



## Testimony

Before the Subcommittee on Technology, Information Policy, Intergovernmental Relations and the Census, Committee on Government Reform, House of Representatives

---

For Release on Delivery  
Expected at time 9:30 a.m. EST  
Tuesday, March 25, 2003

# DATA MINING

## Results and Challenges for Government Program Audits and Investigations

Statement of Gregory D. Kutz, Director  
Financial Management and Assurance



---

---

This is a work of the U.S. government and is not subject to copyright protection in the United States. It may be reproduced and distributed in its entirety without further permission from GAO. However, because this work may contain copyrighted images or other material, permission from the copyright holder may be necessary if you wish to reproduce this material separately.

---



Highlights of GAO-03-591T, a report to the Subcommittee on Technology, Information Policy, Intergovernmental Relations and the Census, Committee on Government Reform, House of Representatives

## Why GAO Did This Study

The Subcommittee asked GAO to testify on its experiences with the use of data mining as part of its audits and investigations of various government programs. GAO's testimony focused on (1) examples and benefits of the use of data mining in audits and investigations and (2) some of the future uses and challenges in expanding the use of data mining in audits of federal programs. Much of GAO's experience with data mining to date relates to its audits of the Department of Defense's (DOD) credit card programs.

## DATA MINING

# Results and Challenges for Government Program Audits and Investigations

## What GAO Found

GAO's data mining work related to audits and investigations of federal government credit card and other programs has identified fraud, waste, and abuse resulting from breakdowns in internal controls. We used these data mining techniques, in conjunction with systematic internal control testing, to make recommendations to federal agencies to develop effective systems and controls that provide reasonable assurance that fraud, waste, and abuse in these credit card and other programs are minimized. For these programs, GAO's data mining often involves extracting information on credit card users or vendors using a set of defined criteria (e.g., vendors that the federal government would not typically do business with) and then having auditors and investigators follow-up on selected transactions or vendors.

Data mining alone is generally not sufficient to identify systemic breakdowns in controls and to provide management with recommendations to improve systems of internal controls. Systemic breakdowns can best be demonstrated using statistical tests of key controls along with a thorough assessment of the overall control environment. Data mining results serve to "put a face" on the control breakdowns and provide managers with examples of the real and costly consequences of failing to properly control these large programs.

Recent GAO audits using data mining of DOD purchase and travel card programs have identified numerous prohibited purchases of goods and services from vendors such as restaurants, grocery stores, casinos, toy stores, clothing or luggage stores, electronics stores, gentlemen's clubs, legalized brothels, automobile dealers, and gasoline service stations.

GAO's use of data mining has expanded beyond the government credit card programs. At the request of several congressional committees and Members, we currently have underway a number of audits and investigations that will utilize data mining, including.

- DOD vendor pay systems
- Army military pay systems
- Department of Housing and Urban Development housing programs
- Department of Energy national laboratories

Challenges to expanding the use of data mining in the federal arena include data integrity and security issues. For example, DOD has long-standing problems with financial systems that are fundamentally deficient and are unable to provide timely and reliable data. Data security issues related to the use of large, detailed databases are another issue that must be considered before undertaking a data mining project. With the right mix of technology, human capital expertise, and data security measures, GAO believes that data mining will prove to be an important tool to help it to continue to improve the efficiency and effectiveness of its audit and investigative work for the Congress.

[www.gao.gov/cgi-bin/getrpt?GAO-03-591T](http://www.gao.gov/cgi-bin/getrpt?GAO-03-591T).

To view the full report, including the scope and methodology, click on the link above. For more information, contact Gregory D. Kutz at (202) 512-9095 or [kutzg@gao.gov](mailto:kutzg@gao.gov).

Mr. Chairman and Members of the Subcommittee:

Thank you for the opportunity to discuss current applications and future possibilities for the use of data mining. We use the term “data mining” to mean analyzing diverse data to identify relationships that indicate possible instances of previously undetected fraud, waste, and abuse. Auditors can use data mining to extract individual, or a series of, questionable transactions from large data files for follow up by auditors or investigators. Data mining can also help serve as a deterrent to those who believe they can get away with fraud because of weak or nonexistent internal control systems.

