stipulating that there are existing statutory protections of intellectual property and business confidentiality that serve the same purpose more effectively. And in Canada, both the Personal Information and Electronic Documents Act (PIPEDA) and the British Columbia Personal Information Protection Act (PIPA) specifically exclude an individual's business contact information from privacy protection. These types of statutes exist worldwide, and they are perhaps the main reason why more countries have not moved to extend privacy protection to corporations on the same terms as they are applied to individuals. The U.S. Supreme Court on March 1, 2011, ruled that corporations are not "persons" for the purpose of the Freedom of Information Act.

Even where the definition of "person" is narrowed to human beings, some countries view "personal" as protecting the privacy only of the living, but not the deceased. Again, most of the definitions are silent on this matter. Hong Kong, Ireland, Japan, New Zealand, Sweden and the United Kingdom specify that only the living are entitled to privacy protection, and New Zealand protects information relating to a death. In the European Union, the Data Protection Directive does not require the extension of privacy protection to the deceased, although individual countries are free to do so at their discretion. The Article 29 Working Party has elaborated their position on data protection for the deceased by stating that such information only requires protection in the event that it can be used to identify living individuals, as would be the case in genetically transmitted diseases such as hemophilia. This demonstrates once more the variance in approaches to defining personal information among the surveyed laws, with some countries preferring a narrower scope while others choose to keep the definition more broad.

Identification

The remaining step of the inquiry looks at the substance of the *identification* requirement of the data subject. As to this point, a consensus appears among the various laws in the surveyed countries. Not one of these laws requires a person actually to become identified. They either leave this as an open-ended possibility, by using the term "identifiable," or they specify that the data are protected if they can cause the data subject to be "identified or identifiable." Thus, the mere possibility of identification can be enough for data to become personal information.

Similarly, none of the laws require a person to be directly identified through the use of the data in question. While nearly half of the definitions of personal information are silent on this matter, many state that possible indirect identification is enough to trigger protection. Essentially, this means that the information in one's possession will need to be treated as personal information even if it does not identify an individual, so long as it can be combined with other available information for that

very purpose. It should be apparent that determining how readily a person can be identified can be a very fact-specific inquiry, and what is not identifiable one year may, a few years later, be determined to be identifiable.

An example used by the Article 29 Working Party envisions a newspaper article about a criminal case that describes some of the details but does not directly name any of the individuals. So long as there is publicly available information, either in court records or in other newspaper articles, that allows one to ascertain the identities of the people involved, then even an article that does not identify the individuals would be deemed, at least by the Article 29 Working Party, to contain personal information. American law generally reaches a different result, as a person named as a possible criminal suspect in a U.S. news article typically has little recourse other than a defamation action.

Applying this "relating to" provision, nonetheless, can be a vexing task in privacy law. It is an area particularly vulnerable to technological developments that place great stress on existing statutory definitions.

Re-identification

In response to laws imposing greater obligations on the custodians of personal information, a practice arose over the years to remove certain information from a compilation of PII in order to "anonymize" or "deidentify" the data so that it could be processed as non-PII. This practice underlies many laws today, which typically require far less protections for "non-personally identifiable information." For example, in the United States, the Privacy Rule implementing the Health Insurance Portability and Accountability Act specifies 18 different categories of identifying information that must be removed in order to "de-identify" health information. However, enterprising researchers in recent years began to demonstrate that it is often possible to "re-identify" supposedly anonymized data.

There have been several well-publicized examples. One involved research by LaTanya Sweeney in Massachusetts, who identified then-Gov. William Weld's medical records using only a state-released "anonymized" data set and a list of registered voters.

More recently, Netflix found it appropriate to cancel a second “Netflix Contest” after researchers were able to identify “anonymized” Netflix viewers in the first “Netflix Contest”—in which it offered $1 million to any researcher who could best improve its recommendation engine—from viewer reviews posted on The Internet Movie Database Web site. Among the characteristics that could be identified were the users' political leanings and, in some instances, even sexual orientation.

