

GAO
Accountability · Integrity · Reliability
Highlights

Highlights of [GAO-05-866](#), a report to the Ranking Minority Member, Subcommittee on Oversight of Government Management, Committee on Homeland Security and Governmental Affairs, U.S. Senate

Why GAO Did This Study

Data mining—a technique for extracting knowledge from large volumes of data—is being used increasingly by the government and by the private sector. Many federal data mining efforts involve the use of personal information, which can originate from government sources as well as private sector organizations.

The federal government’s increased use of data mining since the terrorist attacks of September 11, 2001, has raised public and congressional concerns. As a result, GAO was asked to describe the characteristics of five federal data mining efforts and to determine whether agencies are providing adequate privacy and security protection for the information systems used in the efforts and for individuals potentially affected by these data mining efforts.

What GAO Recommends

GAO is making recommendations to the agencies responsible for the five data mining efforts to ensure that their efforts include adequate privacy and security protections. The agencies responsible for the five efforts we reviewed generally agreed with the majority of our recommendations, but disagreed with others.

www.gao.gov/cgi-bin/getrpt?GAO-05-866.

To view the full product, including the scope and methodology, click on the link above. For more information, contact Linda D. Koontz (202) 512-6240 or koontzl@gao.gov.

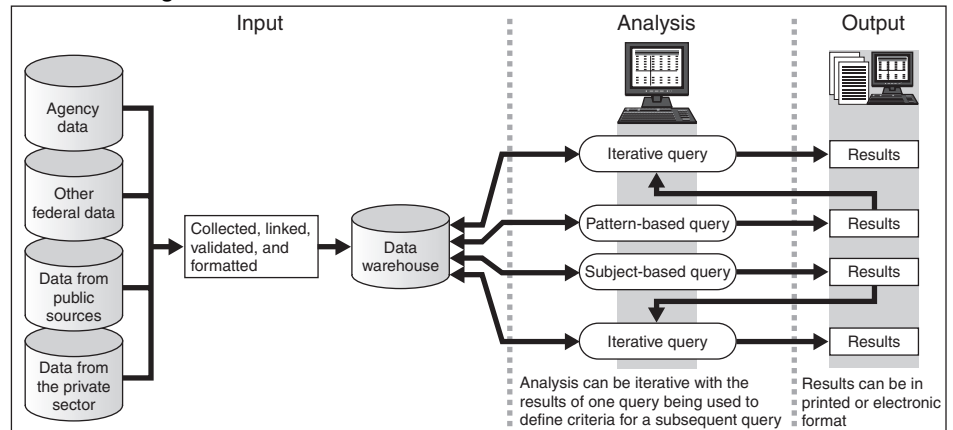
DATA MINING

Agencies Have Taken Key Steps to Protect Privacy in Selected Efforts, but Significant Compliance Issues Remain

What GAO Found

The five data mining efforts we reviewed are used by federal agencies to fulfill a variety of purposes and use various information sources, including both information collected on behalf of the agency and information originally collected by other agencies and commercial sources. Although the systems differed, the general process each used was basically the same. Each system incorporates data input, data analysis, and results output (see figure).

The Data Mining Process



Source: GAO, adapted from Vipin Kumar and Mohammed J. Zaki.

While the agencies responsible for these five efforts took many of the key steps required by federal law and executive branch guidance for the protection of personal information, they did not comply with all related laws and guidance. Specifically, most agencies notified the general public that they were collecting and using personal information and provided opportunities for individuals to review personal information when required by the Privacy Act. However, agencies are also required to provide notice to individual respondents explaining why the information is being collected; two agencies provided this notice, one did not provide it, and two claimed an allowable exemption from this requirement because the systems were used for law enforcement. In addition, agency compliance with key security requirements was inconsistent. Finally, three of the five agencies completed privacy impact assessments—important for analyzing the privacy implications of a system or data collection—but none of the assessments fully complied with Office of Management and Budget guidance. Until agencies fully comply with these requirements, they lack assurance that individual privacy rights are being appropriately protected.