To date, GAO has used data mining as an integral part of our audits and investigations of federal government credit card programs. For these programs, our data mining work has identified fraud, waste, and abuse resulting from breakdowns in internal controls. We used these findings, in conjunction with systemic internal control testing, to make recommendations to federal agencies on actions needed to develop effective systems and controls that provide reasonable assurance that fraud, waste, and abuse in these credit card programs are minimized. My testimony will (1) discuss examples and benefits of the use of data mining in our audits and investigations and (2) some of the possible future uses and challenges to expanding our data mining beyond federal government credit card programs.

### **Use of Data Mining in Federal Government Audits and Investigations**

Data mining has been an integral part of our audits and investigations of federal government purchase and travel card programs. For these programs, data mining has involved obtaining large databases of credit card transactions and related activity and using software to search or “mine” data looking for suspicious vendors, transactions, or patterns of activity. Our data mining often involves extracting information on credit card users or vendors using a set of defined criteria (e.g., vendors that the federal government would not typically do business with) and then having auditors and investigators follow-up on selected transactions or vendors. (See attachment 1 for a list of related GAO products resulting from our data mining.)

We have used data mining for credit card audits in conjunction with our evaluation of the design and effectiveness of internal controls intended to prevent fraud, waste, and abuse in these programs. Our methodology for performing these audits included the following four basic steps:

- gain an understanding of the credit card program;
- make a preliminary assessment of the adequacy of internal controls;
- test the effectiveness of internal controls; and
- identify, using data mining, case studies demonstrating the cause and real life effect of the control breakdowns.

An important element of success in our audits is the integration of our audit and investigative functions. Our auditors and investigators work together on a daily basis on all four steps of the process. In developing effective data mining strategies, we found that it is critical for the auditors and investigators to have a thorough understanding of the program and the related processes and internal controls. Once the process and controls are understood, we then assessed the adequacy of key internal control activities and the overall control environment. For example, in making this assessment for the Department of Defense (DOD) purchase card program, we identified a weak overall internal control environment, including a proliferation of credit cards, which left the program vulnerable to fraud, waste, and abuse. In addition, once vulnerabilities are identified, investigators and auditors work together to identify various schemes that could be used to abuse the program including committing fraud. Our understanding of the program and its vulnerabilities is then used to develop our data mining strategy.

We used data mining and follow on audit and investigative work to demonstrate the effect of systemic breakdowns in internal controls. Data mining alone is generally not sufficient to identify systemic breakdowns in controls and to provide management with recommendations to improve systems of internal controls. Systemic breakdowns can best be demonstrated using statistical tests of key controls along with a thorough assessment of the overall control environment, including existing policies and procedures that govern control activities.

#### Data Mining Criteria and Techniques Used in DOD Purchase and Travel Card Program Audits

The use of purchase cards has dramatically increased in past years as agencies have sought to lower transaction processing costs and eliminate the lengthy processes and paperwork long associated with making small purchases. DOD is promoting department wide use of purchase cards for obtaining goods and services. It reported that for the year ended September 30, 2002, purchase cards were used by about 214,000 cardholders to make about 11 million transactions valued at over \$6.8 billion. Purchase cards may be used for acquisitions at or below the \$2,500 micropurchase threshold, and for payment of items costing over \$2,500 from contracts or other purchase agreements. DOD estimated that in fiscal year 2001, about 95 percent of its transactions of \$2,500 or less were made by purchase card.

