Confusing Osama bin Laden with Johnny Rotten

The U.S. government's passenger screening technologies would mistake the terrorist mastermind for the Sex Pistol. In fact, data broker ChoicePoint possesses better tracking systems.

By Mark Williams Pontin

April 4, 2007

At the end of last February, the U.S. Department of Homeland Security (DHS) launched its Traveler Redress Inquiry Program for the 30,000-plus individuals who in the years since September 11 have been misidentified as possible terrorists by the Transportation Security Administration's (TSA) infamous "no fly" and "selectee" lists. These people may now ask for investigative reviews via an official website, in the hope that the TSA will eventually remove their names.

Alas, the realization of their hopes may be long postponed. Officially, the TSA's much delayed Secure Flight computerized passenger prescreening system will roll out by fall 2008 at the earliest. But TSA administrators have told Congress that full implementation of the system–costing $140 million already and requiring at least $80 million more–may not happen before 2010. Translation: nobody at the DHS and TSA will be taking responsibility for removing any names from the watch lists, and individuals on the lists will continue to undergo extra screening of their persons and carry-ons.

In short, the farce of federal efforts to create an efficient terrorist profiling system to keep terrorists off airplanes–and the farce of privacy-advocacy organizations' reactions to those efforts–will continue. Before September 11, 2001, the U.S. government's list of suspected terrorists banned from air travel held 16 names. Afterward, every government agency indiscriminately dumped information about every potential suspect from its databases onto the watch lists. By March 2003, when the TSA did early tests of CAPPS II (Computer Assisted Passenger Pre-screening System II), the watch lists had expanded to 75,000 names–many of them being, notoriously, common ones like Ted Kennedy and Robert Johnson.

CAPPS II would have required the collection of four pieces of personal data–name, address, telephone number, and birth date–to authenticate travelers' identities. It then would have transmitted that information to commercial data-brokerage companies (primarily AcXiom, the self-styled "industry leader in the use of grid computing") so as to check it against data mined from credit reports, voter-registration cards, driving records, and such, in order to then generate a secret "risk score" for every individual.

Needless to say, all this alone was sufficient to provoke organizations dedicated to civic liberties such as the American Civil Liberties Union (ACLU), the Electronic Frontier Foundation (EFF), and the Electronic Privacy Information Center (EPIC). Additionally, however, the White House, DHS, and Justice Department wanted to expand CAPPS II so as to catch illegal aliens and domestic criminals, despite objections from TSA administrators who had promised to limit the system to profiling foreign terrorists. In the summer of 2004–when some of those administrators were threatening to mutiny, CAPPS II was a year overdue and nonfunctional, and the Government Accountability Office was reporting that the system wouldn't protect individuals' privacy–then-DHS chief Tom Ridge disingenuously suggested to reporters that it would be killed.

In fact, CAPPS II got sent back to the drawing board and scaled down. The requirements for screening for criminals were dropped, and the system was rebranded as Secure Flight. The fights between the TSA and the privacy-advocacy organizations over questions like to what extent the agency could use data collected by the commercial data brokerages have continued. In February 2006, sources at the National Counterterrorism Center told the *Washington Post* that the watch lists had grown to 325,000 names–more than quadruple the 75,000 on the lists in 2003.

For now, the DHS–of which the TSA is a component–may have achieved a temporary end run around the activists. In November 2006, the DHS's privacy office revealed that its Automated Targeting System (ATS) has all along been analyzing the international passenger name records (PNRs) that airlines send DHS screeners. It has also been assigning secret risk scores to individuals based not just on their names, addresses, and seat assignments, but also on how their tickets were paid for, whom the passengers might be traveling with, and what telephone numbers were used to book flights. Some 50 privacy-activist organizations have reacted to this with outrage, claiming that they'd been falsely led to believe that the ATS was being used merely to identify cargo aboard ships, so they'd concentrated their attention on the CAPPS II and Secure Flight systems.

In response, DHS boss Michael Chertoff has insisted that he has talked about Homeland Security collecting and analyzing such information in hundreds of speeches–which he has, although never alongside mentions of the ATS. In a December 8, 2006, *National Journal* article, Chertoff indulged in some minor drollery at the privacy activists' expense: "I've got a new rule. If I want to keep a secret, I give a speech about it. Because if I make a speech, no one picks it up. But if I put it in a document and I slip it under the table, then it gets the front page." The same article reported that Chertoff complained at length about the activists' "penchant for placing great demands on the department, then scolding it for missing deadlines or for being ineffective."

With that last comment, Chertoff actually has a point. Media coverage tends to presume a heavy-handed U.S. government aiming at Orwellian surveillance while fairly uncritically accepting the claims of the privacy-activist organizations. However, although plenty of evidence certainly supports the "heavy-handed U.S. government" thesis, the truth is that it's also logically inconsistent for the activists to insist that individuals' names are all the data that government databases should collect while simultaneously complaining that misidentifications, or false positives, are endemic. Latanya Sweeney, director of the Data Privacy Laboratory at Carnegie Mellon University's School of Computer Science, in Pittsburgh, is a computer scientist who specializes in learning how individuals' personal data is vulnerable and how their privacy may be preserved as data surveillance proceeds–a process she describes as "selective revelation." Sweeney says that false positives will obviously be problematic with the watch lists: "The only

input there is the name, with no secondary data to disambiguate individuals. In addition, the watch-list technology is *totally* unacceptable."

