G A O
Accountability * Integrity * Reliability

United States Government Accountability Office
Washington, DC 20548

February 16, 2007

The Honorable John Conyers, Jr.
Chairman
Committee on the Judiciary
House of Representatives

The Honorable Barney Frank
House of Representatives

Subject: Office of Special Counsel Needs to Follow Structured Life Cycle Management Practices for Its Case Tracking System

The Office of Special Counsel (OSC) is charged with safeguarding the merit system by protecting federal employees and applicants for employment from prohibited personnel practices, such as discrimination, nepotism, and retaliation against whistleblowing.[1] An individual who feels that a prohibited personnel practice has occurred may file a claim with OSC, which OSC then investigates and on which it may seek corrective or disciplinary action through negotiation with agencies or prosecution of claims before the Merit Systems Protection Board.[2] In addition, federal employees, former federal employees, and applicants for federal employment may also disclose to OSC alleged wrongdoing by federal employees (termed whistleblower disclosures), including violations of law, gross mismanagement, or abuse of authority.[3] OSC also provides advisory opinions and enforces Hatch Act restrictions on the political activities of individuals employed by the federal and District of Columbia governments as well as certain state and local government employees in connection with programs financed by federal funds. OSC also prosecutes claims before the Merit Systems Protection Board on behalf of federal employees, former federal employees, and applicants for federal employment under the Uniformed Services Employment and Reemployment Rights Act of 1994 (USERRA), which protects the employment and reemployment rights of federal and nonfederal

[1] Prohibited personnel practices are specified in 5 U.S.C. § 2302(b).

[2] An independent, quasi-judicial agency in the executive branch, the Merit Systems Protection Board serves as the guardian of federal merit principles.

[3] Reprisal for whistleblower disclosure is a prohibited personnel practice.

employees who leave their employment to perform military service and prohibits discrimination against individuals because of their military service.[4] OSC reports annually to Congress on the number of all types of cases it receives, processes, and closes as well as the disposition of those cases.

In the course of two prior reviews at OSC,[5] we found discrepancies in the data generated by OSC's case tracking system—OSC 2000—in the number of cases pending at the beginning of a fiscal year as well as cases received and closed during the year. This report responds to your concerns about the possibility that the data in OSC 2000 may be unreliable and that in turn OSC data on caseloads may be in error. As discussed, our objectives were to (1) identify what actions OSC has taken to help ensure the reliability of its case tracking system and related data and (2) determine whether OSC has corrected the types of data discrepancies we identified during previous work.

To identify what actions OSC has taken to help ensure the reliability of its case tracking system and related data, we reviewed existing documentation on OSC 2000, interviewed knowledgeable OSC officials, and reviewed federal guidance from the National Institute of Standards and Technology, Office of Management and Budget circulars, and guidance from leading information technology organizations. To determine whether OSC has corrected the types of data discrepancies we identified during previous work, we reviewed reports generated by OSC 2000 on the inventory of cases and interviewed knowledgeable OSC officials. Additionally, we selected a random sample of 160 cases from the 3,604 closed cases OSC received from October 1, 2004, through March 31, 2006.[6] Using these cases, we traced electronic data for selected data elements to the source case files to determine the reliability of the data. Our sample was divided into USERRA and non-USRERRA cases as we are

[4] Pub. L. No. 103-353, 108 Stat. 3149, as amended, codified at 38 U.S.C. §§ 4301-4333.

[5] GAO, *U.S. Office of Special Counsel: Strategy for Reducing Persistent Backlog of Cases Should Be Provided to Congress*, GAO-04-36 (Washington, D.C.: Mar. 8, 2004), and *U.S. Office of Special Counsel's Role in Enforcing Law to Protect Reemployment Rights of Veterans and Reservists in Federal Employment*, GAO-05-74R (Washington, D.C.: Oct. 6, 2004).

[6] This period covers the time since we last reviewed the reliability of computer-generated data at OSC.

doing additional work on the agency's USERRA activities.[7] We conducted our work in Washington, D.C., from April 2006 through December 2006 in accordance with generally accepted government auditing standards. Detailed information on our scope and methodology appears in enclosure I.

Results in Brief

Although OSC officials described actions that they said had been taken to help ensure the reliability of OSC 2000 and its related data, they did not provide us with sufficient documentation to demonstrate that fundamental system controls and safeguards are in place and operating as intended. The absence of this documentation can be attributed to OSC's failure to follow a structured system development life cycle approach for OSC 2000—an approach in which system requirements are documented, along with tests of and changes to the system. Failure to follow a structured system development life cycle approach for OSC 2000 is contrary to recognized system development life cycle management practices and increases the risk that the system will not function as intended. Controlling risks in areas such as information security is especially important to protect the personal information of complainants from inadvertent or deliberate misuse, fraudulent use, improper disclosure, corruption, or destruction.

In comparing electronic data in OSC 2000 to the source case files for 158 randomly selected cases, we found that the three data elements used in OSC's annual reports to Congress—date received, date closed, and case type—are sufficiently reliable for reporting purposes but that OSC continues to have small discrepancies in summary data provided to us similar to those previously identified during our work in 2004. These variances in the summary data are primarily caused by inconsistent queries to OSC 2000 and appear to be within OSC's acceptable error rate, which officials have stated is ± 3 percent. However, any untested data elements in OSC 2000 may be in doubt because of OSC's failure to follow structured system life cycle management practices.