These episodes demonstrate that the process of de-identification is not nearly as simple or easy as once may have been believed. Data controllers that wish to de-identify PII are on notice to take greater pains to do so. At this point, however, it is not possible how much is enough, as resourceful researchers will invariably have many tools available to reassemble data if given sufficient motivation. And it is not clear that the answer lies in the “foreseeability” that re-identification is possible, but foreseeability may simply be a function of one’s ingenuity. Note that the tools these researchers used—voter registration lists, Internet databases—were not arcane but were commonplace items that, presumably, were never considered to the de-identifiers as posing a potential risk.

Indeed, Professor Paul Ohm has gone so far as to declare that data can be “useful or perfectly anonymous but not both.” Time will tell whether Prof. Ohm’s provocative formulation is correct or not, but his aphorism highlights the difficulties of anonymizing PII.

Internet Protocol Addresses

A current topic of hot debate is whether a computer user’s Internet Protocol (IP) address should be considered PII. The law appears in flux at the moment, and complicating matters is that regulators and courts are reaching different conclusions.

Privacy regulators in the European Union regard dynamic IP addresses as personal information. Even though dynamic IP addresses change over time, and cannot be directly used to identify an individual, the Article 29 Working Party believes that a copyright holder using “reasonable means” can obtain a user’s identity from an IP

address when pursuing abusers of intellectual property rights. More recently, other European privacy regulators have voiced similar views regarding permanent IP addresses, noting that they can be used to track and, eventually, identify individuals.

This contrasts sharply to the approach taken in the United States under laws such as COPPA where, a decade ago, the FTC considered whether to classify even static IP addresses as personal information but ultimately rejected (http://www.ftc.gov/os/1999/10/64fr59888.pdf) the idea out of concern that it would unnecessarily increase the scope of the law. In the past few years, however, the FTC has begun to suggest that IP addresses should be considered PII for much the same reasons as their European counterparts. Indeed, in a recent consent decree, the FTC included within the definition of "nonpublic, individually-identifiable information" an "IP address (or other "persistent identifier")." And the HIPAA Privacy Rule treats IP addresses as a form of "protected health information" by listing them as a type of data that must be removed from PHI for deidentification purposes.

However, courts are more reluctant to do so. For example, the Irish High Court held in April 2010 that an IP address does not constitute "personal data" when being collected by record companies for the purpose of detecting copyright infringement. And a U.S. federal district court in Washington state also held that an IP address is not PII because it identifies a computer rather than a person. The law is far from settled on this point, however, so lawyers must follow developments closely.

Device Fingerprinting

A newer approach to identifying a particular Internet user is device fingerprinting or, in the online context, "browser fingerprinting." This process focuses on the particular software configuration of a user's browser—the browser type, fonts and other factors—and it happens that the particular combination of such factors on a user's computer is often unique to that user. (The Electronic Freedom Foundation has done useful work in this area). Useful to the entity interesting in "tracking" a user is that no cookie or other code is placed on the user's computer; the tracking is done remotely by using information routinely supplied by the browser to a Web site. The user's name, by the way, is never disclosed, but her device is uniquely identified and

capable of being tracked. This device fingerprinting technology did not even exist just a few years ago. The question, from a legal standpoint, is whether a user's browser configurations are, or will soon become, "PII" for regulatory purposes.

Smart Grid

The "smart grid" will present similar issues in a few years. Once utility companies are capable of monitoring the usage of particular appliances in particular homes, it will be only a matter of time before telltale "identifying" patterns of usage begin to emerge. Energy companies might provide incentives to use certain appliances at off-peak hours; marketers might have a keen interest in knowing which consumers make frequent use of the microwave. The utility companies will surely have some ability to correlate usage patterns with particular customers, but how will definitions of PII factor into the smart grid.