In 1983, the General Services Administration (GSA) awarded a governmentwide master contract with a private company to provide government-sponsored, contractor-issued travel cards to be used by federal employees to pay for costs incurred on official business travel. The intent of the travel card program was to provide increased convenience to the traveler and to reduce the government's cost of administering travel by reducing the need for cash advances to the traveler and the administrative workload associated with processing and reconciling travel advances. Our audits of DOD's travel card program focused on individually billed accounts, which are held and paid by individual cardholders. According to GSA, as of September 30, 2002, DOD had over 1.3

million individually billed travel cardholders who charged \$2.4 billion during the fiscal year.

We assessed controls over the Army, Navy, and Air Force purchase and travel card programs. In each case, we found that a weak overall control environment and breakdowns in key internal control activities left the military services vulnerable to fraud, waste, and abuse. We looked for indications of potential fraud, waste, and abuse as part of our statistical sampling and through nonrepresentative selections of transactions using data mining. Because DOD's purchase and travel card programs involved different key control activities and vulnerabilities, we tailored our data mining techniques to address the unique characteristics of each program. However, we did not look at all potential abuses of either the purchase and travel card and our work was not designed to identify, and we did not attempt to determine, the full extent of potential fraud, waste, and abuse related to the purchase and travel card programs.

For our purchase card audits, we obtained transaction databases for our study period from the purchase card contract banks—U.S. Bank for the Army and Air Force and Citibank for the Navy. For our travel card audits, we obtained transaction databases for the three military services from DOD's travel card contractor—Bank of America. In all cases, control totals from these databases were reconciled to bank or GSA reports to ensure we had a complete and accurate database for our sampling and data mining. Using several database manipulation software tools, we selected transactions or patterns of activity that appeared to represent potential fraud, waste, or abuse. We then conducted additional audit and investigative follow-up based on the nature, amount, timing, and other characteristics of the transactions. In some instances, we also compared (“bumped”) data from different databases to identify anomalies. Our data mining criteria included the following.

#### Nature of the transaction

- Prohibited merchant category codes<sup>1</sup> that should have been blocked, such as jewelry stores, pawn shops, and gambling establishments.
- Personal use, including food, clothing, luggage and accessories, such as sunglasses, purses, and totes.
- Travel related transactions, such as airfare, hotels, and restaurants (for purchase card audits).

---

<sup>1</sup> Merchant category codes (MCC) are established by the banking industry for commercial and consumer reporting purposes. Currently, about 800 category codes are used to identify the nature of the merchants' businesses or trades, such as airlines, hotels, ATMs, jewelry stores, casinos, gentleman's clubs, and theaters.

## Merchants

- Specialty stores, such as hobby shops, sporting goods stores, Victoria's Secret, L.L. Bean and toy stores (e.g., Toys 'R' Us).
- "Dot com" vendors, such as REI, SkyMall, Internet gambling sites, and pornography sites.
- High-end stores, such as Dooney & Bourke, Coach, and Louis Vuitton.
- Department stores, such as Nordstrom and Macy's.
- Other personal use vendors, such as Ticketmaster, Mary Kay Cosmetics, and Avon.
- Gentlemen's clubs and legalized brothels.
- Cruise lines, sporting events, casinos, taxidermy services, and theaters.

## Dollar Amount of Transaction

- Transactions having unusually high dollar amounts (for travel card audits).
- Convenience checks over \$2,500 (for purchase card audits).
- Numerous recurring transactions with the same vendor indicating the need for a contract (for purchase card audits).
- Transactions in round dollar amounts, such as \$330, \$440, etc., indicating possible fee for cash schemes (for travel card audits).
- Multiple, recurring small ATM transactions, indicating possible personal use (for travel card audits).

## Timing of Transactions

- Holiday and weekend transactions.
- End of fiscal year transactions.
- Transactions that were made late at night.
- Multiple transactions on the same day, at same vendor, totaling more than \$2,500, indicating split purchases (for purchase card audits).

## Other Characteristics

- Out of state purchases, when similar items have been purchased locally (for purchase card audits).
- Transaction in which the cardholder and merchant had the same name.
- Cardholders who wrote nonsufficient funds checks (for travel card audits).
- Charged-off accounts, and accounts in salary offset or fixed payment plans (for travel card audits).