We recommend in this report that OSC develop a system development life cycle approach, ensure that such an approach is fully implemented before making additional system changes, and develop consistent system queries.

[7] We removed two cases from our analysis because they involved employees who are not covered by USERRA.

We provided a copy of this report to the Special Counsel for comment. In his comments, the Special Counsel generally concurred with our recommendations. Notwithstanding this concurrence, the Special Counsel disagreed with our finding that OSC had not followed a system development life cycle approach. However, OSC provided no documentation so that we could verify the actions that the Special Counsel described OSC taking. Thus, we made no changes to our recommendation.

Background

According to OSC officials, OSC 2000 was first conceptualized in 1992 to replace OSC's aging case tracking system with a system that utilized the latest client/server relational database architecture. After awarding a contract for the design, development, and deployment of the new case tracking system, OSC decided to terminate the contract because of nonperformance. OSC's in-house information technology staff subsequently completed the OSC 2000 project without a third-party contractor's assistance. OSC 2000 went online in July 1999, 10 months after in-house staff took over the project.

In our March 2004 report, *U.S. Office of Special Counsel: Strategy for Reducing Persistent Backlog of Cases Should Be Provided to Congress*,[8] we reviewed OSC's data by case type for fiscal years 1997 through 2003. During the course of our review, we identified discrepancies, primarily in the beginning and ending inventory of cases in data the agency provided to us. To identify reasons for these discrepancies, we met with OSC's Chief Information Officer (CIO), who said that the methodology used for querying OSC 2000 had limitations, particularly in the data entry operator's use of unreviewed and unverified ad hoc queries. To provide us with accurate data, OSC's CIO developed a software program that offered a more reliable and consistent approach to querying OSC's system. We tested the accuracy and completeness of a sample of cases from OSC's system. On the basis of the results of our tests of required data elements, we determined that the data were sufficiently reliable for the purposes of our report.

In a subsequent report concerning OSC in October 2004, *U.S. Office of Special Counsel's Role in Enforcing Law to Protect Reemployment Rights*

[8] GAO-04-36.

of Veterans and Reservists in Federal Employment,[9] we reviewed OSC data on USERRA cases. During the course of our work, we learned that the number of new USERRA cases we reported in our March 2004 report for fiscal years 2000 and 2002 had been incorrect, as had the number of new USERRA cases reported by OSC in its annual report to Congress for those fiscal years.

OSC Reports Taking Actions to Help Ensure the Reliability of OSC 2000 and Its Related Data but Lacks Documentation of Its Actions

Ensuring the reliability of data produced by any computer system requires documentation and implementation of verifiable controls to ensure that these requirements are being met. OSC officials, including the CIO, provided several policies and described actions they had taken, including implementing and operating system safeguards to ensure that OSC's case tracking system produces reliable data. However, OSC could not produce sufficient documentation to provide reasonable assurance that it had taken actions or that verifiable controls to help ensure the reliability of data were in place and functioning as intended.

Actions OSC Officials Reported Taking to Ensure Reliable, Accurate, and Complete Data

According to a senior OSC official, OSC has built a number of safeguards into OSC 2000—some that operate automatically and others that operate manually by routine staff review—that are intended to ensure the reliability and accuracy of the data entered into OSC 2000 as well as the reports the system generates based on those data. For example, the official stated that most data fields will only accept data from a drop-down list programmed to a table for that field. Certain data fields have restrictions on them as to what can be typed in (e.g., a date field entry must be a valid date and not one prior to the date received). A case cannot be closed unless all the allegations connected with that case have been closed out. Users of the system cannot delete a case; this action can only be taken by the System Administrator. In addition, to safeguard against accidental deletions, allegations and certain actions cannot be deleted from the system. According to this official, most important for data integrity are the constraints built into the architecture of the system that operate automatically. These include referential restrictions (i.e., an action code cannot be entered into a case on a date before the received date; a right-to-sue letter code can only be entered in a reprisal case; and

[9] GAO-05-74R.

security restrictions are based on the identification of the staff performing the data entry so that for example, only the relevant supervisor can approve a corrective actions screen).

To ensure that data are entered correctly and consistently, OSC's CIO stated that users are held accountable for the accuracy and completeness of the data. For example, according to a data entry policy dated October 2002, all office heads are responsible for the accuracy of the data entered for matters and cases under their supervision and are to be held accountable when records are incomplete or inaccurate. Likewise, each attorney must certify in writing that the computer record for a matter or case is complete and accurate before the file can be closed. In addition, the official stated that processing steps and standard forms are used to ensure that data are entered into the system consistently. Concerning the completeness of information entered into OSC 2000, the official said that with electronic filing of complaints, more constraints are built into the electronic forms, so that if certain critical information is missing, the complainant cannot submit the form.[10] With paper filing, staff enter information into OSC 2000, and if information is missing, they have to contact the complainant.

OSC officials said that regular OSC 2000 user workgroup meetings, attended by representatives of all OSC work units, are used as a forum for raising and solving problems with the system as well as approving enhancements to it. OSC's CIO said that although OSC does not have written protocols for changes to OSC 2000, it has an established process of using the workgroup to review, approve, and test the changes. Under this process, the CIO is responsible for making minor changes to the system and documents those changes in handwritten notes. The CIO said that minor changes include adding a column or another search function. Major changes—such as the electronic filing of complaints, the use of bar codes for documents, changes to the data dictionary, and changes to the work flow diagram—must be authorized by the user workgroup. The CIO said

[10] OSC accepts complaints electronically for prohibited personnel practices (OSC Form 11) and whistleblower disclosures (OSC Form 12).

that once the workgroup approves a major change, he will design it and return to the workgroup for approval before making changes.[11]

Finally, the CIO said that while there have been no problems (e.g., system “crashes”) that would affect the quality of the data to date, OSC 2000 has both a primary and a backup system. He further stated that OSC 2000 is backed up to a backup server twice a day, so that if the primary system goes down, only 4 hours of work would be lost. OSC also has a tape backup off-site, according to the CIO.