Sweeney means that the U.S. government watch lists, besides containing common names like Ted Kennedy, depend on variations of a phonetic algorithm called Soundex. As she says, "Soundex is an old patent that's been used for a long time, whenever they have two databases where they're trying to match up records." Indeed, Soundex dates back to a time when Hollerith punch cards were the newest thing in computing technology. Developed to index and retrieve soundalike surnames with different spellings (like Rogers and Rodgers) scattered throughout an alphabetical list, Soundex was first used so that U.S. government clerks could retroactively analyze the 1890 U.S. census results. Soundex works by taking the first letter of a name, dropping all vowels, assigning a number to each of the next three consonants (with similar-sounding consonants like *s* and *c* getting the same numbers), then dropping any remaining consonants. Thereby, the algorithm reduces all names to a letter followed by three numbers.

Consequently, Soundex assigns to the name Laden the code L350, as it does Lydon, Lawton, and Leedham. This is, in other words, an algorithm so deficient for identification purposes that it confuses al Qaeda's Osama bin Laden and the Sex Pistols' Johnny (Lydon) Rotten. To see for yourself how poorly Soundex performs, go to nofly.s3.com, where S3 Matching Technologies has combined the algorithm with a list of potential-terrorist names recorded in U.S. government databases. "The U.S. government obviously updates its lists every day, so we don't suggest this is up-to-date," says James Moore, a company spokesperson. "But we got the best available data on who'd be on terrorist watch lists from various private intelligence agencies." Using Soundex and S3 Matching Technologies' version of the watch list reveals that the names Jesus Christ and George Bush resemble terrorists' names enough that they're assigned to the no-fly or selectee list.

How does the U.S. government rationalize using such error-prone technology for its watch lists? Sweeney says, "Whomever I ask–whether it's DHS, DARPA, the Department of Justice–everybody essentially says, 'We're just going to plow ahead.' At the DOJ, the answer I get is, 'It'll get solved when we use biometrics.' Their belief is that the current problem will disappear because you'll show your driver's license and match your fingerprint against your fingerprint's stored image on your license." Sweeney half-seriously proposes a hypothetical solution to the watch-list problem. "I've told ChoicePoint that they ought to go into the watch-list business."

Alongside Lexis-Nexis and AcXiom, ChoicePoint is one of the big-three data-brokerage corporations and in many ways the most interesting of them. Evan Hendricks, editor-publisher of the Washington-based *Privacy Times*, says, "Though most Americans don't know about ChoicePoint, it's a company that knows a lot about hundreds of millions of Americans." Would ChoicePoint have a minimum of four data points–name, address, social-security number, and birth date–for almost every adult U.S. citizen, and therefore have enough information to differentiate among, say, any five people with names whose Soundex hashes would come out the same? Hendricks answers, "That's certainly true. So would the three main credit-reporting companies." However, Hendricks continues, whereas the big-three credit-reporting agencies–Experian, Trans Union and Equifax–calculate individuals' credit scores, ChoicePoint defines itself as a data-aggregation company in the business of selling actionable intelligence to both industry and government, with credit-related information being only a subset of that whole.

ChoicePoint doesn't only possess copious records on U.S. citizens (its subsidiary, VitalChek Network, provides the technology to process and sell birth, death, marriage, and divorce records in

every U.S. state). It has also acquired data on some 300 million citizens of Mexico, Brazil, Colombia, Argentina, Nicaragua, Guatemala, Honduras, El Salvador, and Costa Rica–a fact that emerged in 2003, after the company disclosed that it had bought data (reportedly including even passport numbers and unlisted phone numbers) on Mexico's entire roll of 65 million registered voters. Whether ChoicePoint still retains this information is unclear since, as part of its $67 million annual contract with the U.S. Department of Justice, the company was supplying the information to the U.S. government, and Mexico, Nicaragua, and Costa Rica responded with arrest warrants–and in Mexico's case, threatened to bring charges of treason–against the local individuals who'd sold the data to ChoicePoint. In June 2003, the company claimed to the relevant countries that it had expunged their citizens' information from its databases.

Other products and services provided by ChoicePoint include the DNA identification of the victims of the Word Trade Center attacks on September 11 via its subsidiary, the Bode Technology Group (sold off by ChoicePoint in March), and SmartSearch, which performs "wildcard searches" that can construct a comprehensive personal profile in minutes, starting with only a first name or partial address. More controversially, ChoicePoint subsidiary Database Technologies (also known as DBT Online) was contracted to assemble a list of voters barred from voting by the state of Florida and was responsible for an alleged 57,700 people–primarily African-American and Hispanic Democrats–being incorrectly listed as felons during the U.S. elections of 2000. Other ChoicePoint divisions deliver all types of credential verification, employment-background screenings, drug testing, criminal records, motor-vehicle records, mortgage-asset research, tenant screening, database software, medical information, and services for the life- and health-insurance fields.

Would it be a good or bad thing, on balance, if the government were to farm out to ChoicePoint the administration of the watch lists in their entirety? Hendricks says, "I think it would overall be a very bad thing." In this, Hendricks echoes the general sentiment among privacy advocates. At the Big Brother Award ceremonies held annually by U.K.-based Privacy International, ChoicePoint has twice been a winner: in 2001, as "Greatest Corporate Invader" for "massive selling of records, accurate and inaccurate to cops, direct marketers and election officials," and again in 2005, as "Lifetime Menace Award" for its continuing efforts to build dossiers on individuals.

It's extraordinary, of course, that private corporations should have had the means to accumulate–and trade–more personal data on Americans than the U.S. government possesses. Moreover, limited avenues for rectification and reparation are available to citizens in the face of the sea of errors that exist in these corporations' databases. Nevertheless, in substantial measure, the privacy activists played no small part in bringing this extraordinary situation to pass, as we shall see in the second part of this article next week.