To fully develop the case study examples that we included in our reports required extensive collaboration on the part of auditors and investigators. It is clear that data mining techniques, although a powerful tool by themselves, are best used in combination with strategies that create a synergy between teams of auditors and investigators to identify and develop case studies on the causes and effects of any control breakdowns. Our auditors have expertise in financial systems, data manipulation, and evaluating internal controls. Our investigators are federal agents with years of law enforcement experience, particularly in the area of detecting financial crimes. Further, we found that the experience gained with each successive audit increased the knowledge base of our auditors and investigators and improved the overall data mining results.

## Data Mining Results in DOD Purchase and Travel Card Program Audits

Data mining “puts a face” on the control breakdowns and provides managers with examples of the real and costly consequences of failing to properly control these large programs. Recent GAO audits using data mining of DOD purchase and travel card programs have identified numerous prohibited abusive or questionable purchases of goods and services from vendors such as restaurants, grocery stores, casinos, toy stores, clothing or luggage stores, electronics stores, gentlemen’s clubs, legalized brothels, automobile dealers, and gasoline service stations.

Specific examples of abusive and questionable activity identified as a result of the previously discussed data mining criteria and techniques include

- Nature of the transaction: blocked merchant category code (MCC) – As part of our audit of the Army purchase card program, we identified a cardholder transaction for \$630 that was coded as being from an escort service, which should have been a blocked MCC code. As part of our investigation we determined that this was an unauthorized, potentially fraudulent transaction, and that the cardholder was also being investigated for possible theft of chapel funds.
- Merchants – Gentlemen’s Clubs and Brothels – We found that DOD cardholders used their government travel cards at legalized brothels in Nevada and at gentlemen’s clubs that provide adult entertainment. We initially identified this

abusive use of the travel card based on our interviews with cardholders. Subsequently, we used this information to refine our data mining and identify a substantial number of these transactions.

- Merchants – Taxidermy Services – An Air Force cardholder used the purchase card to prepare a shoulder mount of a mule deer head. The deer was a “road kill” that was found on the roadside by an approving official who approved the purchase of taxidermy services. The deer head was hung on the wall in the Natural Resources Office. The cardholder, approving official, and two other employees occupy the office where the deer head currently hangs.
- Dollar Amount of Transaction: High Dollar Purchases – For the Army travel program, we found that a cardholder’s spouse used his government travel card to make two payments of \$2,050 each to Budget Rent-A-Car for the purchase of a used automobile.
- Dollar Amount of Transaction: Recurring Purchases – During fiscal year 2001, the Navy purchased over \$1 million from 122 different vendors using the purchase card. In total, these vendors were paid about \$330 million. However, despite this heavy sales volume, the Navy had not negotiated reduced-price contracts with any of the vendors.
- Timing of Transaction – In an audit of the Navy purchase card program, we identified about \$12,000 in potentially fraudulent fiscal year 2000 transactions. These purchases occurred primarily between December 20 and December 26, 1999, and included an Amana range, Compaq computers, gift certificates, groceries, and clothes.

In addition, we used data mining techniques to identify 220 cardholders that abused their travel card or had been involved in potentially fraudulent activity and who had severe financial problems. We compared records for these cardholders with DOD databases that included security clearance information. Based on this analysis, we found that 97 of 220 individuals with severe financial problems continued to maintain secret or top-secret security clearances at the end of our respective audits.

#### Data Mining Results at Other Federal Agencies

We have used data mining techniques to help assess the controls over various programs at the Departments of Housing and Urban Development (HUD) and Education and the Federal Aviation Administration, among others. Further, our October 2001 Executive Guide entitled, *Strategies to Manage Improper Payments: Learning From Public and Private Sector Organizations* (GAO-02-69G), discusses the use of data mining techniques by various state and federal programs as part of a research-based approach to fraud prevention and detection. For example, the Illinois Department of Public Aid used data mining techniques to identify health care providers that were billing for services provided in excess of 24 hours in a single day. Their analysis identified 18 providers that had billed over 25 hours for at least 1 day during the 6 months ended December 31, 1999.