OSC Could Not Provide Documentation to Verify the Actions It Took or the Existence of Sufficient Controls

Although OSC officials provided copies of several policies and described actions OSC has taken to ensure the quality of the data it generates, they did not provide sufficient documentation for us to verify the agency’s stated actions or that it had controls in place. For example, OSC did not provide design specifications or documentation about OSC 2000’s functional requirements. Such requirements are typically contained in a formal document that specifically describes what the system is supposed to do. As we have previously reported,[12] this detailed documentation is important because it is used for developing thorough test plans, maintaining the system, and ensuring that risks associated with building and operating the system are adequately controlled. OSC officials said that the agency does not have documentation describing the functionality of the system.

In addition, guidance from federal agencies (e.g., the National Institute of Standards and Technology) and leading information technology organizations discusses the need for organizations to adopt a structured,

[11] As a security control, only the CIO has the level of access necessary to change the code in the system. To verify any changes the CIO made to the system, the System Administrator first looked at the code to verify that it contained no mistakes, then went into the system to verify that the change the CIO made was present and to ensure that it worked properly. According to an OSC official, the System Administrator verified and tested most changes.

[12] See GAO, *Treasury Automation: Automated Auction System May Not Achieve Benefits or Operate Properly*, GAO/IMTEC-93-28 (Washington, D.C.: Apr. 27, 1993).

or System Development Life Cycle (SDLC), approach.[13] An SDLC approach requires organizations to document the phases of the development life cycle for automated information systems and their software applications, including any changes that are made to the systems or their software. As we previously reported,[14] ensuring an information system's reliability is one reason for following an SDLC approach. Federal guidance recommending that agencies follow best practices for automated information systems was issued before OSC 2000 became operational in July 1999. OSC officials did not provide documentation of an SDLC approach that would guide how OSC 2000 was defined, designed, developed, tested, implemented, and maintained. OSC provided the CIO's handwritten notes identifying problems fixed, generally by date (e.g., change the remarks field in the actions table to unlimited length) as documentation of changes to the system. According to the CIO's notes, OSC has made literally hundreds of changes to the system. OSC officials stated that they recognized the importance of an SDLC approach and would work on developing SDLC documentation.

Also, OSC did not provide documentation of the testing of changes to the system. As we previously reported, it is important that testing of an automated information system be fully documented, with traceability of test cases to the system requirements and the acceptance criteria.[15] Such traceability is not possible without functional requirements documentation. Also, without documentation, which according to guidance from a leading information technology organization should occur within the system's architecture, the history of system changes can be lost if staff changes occur, thus making future system modifications or problem corrections more time-consuming and costly. Future systems modifications are already being planned. OSC officials said that the agency

[13] An SDLC approach generally includes the following phases: (1) initiation (the recognition of a problem and the identification of a need); (2) definition (the specification of functional requirements and the start of detailed planning); (3) system design (specification of the problem solution); (4) programming and training (the start of testing, evaluation, certification, and installation of programs); (5) evaluation and acceptance (the integration and testing of the system or software); and (6) installation and operation (the implementation and operation of the system or software, the budgeting for it, and the controlling of all changes and the maintenance and modification of the system during its life).

[14] GAO, *OPM's Central Personnel Data File: Data Appear Sufficiently Reliable to Meet Most Customer Needs*, GAO/GGD-98-199 (Washington, D.C.: Sept. 30, 1998).

[15] GAO/GGD-98-199.

plans to convert OSC 2000 to a Web-based system but that such a conversion depends on funding.

Finally, as OSC accepts some complaints electronically, information security is an important consideration because, as we have previously reported,[16] the same speed and accessibility that create the enormous benefits of the computer age can, if not properly controlled, allow individuals and groups with malicious intent to intrude into inadequately protected systems and use this access to obtain sensitive information, commit fraud, disrupt operations, corrupt data, or launch attacks against other computer networks and systems. Effective information security controls affect the integrity, confidentiality, and availability of sensitive information—such as personal information on complainants—maintained by OSC.[17] These controls are essential to ensuring that information is adequately protected from inadvertent or deliberate misuse, fraudulent use, improper disclosure, corruption, or destruction. OSC officials provided a security control policy dated October 2002 from the OSC 2000 user manual that states that OSC 2000 has five "areas of security control,"[18] which touch on security controls but are not in and of themselves sufficient as verifiable controls. OSC's CIO described backing up OSC 2000 to a server twice a day and having tape backup off-site, both of which are procedures discussed in federal guidance.[19] However, OSC did not provide documentation of detailed information security controls or standards that we could review.

[16] GAO, *Information Security: Continued Progress Needed to Strengthen Controls at the Internal Revenue Service*, GAO-06-328 (Washington, D.C.: Mar. 23, 2006).

