

STEVEN C. MCCRAW  
ASSISTANT DIRECTOR  
OFFICE OF INTELLIGENCE  
FEDERAL BUREAU OF INVESTIGATION

HOUSE GOVERNMENT COMMITTEE  
SUBCOMMITTEE ON TECHNOLOGY  
INFORMATION POLICY, INTERGOVERNMENTAL  
RELATIONS AND THE CENSUS

"DATA MINING: PROTECTING THE HOMELAND,  
SAFEGUARDING AMERICAN VALUES"

MAY 6, 2003

Good morning Mr. Chairman and Members of the Subcommittee:

My name is Steve McCraw, and I am the Assistant Director of the FBI's Office of Intelligence. I am pleased to have an opportunity to appear before you today to discuss the FBI's use of "data mining" and the safeguards it has in place to protect the privacy and personal information of American citizens.

First, I would like to take this opportunity to thank you and the Members of the Subcommittee for the support you have provided to the FBI in modernizing our information technology infrastructure. The FBI for too long has not been able to capitalize on the tremendous advances in information technology to quickly locate essential elements of information and identify previously unknown

links, relationships, and associations hidden within the vast amount of data collected by the FBI in the performance of its investigative responsibilities.

Second, I would like to thank the Subcommittee for this opportunity to address the issue of "data mining" and personal privacy. As witnessed both here in Congress and in the press, it is clearly an issue in the hearts and minds of the American people. The FBI is proud of its commitment to privacy concerns and I am happy to discuss the efforts made in this regard.

#### "Data Mining"

\_\_\_\_\_As background to this issue, I believe it is important to understand the term "data mining," as it is commonly used today. "Data mining" is defined as technology that facilitates the ability to sort through masses of information through databases exploration, extract specific information in accordance with defined criteria, and then identify patterns of interest to its user. In general it refers to the ability to work with larger amounts of data, at faster speeds, in ways that were previously not possible computationally due to size or speed limitations. The private sector often uses data mining to make sense of the wide breadth of data that companies and industries have available. For example, data mining is often used by industry to analyze potential goods and services that are in demand. Companies that use data mining shorten response time to market changes, which allows for better alignment of their products with their customer's needs. Both government and industry have also successfully used data mining to identify and protect against fraud. A key principle of the FBI's information technology business plan

calls for the use of effective industry practices and off the shelf private sector technology to allow the FBI to work more efficiently, more effectively, and more economically. Further, prior to procuring these technologies we carefully review the purchase to ensure they meet privacy laws, policies and regulations.

In recent debates, however, people have begun to use the term "data mining" as a shorthand reference to the specter of abusive searches through vast amounts of publicly available data on innocent private citizens. As your letter requesting FBI attendance at this hearing astutely recognized, these are not the same things. The term "data mining" should not be viewed as always connoting any such abuse.

#### Uses of "Data Mining"

\_\_\_\_\_The United States Constitution and the United States Congress, through legislation, have carefully delineated acceptable conduct in law enforcement investigations and intelligence activities. The FBI has an unwavering commitment to adhere to those requirements, as well as those mandated by Federal regulations and the Attorney General guidelines. Whether the work is performed manually or in an automated fashion, the commitment does not change.

I'd like to provide some examples of the ways in which the FBI uses, or plans to use, "data mining":

- The FBI's Integrated Automated Fingerprint Identification System (commonly known as IAFIS) has long been available to the law enforcement community to permit verification of arrestee identity. When necessary, IAFIS will compare the

fingerprints of a non-cooperating arrestee against millions of known fingerprints in order to provide law enforcement with the individual's identity. It also conducts pattern searches to link on unknown latent fingerprints taken from a crime scene with known individuals.

- The new SCOPE/Integrated Data Warehouse project will allow an agent or analyst to search within the FBI's existing datasets (that is, information collected in lawful investigations and stored in computer format) for links, associations, and relationships among the individuals.
- The Information Sharing Initiative begun in St. Louis will permit sharing of state, local, and Federal data to enable officers to quickly search through multiagency investigative data to identify links between subjects of terrorism and criminal investigations.

#### Risk Assessment Technology

The Subcommittee has specifically inquired about the use of risk assessment technology. To the FBI, this term connotes the use of technology to identify individuals who present a particular risk, for example, a risk to the national security. As a general rule, the FBI until recently has not been able to participate in the use of such technology because of inadequate information technology capabilities.

The FBI intends to use the advances in risk technology to identify FBI employees whose activities include a pattern of possible misuse of the FBI's computer systems to access information. The use of this technology and increased audit functions will substantially increase FBI internal security. This ability would have assisted in earlier detection of convicted spy, former FBI Special Agent Robert Hanssen.

The FBI will also leverage risk assessment technology to increase its oversight of human sources. As part of a reengineering project to expand the FBI's Human Intelligence base while providing greater oversight, the FBI will utilize information technology to identify potential problems in the operation of human sources.

Ensuring the appropriate controls to protect the privacy of Federal government data, public source data that contains personal information on American citizens, as well as a combination of the two remains a top priority. Certainly, the FBI uses information collected by public source companies to obtain information on individuals during the course of its terrorist, criminal, and foreign intelligence investigative activities. Often, these systems provide leads that enable the FBI to save valuable investigative and analytical time and resources. Again, acceptable conduct in the collection and use of this information has been clearly defined by statute and Attorney General guidelines. Although public source data is a useful tool, the FBI is well aware that the data is sometimes outdated and/or inaccurate and follow-up investigation to confirm the information is essential.

However, the FBI neither has in the past nor intends in the future to purchase personal information on American citizens who are not part of an ongoing investigation from public source companies and placed it within the FBI's system of records. The FBI has no interest in gathering data on law abiding citizens. Such information is not required to protect the nation.

In closing, I want to thank you for the opportunity to testify before you today and I look forward to any questions you may have for me.