As a result, the Illinois Department of Public Aid Office of Inspector General planned to refer serious cases to appropriate law enforcement agencies and take administrative action against the less serious violators.

Additional examples of the results of our data mining at other agencies include the following:

- At the Department of Education, we performed a variety of data mining queries and found that three schools fraudulently disbursed about \$2 million in Pell Grants to ineligible students and another school improperly disbursed about \$1.4 million in Pell Grants to ineligible students.
- At the Department of Housing and Urban Development (HUD), we identified a scheme where only one-third of the work paid for by HUD to replace a concrete sidewalk was actually performed. As a result, more than \$164,000 of the \$227,500 billed and paid for appeared to be fraudulent.

### **Future Use of Data Mining and Related Challenges**

Our use of data mining has expanded beyond government credit card programs. This expansion provides opportunities for significant impact and improvements in other programs but also presents other challenges. At the request of several congressional committees and Members, we currently have a number of audits, which will utilize data mining. These audits include the following.

- DOD Vendor Pay Systems – This effort is an evaluation of the adequacy and effectiveness of DOD’s controls over its vendor pay processes. With reported annual vendor payments in excess of \$77 billion, this program entails most of DOD’s disbursements for items (excluding major weapons systems).
- Army Military Pay Systems – This effort is an evaluation of the Army’s controls over the payroll payments to military members. For fiscal year 2002, Army’s reported payroll was about \$32 billion.
- Centrally-billed travel accounts – These accounts are used primarily to purchase transportation including airline tickets. This activity was about \$1.5 billion for fiscal year 2002.
- Governmentwide purchase card program – We are evaluating whether the federal government is effectively managing its procurements of \$15 billion in goods and services using purchase cards.
- HUD single and multifamily properties - As a follow-on to previous work, we are evaluating the propriety of payments made related to HUD-owned single and multifamily properties.
- Department of Energy contractor-managed national laboratories - In response to allegations of improprieties at the Los Alamos national laboratory, we are assessing internal controls over disbursements and whether purchases made are a valid use of government funds at selected other laboratories.

For each of these audits, we are in the process of developing and/or executing data mining strategies to assist with the identification of breakdowns in controls or the inefficient use of federal funds. In addition, in response to a congressional request, we are preparing a guide to assist federal agencies in their efforts to audit internal controls of government purchase card programs. We have found that as government purchase card use grows, federal and state and local government auditors are increasingly being asked to do more audits of these programs. Building on the lessons learned from our purchase card work, our guide is intended to provide a blueprint for other auditors to use when auditing purchase card programs. This guide will include a section on data mining and related follow-up.

For the credit card work to date, we have used databases provided by the contractor banks. We found that the data quality is high, thus allowing us to do efficient and effective data mining. However, a challenge with federal government databases is that the quality and availability of information from which to mine data is often poor. For example, we have previously reported that DOD's financial systems are fundamentally deficient and are unable to provide data in a timely and reliable manner for decisionmaking. These data problems result in the following challenges for future data mining.

- For DOD, data needed for effective data mining may not be available in any one system. Consequently, obtaining and reconciling data from numerous databases is necessary to develop populations from which to data mine. In addition, because of the large volume of transactions involved in many DOD program areas, storing and conducting data mining queries of such large files may present a significant challenge.
- Because databases do not reconcile to independent, reliable sources, the completeness of databases used for data mining is questionable.
- Many agencies have known problems with data reliability.

In most cases these issues can be overcome, but they result in less productive data mining, and increase the cost of doing the work.

Other challenges lie in the area of data security and privacy protection. For example, as part of our extensive use of many detailed databases to assess the controls over DOD's credit card programs, we developed strict protocols to protect the sensitive data included in the databases. We were especially concerned with protecting active credit card account numbers and individual social security numbers. Data security issues must be addressed before embarking on audits involving data mining.