[17] Information system general controls affect the overall effectiveness and security of computer operations as opposed to being unique to any specific computer application. These controls include security management, operating procedures, software security features, and physical protections designed to ensure that access to data is appropriately restricted, only authorized changes to computer programs are made, incompatible computer-related duties are segregated, and backup and recovery plans are adequate to ensure the continuity of operations.

[18] The first area describes procedures for entering data correctly and consistently. The second states that a logon identification and password are needed to ensure that the person who logs in has the right credentials to use the system. The third discusses the user profile and security level to work in conjunction with rules, constraints, and triggers. The fourth identifies an audit trail for deleted data. The final area states that reports allow OSC officials to verify that data are entered correctly and completely.

[19] See the National Institute of Standards and Technology's Annex 1 to its Special Publication 800-53, *Recommended Security Controls for Federal Information Systems*: *Minimum Security Controls Low Baseline* (Gaithersburg, Md.: 2005), p.13.

OSC Could Not Provide Requirements for Ensuring Data Quality

The information disseminated by federal agencies is a critical strategic asset. Given the widespread use and impact of federal information, it is important for it to meet basic quality standards. In response to questions about its requirements for ensuring data quality, including the results of any reviews of the quality of the data, OSC officials did not provide sufficient documentation to provide us with reasonable assurance that they had taken certain actions. OSC provided a data entry policy, dated October 2002, which we discussed earlier, that identifies OSC's policies for ensuring that accurate and complete data are being entered into OSC 2000 and holding office heads responsible for the accuracy of the data entered into the system. The policy, however, does not provide accompanying procedures that identify which data elements are required to be entered for the computer record to ensure completeness or describe who is to conduct periodic reviews of the completeness and accuracy of the data. Without such data quality control procedures, there is no assurance that the data are complete and accurate or that OSC employees are being held accountable for the policy. It also does not describe quality control measures such as acceptable error rates or how these measures will be used to assess the quality of the data.

Selected OSC 2000 Data Are Sufficiently Reliable for Reporting to Congress, but OSC Continues to Have Data Discrepancies Similar to Those We Previously Identified

Of 11 unique selected data elements reviewed,[20] 3 are used in OSC's annual reports to Congress—date received, date closed, and case type—and are sufficiently reliable for reporting purposes. However, OSC continues to have discrepancies in summary data provided to us similar to those previously identified during our work in 2004. These small variances are primarily caused by inconsistent queries to OSC 2000 and appear to be within OSC's acceptable error rate, which officials have stated is ± 3 percent.

[20] We reviewed 11 out of 90 unique data elements in OSC 2000's data dictionary: date received, date closed, case type, case subtype, agency name, source code, action office, corrective action type, allegation, name (complainant/subject), and personnel action. However, we did not review each of these 11 data elements for both USERRA and non-USERRA cases; we discuss 5 data elements that are common to both (date received, date closed, case type, action office, and corrective action type). In addition, we reviewed 3 data elements that were unique to either USERRA cases (case subtype, agency name, and source code) or non-USERRA cases (allegation, name and personnel action).

OSC Data on Number of Cases Received and Closed by Case Type Are Sufficiently Reliable for Reporting to Congress

Three of the 11 unique selected data elements reviewed are used in OSC's annual reports to Congress—date received, date closed, and case type—and are sufficiently reliable for reporting purposes. Six of the other 8 data elements we reviewed als.o were sufficiently reliable, and another is sufficiently reliable for non-USERRA cases (i.e., those concerning prohibited personnel practices, whistleblower disclosures, and Hatch Act allegations).[21] Another data element, source code, which only applies to USERRA cases, is generally unreliable as it would match in less than 7 percent of cases. We compared electronic data in OSC 2000 to the source case files for 158 randomly selected cases received from October 1, 2004, through March 31, 2006.[22] For the purposes of this report, we assessed reliability by the amount of agreement between the data in OSC 2000 and the source case files. We did not evaluate the accuracy of the source case files for the data elements reviewed.

Our random sample for closed USERRA and non-USERRA cases was sufficient for us to comment on the reliability of selected data elements in closed cases in OSC 2000 (see enc. I for a discussion of our sampling methodology).[23] We excluded cases from the original sample size (i.e., 64 USERRA and 94 non-USERRA cases) when information was missing from the case file and prevented the comparison of data in OSC 2000 to the source case files for a particular data element. For data elements pertaining to time (i.e., date received and date closed for both USERRA and non-USERRA cases), we did not include differences between the data in OSC 2000 and the case files when they were off by 1 day; only differences of more than 1 day were included. (See enc. II for the results of our file review for all data elements reviewed.)

For date received in USERRA cases, there is at least a 95 percent chance that a match will occur between OSC 2000 and the case files in more than

[21] That data element, corrective action, is not sufficiently reliable for USERRA cases (i.e., would expect a match between OSC 2000 and the case file in more than 76 percent of cases).

[22] We removed 2 cases from our analysis of 160 cases because they involved employees not covered by USERRA.

[23] Because we sampled a portion of OSC closed cases, our results are estimates of all closed cases and are subject to sampling error. For example, we are 95 percent confident that for USERRA cases the date closed data element has a match rate of at least 92 percent. This one-sided confidence interval indicates that there is at least a 95 percent chance that a match will occur between OSC 2000 and the case files for this data element in USERRA cases in at least 92 percent of the cases.

