Data Privacy Fix Broader Than Social Security Numbers

May 01 2008

Posted By: Mark F. Foley

Practice Area: Information Technology, Data Privacy and Security

Editor's note: Gov. Jim Doyle has called for state agencies to replace Social Security numbers with random identifiers for the administration of state programs.

On March 5, the Wisconsin Assembly passed Bill AB 771, which prohibits any state agency from using a Social Security number as an identifier unless such use is required by state or federal laws or regulations, or is otherwise authorized by law. If enacted by the Senate and signed by the Governor, this bill will join many other laws in Wisconsin and elsewhere that limit the use of SSNs, but the issue involved is broader than SSNs alone. The passage of this bill should remind everyone of the need to apply the "Use Limitation Principle" to all information technology activities.

The bill responds to several high profile, unauthorized disclosures of SSNs by state agencies. In late 2006, a contractor mailed 170,000 residents Wisconsin tax forms with SSNs printed on the label. In November 2007, the names, e-mail addresses, and SSNs for 200 faculty and staff of the University of Wisconsin-Madison's Division of Information Technology were published in a web database created for "statistical analysis." In January 2008, the Wisconsin Department of Health and Family Services mailed information to 260,000 elderly, disabled, and low-income individuals with their SSNs visible.

The main concern about such disclosures, whether by government or businesses, is that SSNs can be used for identity theft. Criminals can use SSNs with correct, incorrect, or made-up names to create new identities and open credit accounts. Credit agencies are not required to, and generally do not, give notice to an account holder when his SSN is used to open an account in a different name. As a result, SSNbased fraud can go undetected for years until a lender denies the real SSN owner credit or tries to collect from the real owner a debt she did not create.

If the purpose of AB 771 is to prevent similar disclosures of SSNs in the future, it is not likely to succeed. This is because both state agencies involved are authorized or required by law to collect and use SSNs for their activities. These agencies will still have the SSNs and the data will still be at risk. The problem, and the solution, lies elsewhere.

Useful Limitations

Unauthorized uses or disclosures of SSNs often result from violation of the "Use Limitation Principle." That is, to best protect privacy interests, data should be collected only for a specified limited purpose and not used for any other purposes. The federal government created SSNs in 1935 to track individual taxpayer's eligibility for Social Security benefits. Since these benefits are based on income earned, the Internal Revenue Service also adopted SSNs to identify individual taxpayers.

SSN cards state on their face: “Not to be used for identification.” Nevertheless, many agencies and companies have created databases unrelated to Social Security or payroll using the SSN as the primary key to identify a record or data subject, and then tie all their data to that field. This was an easy shortcut for database design since there are no duplicate SSNs and alternative strategies would require the database designer to devise a system for generating identification numbers and assuring that they remained unique.

But use of the SSN as the recurring key identifier in credit, education, financial, and other databases has led to much of the identity theft risk we face today. When that number is publicly disclosed from any source, it can be used, or abused, in connection with every other database using that same identification number. The broader the use of SSNs, the greater the risk to the individual.

The “Use Limitation Principle” would bar the use of a SSN for anything but its original purpose. Although you might still need the SSN somewhere in your payroll database to report earnings and tax withholding to the government, you would not use the SSN as your primary employee ID and would not use it to link various subcategories of data. Rather, you would develop one or more unique employee identifiers that do not include and are not based on the SSNs. Then, if data containing your identifiers are lost or stolen, the risks of data compromise are limited to your own database, and the risks of identify theft or other misuse are much reduced. And you would not allow, much less encourage, use of a SSN as a user ID or password.

Implementing the Use Limitation Principle requires security in the form of access control. Only those persons with a need to use specific fields of information should have access to those fields. Only information relevant to a specific task should be provided to data processors. For example, the vendor that mailed forms for the Department of Revenue needed a list of names and addresses; it didn’t need SSNs to do its job. If the vendor had not received the unneeded SSN data, it could not have inadvertently disclosed it.

Both the process of providing information to the mail fulfillment vendor and the individuals involved in sending and receiving the data, should have been better attuned to the Use Limitation Principle. Similarly, in a typical “lost laptop” case, a user downloads large amounts of information from a general company database to an Excel spreadsheet so that the user can work at home. When the laptop is lost or stolen, all these data are compromised. Often the facts emerge that the data were downloaded en masse because it was easy, not because the user actually needed all the data fields or records available.

What To Do?
What does your organization do? When designing databases, do you instruct the developers to use key fields that do not involve SSNs or other sensitive identifiers? When designing access controls, reports, data export scripts, etcetera, do you require an analysis of what data fields are actually necessary for the person receiving access or a data subset? Do you conduct privacy and security assessments of the requested data downloads? Do you insist upon business processes that incorporate best privacy and security practices, such as automatic encryption of all data saved to a laptop or removable media? Do you have sufficient granularity in the levels of access control in your IT systems? Do you train employees always to use the most limited subset of data suitable to the task at hand?

The best practices are not hard to identify, although they can be difficult to implement. All databases, reports, scripts, and data access rules should be designed with the Use Limitation Principle in mind. Collect only the information that is needed for the task at hand, not everything you might someday think about using. Organize the data around key fields that do not lend themselves to reuse or misuse, especially SSNs and other official identification numbers. Use technological, physical, and organizational means to limit each user to access only the data necessary for a particular task. Automatically encrypt all sensitive data when it is first exported, downloaded, or saved. Train. Audit. Repeat.