## **Conclusions**

The use of data mining is a critical component of the audit and investigation of certain federal programs. The results of data mining show real consequences or effect of breakdowns in internal controls. In addition, data mining results contribute greatly to

the development and implementation of recommendations to management on improvements in controls that can provide assurance that fraud, waste, and abuse is minimized. We are in the process of moving beyond the use of data mining for government credit card programs to other areas of interest to the Congress. We are just beginning to make full use of data mining strategies. With the right mix of technology, human capital expertise, and data security measures, we believe that data mining will prove to be an important tool to help us to continue to improve the efficiency and effectiveness of our audit and investigative work for the Congress.

### **Contacts and Acknowledgments**

For future contacts regarding this testimony, please contact Gregory D. Kutz at (202) 512-9095. Individuals making key contributions to this testimony included Francine DeVecchio, Steve Donahue, Gayle Fischer, Geoffrey Frank, John Kelly, Mai Nguyen, John Ryan, Kara Scott, and Scott Wrightson.

Attachment 1

**Related GAO Products**

*Travel Cards: Control Weaknesses Leave Navy Vulnerable to Fraud and Abuse.* GAO-03-147. Washington, D.C.: December 23, 2002.

*Travel Cards: Air Force Management Focus Has Reduced Delinquencies, but Improvements in Controls Are Needed.* GAO-03-298. Washington, D.C.: December 20, 2002.

*Purchase Cards: Control Weaknesses Leave the Air Force Vulnerable to Fraud, Waste, and Abuse.* GAO-03-292. Washington, D.C.: December 20, 2002.

*Travel Cards: Control Weaknesses Leave Army Vulnerable to Potential Fraud and Abuse.* GAO-03-169. Washington, D.C.: October 11, 2002.

*Travel Cards: Control Weaknesses Leave Navy Vulnerable to Fraud and Abuse.* GAO-03-148T. Washington, D.C.: October 8, 2002.

*Financial Management: Strategies to Address Improper Payments at HUD, Education, and Other Federal Agencies.* GAO-03-167T. Washington, D.C.: October 3, 2002.

*Purchase Cards: Navy Is Vulnerable to Fraud and Abuse but Is Taking Action to Resolve Control Weaknesses.* GAO-02-1041. Washington, D.C.: September 27, 2002.

*Travel Cards: Control Weaknesses Leave Army Vulnerable to Potential Fraud and Abuse.* GAO-02-863T. Washington, D.C.: July 17, 2002.

*Purchase Cards: Control Weaknesses Leave Army Vulnerable to Fraud, Waste, and Abuse.* GAO-02-844T. Washington, D.C.: July 17, 2002.

*Purchase Cards: Control Weaknesses Leave Army Vulnerable to Fraud, Waste, and Abuse.* GAO-02-732. Washington, D.C.: June 27, 2002.

*FAA Alaska: Weak Controls Resulted in Improper and Wasteful Purchases.* GAO-02-606. Washington, D.C.: May 30, 2002.

*Government Purchase Cards: Control Weaknesses Expose Agencies to Fraud and Abuse.* GAO-02-676T. Washington, D.C.: May 1, 2002.

*Education Financial Management: Weak Internal Controls Led to Instances of Fraud and Other Improper Payments.* GAO-02-406. Washington, D.C.: March 28, 2002.

*Purchase Cards: Continued Control Weaknesses Leave Two Navy Units Vulnerable to Fraud and Abuse.* GAO-02-506T. Washington, D.C.: March 13, 2002.

*Purchase Cards: Control Weaknesses Leave Two Navy Units Vulnerable to Fraud and Abuse.* GAO-02-32. Washington, D.C.: November 30, 2001.

*Purchase Cards: Control Weaknesses Leave Two Navy Units Vulnerable to Fraud and Abuse.* GAO-01-995T. Washington, D.C.: July 30, 2001.

(192095)