81 percent of cases and for non-USERRA cases in more than 83 percent of cases. For USERRA cases, date received could be verified in OSC 2000 data for 63 cases where information was present in the case file; the average difference between the case file and OSC 2000 was less than 1 day. Of those 63 cases, 7 were off by more than 1 day. Across the 94 non-USERRA cases, the average difference between OSC 2000 and the case file in the date received data element was about ±2 days. Of those 94 cases, 10 cases were off by more than 1 day. The small average difference in days for date received combined with the matching rate between OSC 2000 and the case files demonstrates sufficient data reliability for using date received to report the number of new cases to Congress.

For date closed in USERRA cases, there is at least a 95 percent chance that a match will occur between OSC 2000 and the case files in more than 92 percent of cases and for non-USERRA cases in more than 91 percent of cases. For USERRA cases, date closed could be verified in OSC 2000 data for 63 cases where information was present in the case file; the average difference between the case file and OSC 2000 was less than 1 day. Of those 63 cases, 1 was off by more than 1 day. For non-USERRA cases, date closed could be verified in OSC 2000 data for 93 cases where information was present in the case file; the average difference between OSC 2000 and the case file was less than 1 day, and no cases were off by more than 1 day. The small average difference in days for date closed combined with the matching rate between OSC 2000 and the case files demonstrates sufficient data reliability for using date received to report the number of new cases to Congress.

For case type in USERRA cases, there is at least a 95 percent chance that a match will occur between OSC 2000 and the case files in more than 96 percent of cases and for non-USERRA cases in more than 95 percent of cases. Case type could be verified in OSC 2000 data for all 64 USERRA cases reviewed and for all 94 non-USERRA cases reviewed.

OSC Continues to Have Small Data Discrepancies Similar to Those We Identified in Prior Work

As discussed earlier, we previously identified data discrepancies, primarily in the beginning and ending inventory of cases by fiscal year during our work for our March 2004 report.[24] According to OSC's CIO, since our work in 2004, OSC has developed a standard structured query language (SQL)[25] methodology as one of its management tools to query OSC 2000 for caseload information. Although we did not test the SQL code on OSC 2000 to determine whether it worked as intended, OSC provided us with data for the beginning and ending inventory of cases for fiscal years 2004 and 2005. These data showed (by type of case) cases pending at the beginning of both fiscal years, cases received for both years, and cases closed for both years. For cases concerning prohibited personnel practices, whistleblower disclosures, and the Hatch Act, these data showed small discrepancies between the numbers of cases as determined by the SQL methodology and as determined by data recorded in OSC 2000.

In addition, in providing USERRA data primarily for fiscal years 2005 and 2006 for another ongoing GAO engagement, we found that when OSC queried OSC 2000 by case identification number, which is to include the fiscal year (e.g., "05" or "06") in which the data were entered into the system, the result for fiscal year 2005 was 109 cases. When OSC later queried OSC 2000 by date received (i.e., "date received between 10/1/2004 and 9/30/2005"), the result for fiscal year 2005 was 111, because 2 cases that were received near the end of the fiscal year had been entered at the start of the new fiscal year and received case identification numbers of "06." Thus querying by date produced a different result than querying by case identification number, indicating a lack of standardized SQL queries. A senior OSC official acknowledged another instance where system queries produced different results, and OSC officials agreed that they needed to use standardized SQL queries and said that OSC was working on correcting the problem as part of its computer system upgrades. In these instances, it is not clear that OSC information technology or management officials reviewed the data for accuracy.

Conclusion

Agencies that follow structured system development life cycle practices are able to demonstrate that fundamental system controls and safeguards are in place and operating as intended. Because OSC has not followed

[24] GAO-04-36.

[25] SQL is a popular computer language used to create, modify, retrieve, and manipulate data from relational database management systems.

such a structured approach, the reliability of its case tracking system is in question, and the risks increase that personal information in complaints captured in that system could be vulnerable to inadvertent or deliberate misuse, fraudulent use, improper disclosure, corruption, or destruction. Although the continued discrepancies we found in computer data generated by OSC 2000 were within OSC's acceptable error rate, without the use of consistent SQL queries, OSC does not have assurance that future discrepancies will remain within the acceptable rate. As the agency is planning future system modifications, OSC has an opportunity to implement relevant federal guidance for information technology systems.

Recommendations for Executive Action

We recommend that the Special Counsel direct OSC's CIO to take the following actions:

- Define an SDLC approach that is consistent with relevant federal guidance and practices at successful information technology organizations, including the minimum security requirements outlined in National Institute of Standards and Technology guidance.
- Ensure that the SDLC is fully implemented as part of any planned changes to or replacements for OSC 2000.
- Develop and utilize consistent standard SQL queries for reporting on the inventory of cases.

Agency Comments and Our Evaluation

We provided a draft of this report to the Special Counsel for his review and comment. In written comments, the Special Counsel stated that he fully concurred that formal systems documentation needs to be updated before redesigning and redeveloping OSC 2000 to be Web enabled and to develop consistent queries to reduce data discrepancies. Notwithstanding this concurrence, the Special Counsel also stated that he disagreed with our finding that OSC had not followed structured life cycle management practices for the development of OSC 2000 and went on to describe tests that he said OSC had run on the system. However, OSC's inability to produce documentation of the phases of OSC 2000's development life cycle or of the testing of changes to the system precluded us from verifying the actions that the Special Counsel described OSC as having taken. It is widely understood that documentation of life cycle management activities is as important as the actual execution of those activities. Further, without such documentation, OSC will likely find making future system modifications or problem corrections more time-consuming and costly than it would with adequate documentation. We believe that the report accurately reflects what we found and were able to verify. As such, we continue to believe that OSC needs to define and document a structured life cycle management approach that includes the minimum requirements outlined in National Institute of Standards and Technology guidance. A copy of OSC's written response is included in enclosure III.

