PUTTING YOUR PRIVACY PEOPLE BACK TO SLEEP WITH DE-IDENTIFICATION

ROBERT L. ROTHMAN*

We privacy professionals just LOVE data flow maps. Data flow maps show in colorful PowerPoint lines and cute little icons the origin of personal information, how it gets from point A to point B, together with all the stops and processing that takes place along the way. All those fancy arrows make us confident that we understand exactly what is happening with our company's personal data so we can determine which of the plethora of privacy rules that exist today are applicable. For this reason, chief privacy officers (CPOs) around the world dispatch emissaries to various parts of their organizations to map the travels of human resource data, customer data, shareholder data, etc. Upon completion of each map, necessary safeguards and contracts are put in place, required filings are made and whatever else that has to happen to comply with the applicable laws is done. Eventually the CPO has a whole atlas of data flow maps together with privacy solutions for each of the issues identified, all of which allow him to sleep soundly each night.

A series of studies by the Ponemon Institute have been awakening those CPOs from their nocturnal slumbers like the Ghost of Christmas Future. While nothing has disturbed the information in our beloved data flow maps, it turns out that they often reflect only the company's standard ongoing operations. It seems that IT departments are sending personal information all over the world in connection with "testing" new systems, modifications or fixes to existing systems, etc. And this appears to be happening, unknown to most CPOs, on a very regular basis. Consider these findings:

- Eighty percent of respondents in the U.S. and 77 percent in the U.K. report that they use real production data as part of their application development and testing process.
  Ponemon Institute: *Data Security in Development & Testing*, July 31, 2009, p2.

- In the financial services industry, the numbers may even be slightly higher: Over 83 percent of [financial services] companies use real (live) customer or employee information in development and testing, and 51 percent of these companies admit they **do not take** appropriate steps to protect real data used in development and testing.
  Ponemon Institute: *Privacy & Data Protection Practices Benchmark Study of the Financial Services Industry*, January 31, 2010, p2.

- Customer and consumer records are typically the largest data files and are most frequently used for development and testing purposes. Eighty-nine percent of companies that use live data use customer records and 74 percent use consumer lists.
  Ponemon Institute: *The Insecurity of Test Data: The Unseen Crises*, December 11, 2007, p4

- Seventy-one percent of respondents in the U.K. and 79 percent of respondents in the U.S. say they use files with more than one terabyte of real data in development and testing.
  Ponemon Institute: *Data Security in Development & Testing*, July 31, 2009, p7.

- Seventy percent of U.S. respondents and 60 percent of U.K. respondents send real data to third-party organizations for development and testing.
  Ponemon Institute: *Data Security in Development & Testing*, July 31, 2009, p7.

These Ponemon Institute findings have led sleepless CPOs to have some very frank discussions with their IT Departments:

CPO: *Why didn't you tell us we're sending real customer personal information to our IT department in Ireland?*

IT Department: *You didn't ask. Besides, it's just for testing, not for production.*

CPO: *But I had three people here interviewing you about data flows and you didn't mention it.*

IT Department: *They didn't ask. Besides, it's only for testing and we didn't send more than 10,000 customer records.*

CPO: *And I now understand that Ireland sends the same personal data to an Indian software development company for some sort of specialized test process Ireland doesn't do.*

IT Department (exasperated): *Of course, how else would you suggest we do it?*

CPO (excitedly): *Don't you realize there are legal consequences to sending that personal data overseas? Don't you realize that our privacy statement doesn't tell customers we are using their data to test our own systems and will be sending it overseas? Don't you realize that when the American personal data is sent to Ireland it may become subject to the Irish Data Protection Act and in order to send it to India you need to put in place an EU Standard Clause Contract between the Irish company and the Indian company or our U.S. entity has to register with the U.S. Department of Commerce for Safe Harbor and the Irish company has to send the personal information to the U.S. company first which can transfer it to the Indian company once a qualifying Onward Transfer Agreement has been put in place or that the whole company has to develop Binding Corporate Rules which have to be approved by all the relevant data protection authorities in Europe? And what about the breach statutes? Huh? Are you sure we will get notice if there is an unauthorized access to any of that American personal information while it is in Ireland or India? We have an obligation to give notice to the American customers if that happens! The sky is falling!!!!!*