Questions for the Future

These developments regarding reidentification, IP addresses, browser fingerprinting and the smart grid provide examples of how new technological developments can cause the reidentification of data previously deemed non-PII. Other issues abound, such as the extension of privacy to photographic data, especially as it relates (http://gigaom.com/2010/02/26/eu-google-street-view/) to Google's Street View map service, as well as geographic location information derived (http://www.manatt.com/NewsEmail.aspx?id=12032#Article2) from a new generation of mobile devices. In none of these cases has a legislature changed a statutory definition; each involves the application of a previously-established definition in light of new technology. In this way, the process of re-identification can be said to enable technology to redefine PII.

Is there a limit to how technology can redefine PII? To how much effort must a re-identifier go, or, put differently, is there some reasonable limit that a de-identifying entity can assume applies when attempting to render data non-identifiable? Or, is the problem a limit on people's ability to imagine or foresee how a re-identifier might go about her task?

Existing statutory laws neither ignore this problem nor resolve it. The laws often contain limitations to how practicable such indirect identification must be, and this is where different approaches are taken in the laws surveyed. For example, Hong Kong's Personal Data Protection Ordinance and Poland's Act on the Protection of Personal Data stipulate that data will not be protected if indirect identification is not practical or if it requires unreasonable cost, time or manpower, respectively. Of course, practicality and reasonability are unspecific concepts. Again, the Article 29 Working Party offers some guidance on this topic within the European Union by utilizing a cost-benefit analysis. Accordingly, the mere hypothetical possibility of identification is not enough, and one should consider the cost of conducting the search, the expected benefit of identifying the person and the interests at stake in order to determine whether a person is identifiable.

Still, this leaves many questions unanswered. A calculation of costs and benefits will change as technology creates new ways of combining information or researchers become more clever. Remember that Ms. Sweeney needed only a registered voter list to identify Gov. Weld's medical records, something plainly not foreseen by Massachusetts authorities but, in hindsight, perhaps not so surprising. How "practical" is device fingerprinting today or will it be in two years?

Conclusion

The different ways that similar statutory language is applied around the world causes problems in practice. Any business that conducts operations in more than one country faces a continuing challenge of understanding and complying with legal terms that are applied differently across borders. And, after understanding the differing definitions, they must then comply with the corresponding policies that govern data in each country.

Going forward, with new technological advances being made on a regular basis, these definitions of personal information, and the type of data they cover, will be reshaped, refined and revised. There is a strong likelihood that the driver of these "redefinitions" will be the technological developments themselves. That is, even where statutory definitions provide what legislators intended to be clear

classifications, changes in technology may be, in effect, “amending” these statutes without any legislative action. The future course of such “blurring” is a trend worth watching.

The IAPP hopes that the compendium of laws attached to this article will prove to be a helpful contribution to the discussion.

Go to compendium (https://iapp.org/resources/article/compendium-the-changing-meaning-of-personal-data).

William B. Baker is a partner at Wiley Rein LLP in Washington, DC. Anthony Matyjaszewski is a member of the University of Maine Law School class of 2011.

The research for this article was first presented at the IAPP Privacy Academy in Baltimore on September 30, 2010, by Mr. Baker. An earlier version of this document was distributed at that presentation.

Tags: Education (/tag/education), Financial (/tag/financial), Government (/tag/government), Health Care (/tag/healthcare), HR (/tag/hr), Marketing (/tag/marketing), Africa (/tag/africa), Asia-Pacific (/tag/apac), Canada (/tag/canada), EU (/tag/eu), Latin America (/tag/latin-america), U.S. (/tag/u-s), Infosecurity (/tag/infosecurity), Internet of Things (/tag/internet-of-things), Personal Privacy (/tag/personal-privacy), Privacy Operations Management (/tag/privacy-operations-management), Privacy Research (/tag/privacy-research)

Pease International Tradeport, 75 Rochester Ave, Suite 4
Portsmouth, NH 03801 USA • +1 603.427.9200