We will send copies of this report to the Special Counsel, interested congressional committees, and other interested parties. Copies will be made available to others upon request. This report will also be available at no charge on GAO's Web site at http://www.gao.gov.

If you or your staff have questions about this report, please contact me on (202) 512-9490 or by e-mail at stalcupg@gao.gov. Contact points for our Offices of Congressional Relations and Public Affairs may be found on the last page of this report. Key contributors to this report are listed in appendix IV.

George H Stalcup

George H. Stalcup
Director, Strategic Issues

Enclosures

Enclosure I: Scope and Methodology

Actions OSC Has Taken to Help Ensure the Reliability of Its Case Tracking System and Related Data

To identify what actions the Office of Special Counsel (OSC) has taken to help ensure the reliability of its case tracking system and related data, we reviewed existing information about the data and the system that produced them and interviewed knowledgeable OSC officials, including OSC's Chief Information Officer and System Administrator, using GAO's standard interview questions related to data reliability assessments. We also reviewed prior GAO reports, National Institute of Standards and Technology guidance, Office of Management and Budget circulars, and guidance from leading information technology organizations.

Determination of Whether OSC Has Corrected the Types of Data Discrepancies We Identified during Previous Work

To determine whether OSC has corrected the types of data discrepancies we identified during previous work, we reviewed relevant documentation and interviewed knowledgeable OSC officials. We also traced electronic data for selected data elements to the source case files.

Comparison of Electronic Data to the Source Case Files

To compare electronic data to the source case files, we first selected which data elements to include in our review. We selected the specific data elements for review by asking knowledgeable OSC officials to identify data elements critical for processing claims filed by federal employees under the Uniformed Services Employment and Reemployment Rights Act of 1994 (USERRA) and non-USERRA cases (i.e., prohibited personnel practices, whistleblower disclosures, and Hatch Act allegations).[1] Of the 90 unique data elements that appear in OSC's data dictionary,[2] we selected 11 data elements to review: date received, date closed, case type, case subtype, agency name, source code, action office, corrective action type, allegation, name (complainant/subject), and personnel action.

For USERRA cases, all but one of these data elements, action office, was identified by OSC officials as critical to processing a case. For the non-

[1] Our sample was divided into USERRA and non-USERRA cases, as we are doing additional work on USERRA at OSC.

[2] We counted remarks as a single data element; remarks might have contained different content in different tables.

USERRA data elements, because we had so few data elements that were considered critical for processing a case, we included three data elements—personnel action code, action office, and corrective action type—that an OSC official said were not necessarily comparable between the case file and OSC 2000. We included these three to confirm that they were not comparable (i.e., that we would not find them present in either the electronic data or the case file or both).

We did not review each of these 11 data elements for both USERRA and non-USERRA cases; we reviewed 5 data elements that were common to both (date received, date closed, case type, action office, and corrective action type). In addition, we reviewed three data elements that were unique to either USERRA cases (case subtype, agency name, and source code) or non-USERRA cases (allegation, name, and personnel action).

Our review focused on a randomly selected sample of 160 cases from the 3,604 closed cases OSC received from October 1, 2004, through March 31, 2006, as it covers the period since we last reviewed the reliability of computer-generated data at OSC.[3] Of the 3,604 closed cases, 175 were USERRA cases and 3,429 were non-USERRA. Our randomly selected sample was for 66 USERRA and 94 non-USERRA cases or 160 cases. Although we compared electronic data in OSC 2000 to the case files for all 160 cases in our sample, we removed from our analysis 2 cases from our original sample of 66 USERRA cases, leaving 64, because we learned that they were filed by Transportation Security Administration security screeners and supervisory security screeners, who are not covered by USERRA—for a total of 158 cases in our sample.[4] In addition, for each data element, we excluded cases if information was missing from the case file, thus preventing a comparison between data in OSC 2000 and the case file.

[3] GAO, *U.S. Office of Special Counsel: Strategy for Reducing Persistent Backlog of Cases Should Be Provided to Congress*, GAO-04-36 (Washington, D.C.: Mar. 8, 2004), and *U.S. Office of Special Counsel's Role in Enforcing Law to Protect Reemployment Rights of Veterans and Reservists in Federal Employment*, GAO-05-74R (Washington, D.C.: Oct. 6, 2004).

[4] Transportation Security Administration security screeners and supervisory security screeners are not covered by USERRA. See, *Spain v. Department of Homeland Security*, 99 M.S.P.R. 529 (2005), citing to *Conyers v. M.S.P.B*, 388 F3d. 1380 (Fed. Cir. 2004). The Transportation Security Administration, however, voluntarily permits OSC to investigate USERRA claims.

For data elements pertaining to time (i.e., date received and date closed for both USERRA and non-USERRA cases), we also did not include differences between the data in OSC 2000 and the case files when they were off by 1 day, only differences of more than 1 day.

For the purposes of this report, we assessed reliability by the amount of agreement between the data in OSC 2000 and the source case files. We did not evaluate the accuracy of the source case files for the data elements reviewed.

We conducted our work in Washington, D.C., from April 2006 through December 2006 in accordance with generally accepted government auditing standards.