IT Department (calmly): *Not to worry. We have all the legal issues covered. The Irish company signed our standard non-disclosure agreement that I'm sure had been approved at some point by legal. Besides, they said they would be careful.*

CPO: *Arghhhhhhhhhh!!! Didn't you attend any of the privacy training sessions?!!*

CPO (trying to calm down): *Okay, let's skip the legal issues for a moment. Can you please tell me why in the world you sent real customer data instead of making up a few names and using those?*

IT Department: *Well, in order to test the system thoroughly we needed at least 10,000 customer records of various types. All the normal data fields would have to be filled in. To make up a customer record, it would have taken 3.5 minutes per record for a total of 35,000 minutes, which is 583.3333 hours or 72.91667 man days or …*

CPO (interrupting): *Okay, okay. I guess I can understand budget issues. Any other reason?*

IT Department: *Yes. In order to make certain table C operates properly it is necessary to use the data elements set forth on the eighth column of table WW together with the alternate algorithm, and they have to be realistic.*

CPO: *I don't understand. Can you clarify?*

IT Department: *Of course, I'd be happy to. You see, the Asymmetric Digital Subscriber Line is measured in MIPS. CADE, considering the Peltier effect, has adverse consequences for the ECOphlex root partition, when described in ASCII. Therefore, Ανοδικά κινήθηκε στην τελευταία συνεδρίαση της εβδομάδας ο γενικός δείκτης τιμών του Χρηματιστηρίου Αθηνών. Εκλεισε στις. Furthermore, 成为全球最大的实验室和处方数据 …*

CPO: *Okay, okay.*

An increasing number of insomniac CPOs and frustrated CIOs who have been through these kinds of conversations are now turning to de-identification of personal information to solve their privacy problems while giving their IT departments the latitude they need to deal with the technical test issues. One form of de-identification is anonymization, where all personal identifiers are permanently removed and it is not possible to reverse the process. Another form is pseudonymization, where the process can be reversed with the proper "key." Programs are available that can quickly and effectively anonymize large databases by essentially mixing up the various data elements. The different elements (e.g., first name, last name, month of birth, day of birth and year of birth) are each shuffled to create, what in this case would be, fictitious people with fictitious but plausible names and birth dates. Yet, because of the algorithms used, if a real individual showed up in the database more than once, the system would assign that individual the same false name each time the individual appeared in the anonymized database.

This approach has advantages to both those who speak IT and those who speak privacy. For IT speakers, the database is realistic and suffers from the same kind of defects and other characteristics that an actual database has. This helps assure that any IT test regime will actually be effective. For privacy speakers, the anonymized database has no information that falls within any major definition of personal data. Thus, in the conversation recorded on the previous page there would be no American privacy statement problem because no actual customer personal information was sent to Ireland or was used in system testing; it was only data of fictitious persons. Even if the data became subject to Irish privacy laws there would be no problem sending it to India because it is not "data related to an identified living person" and, because the data was anonymized, as opposed to pseudonymized, there is not even an issue with "data related to an *identifiable* living person." That means no worries about EU standard clause contracts, no Safe Harbor filings, no consents, and no Binding Corporate Rules to draft and get approved. Finally, the American state breach laws would not be applicable because no personal data of a state resident was involved. Even if the data were published on the front page of the *Irish Times* or *Hindustan Times*, no breach notifications to U.S. customers would be necessary.

De-identification of personal data allows organizations to sidestep complex and costly privacy regulation. It allows IT organizations to get on with necessary testing without unforeseen delays or costs due to compliance requirements or breaches. And as for us privacy people, we enjoy renewed faith in our beloved data flow maps knowing the IT department isn't sending actual personal data around the world. And we may *finally* be able to get in those forty winks.

Sponsored by

Compuware

THE TECHNOLOGY
PERFORMANCE COMPANY

compuware.com