Enclosure II: Review of Selected Data Elements from OSC 2000 to Determine the Reliability of the Data

We compared electronic data for 11 selected data elements in OSC 2000 to the source case files for 158 randomly selected closed cases received from October 1, 2004, through March 31, 2006. For the purposes of this report, we assessed reliability by the amount of agreement between the data in OSC 2000 and the source case files.[1] Our random sample for USERRA and non-USERRA cases was sufficient for us to comment on the reliability of the 11 selected data elements for closed cases in OSC 2000 (see enc. I for a discussion of our sampling methodology).[2]

Data Elements Reviewed in USERRA Cases

For USERRA cases, we reviewed the following data elements: date received, date closed, agency name, case type, case subtype, source code, action office, and corrective action type. We excluded cases by data element from the original sample size of 64 cases when information was missing from the case file. Of the 64 cases reviewed, 5 were missing information from the source case file for at least one of the data elements. For one data element reviewed for USERRA cases, source code, data were not sufficiently reliable because of a high degree of incompleteness in OSC 2000 (i.e., OSC 2000 generally did not contain the data). Another data element, corrective action, is not sufficiently reliable for USERRA cases (i.e., would expect a match between OSC 2000 and the case file in more than 76 percent of cases) but is sufficiently reliable for non-USERRA cases (would expect a match between OSC 2000 and the case file in more than 89 percent of cases). Table 1 shows a breakdown of the eight data elements reviewed for USERRA cases by sample size, matches between the case file and OSC 2000, and the percentage to which we are 95 percent confident of such a match for all USERRA cases.

[1] We did not evaluate the accuracy of the source case files for the data elements reviewed.

[2] Because we sampled a portion of OSC cases, our results are estimates of closed cases and are subject to sampling error. For example, in table 1, we are 95 percent confident that for USERRA cases the date closed data element has a match rate of at least 92 percent. This one-sided confidence interval indicates that there is at least a 95 percent chance that a match will occur between OSC 2000 and the case files for this data element in USERRA cases in at least 92 percent of cases.

Table 1: Breakdown of Data Elements Reviewed for USERRA Cases

Data element	Sample size[a]	Matches between the case file and OSC 2000		One-sided 95 percent confidence interval for percentage matching[b]
		Number	Percent	
Date received	63	56	89	>81
Date closed	63	61	97	>92
Case type	64	64	100	>96
Case subtype[c]	50	48	96	>89
Agency name	64	64	100	>96
Source code	63	1	2	<7
Action office	63	62	98	>93
Corrective action type	63	53	84	>76

Source: GAO analysis of OSC 2000 data.

[a] We generally excluded cases from the original sample size of 64 cases when information was missing from the case file, which accounts for any differences from 64.

[b] We are 95 percent confident that a match will occur between OSC 2000 and the case files for data elements in closed USERRA cases more or less than the percentage shown.

[c] USERRA cases are either referral or demonstration project cases, and only demonstration project cases have case subtypes. Thus for case subtype, we excluded 11 cases that were referral cases for which case subtype is not a relevant data element. We also excluded 3 cases because information was missing from the case files.

Two of the USERRA data elements concerned time: date received and date closed. Date received could be verified in OSC 2000 data for 63 cases where information was present in the case file; the average difference between the case file and OSC 2000 was less than 1 day. Of those 63 cases, 7 were off by more than 1 day.[3] The greatest difference in date received between the case file and the date in OSC 2000 was 14 days. Similarly, date closed could be verified in OSC 2000 data for 63 cases where information was present in the case file; the average difference between the case file and OSC 2000 was less than 1 day. Of those 63 cases, 1 was off by more than 1 day, and the difference in date closed between the case file and the date in OSC 2000 was 6 days.

[3] We excluded cases that were off by 1 day for date received, because that difference could include a case that was received late one afternoon and entered the following morning.

Non-USERRA Data Elements Reviewed

For non-USERRA case types, we reviewed the following data elements: date received, date closed, allegations, name (complainant/subject),[4] case type, personnel action, action office, and corrective action type. We excluded cases from the original sample size of 94 cases when information was missing from the case file. Of the 94 cases reviewed, 15 were missing information from the source case file, and we excluded 24 others because they were whistleblower disclosure and Hatch Act cases, which generally do not contain personnel actions.[5] Table 2 shows a breakdown of the eight data elements reviewed for non-USERRA cases by sample size, matches between the case file and OSC 2000, and the percentage to which we are 95 percent confident of such a match for all USERRA cases.

Table 2: Breakdown of Data Elements Reviewed for Non-USERRA Cases

		Matches between the case file and OSC 2000		
Data element	Sample size[a]	Number	Percentage	One-sided 95 percent confidence interval for percentage matching[b]
Date received	94	84	89	>83
Date closed	93	89	96	>91
Allegation	94	93	99	>95
Name (Complainant/subject)	94	94	100	>97
Case type	94	93	99	>95
Personnel action[c]	55	53	96	>89
Action office	91	89	98	>93
Corrective action type	93	88	95	>89

Source: GAO analysis of OSC 2000 data.

[a] We generally excluded cases from the original sample size of 94 cases when information was missing from the case file, which accounts from any differences from 94.

[b] We are 95 percent confident that a match will occur between OSC 2000 and the case files for data elements in closed non-USERRA cases more or less than the percentage shown.

[4] For prohibited personnel practices and whistleblower disclosures, there would be a complainant filing a claim or making a disclosure, whereas with a Hatch Act case the focus of the allegation would be the subject.

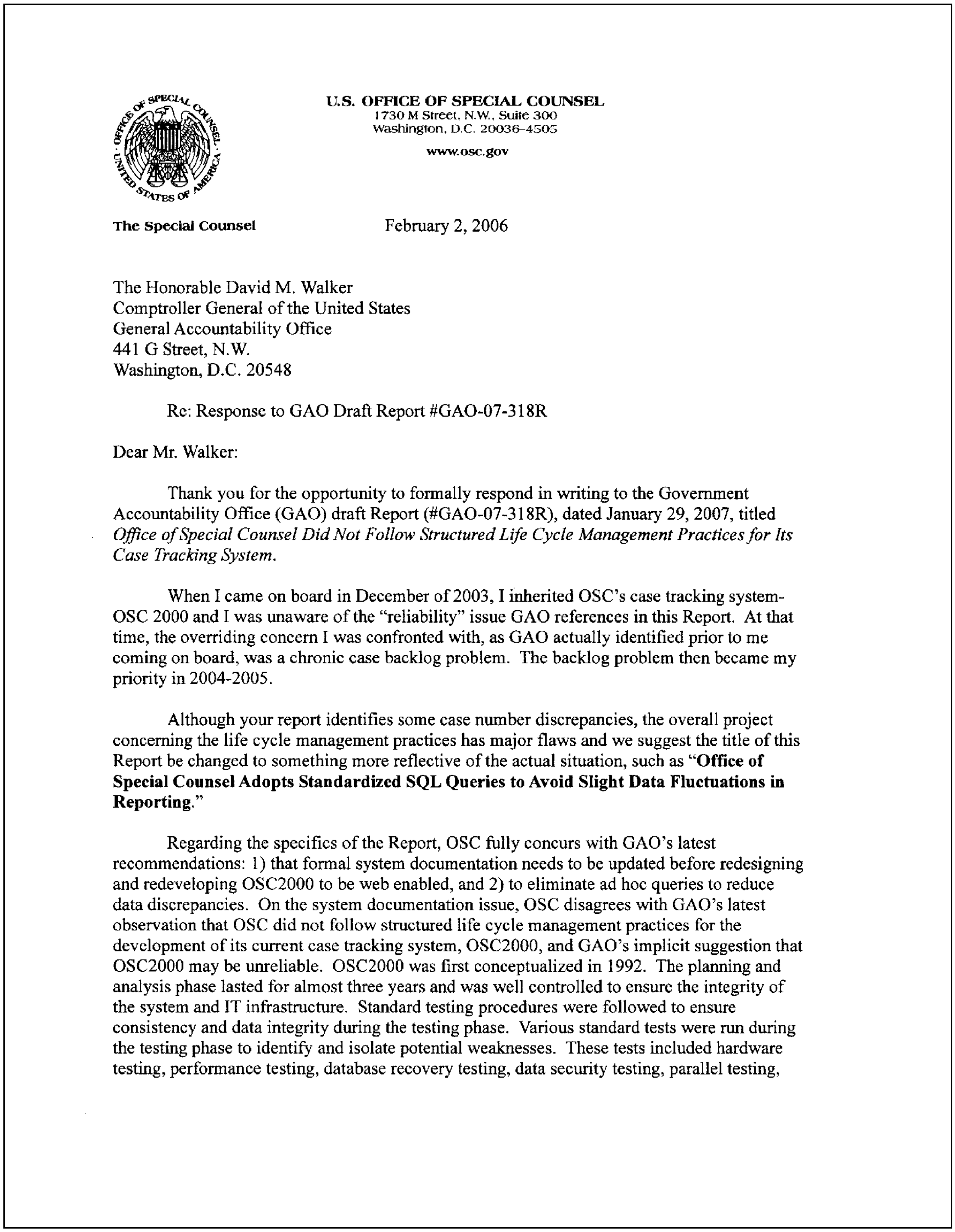
[5] One whistleblower disclosure case included information in the case file that we could verify with data in OSC 2000.

[c] For personnel action code, of the 94 cases, we excluded a total of 39 from the sample. We excluded 24 cases because they were whistleblower disclosure or Hatch Act cases, which generally do not contain personnel actions, although 1 whistleblower disclosure case included information in the case file to verify data in OSC 2000. We excluded 15 others because they were missing information in the case file.

As in the USERRA cases, two data elements concerned time: date received and date closed. Across the 94 non-USERRA cases, the average difference between OSC 2000 and the case file in the date received data element was about ±2 days. Of those 94 cases, 10 cases were off by more than 1 day.[6] The greatest difference in date received between the case file and the date in OSC 2000 was at least 86 days. Date closed could be verified in OSC 2000 data for 93 cases where information was present in the case file; the average difference between OSC 2000 and the case file was less than 1 day, and no cases were off by more than 1 day.

[6] We excluded cases that were off by 1 day for date received, because that difference could include a case that was received late one afternoon and entered the following morning.

Enclosure III: Comments from the Office of Special Counsel



U.S. OFFICE OF SPECIAL COUNSEL
1730 M Street, N.W., Suite 300
Washington, D.C. 20036-4505

www.osc.gov

The Special Counsel

February 2, 2006

The Honorable David M. Walker
Comptroller General of the United States
General Accountability Office
441 G Street, N.W.
Washington, D.C. 20548

Re: Response to GAO Draft Report #GAO-07-318R

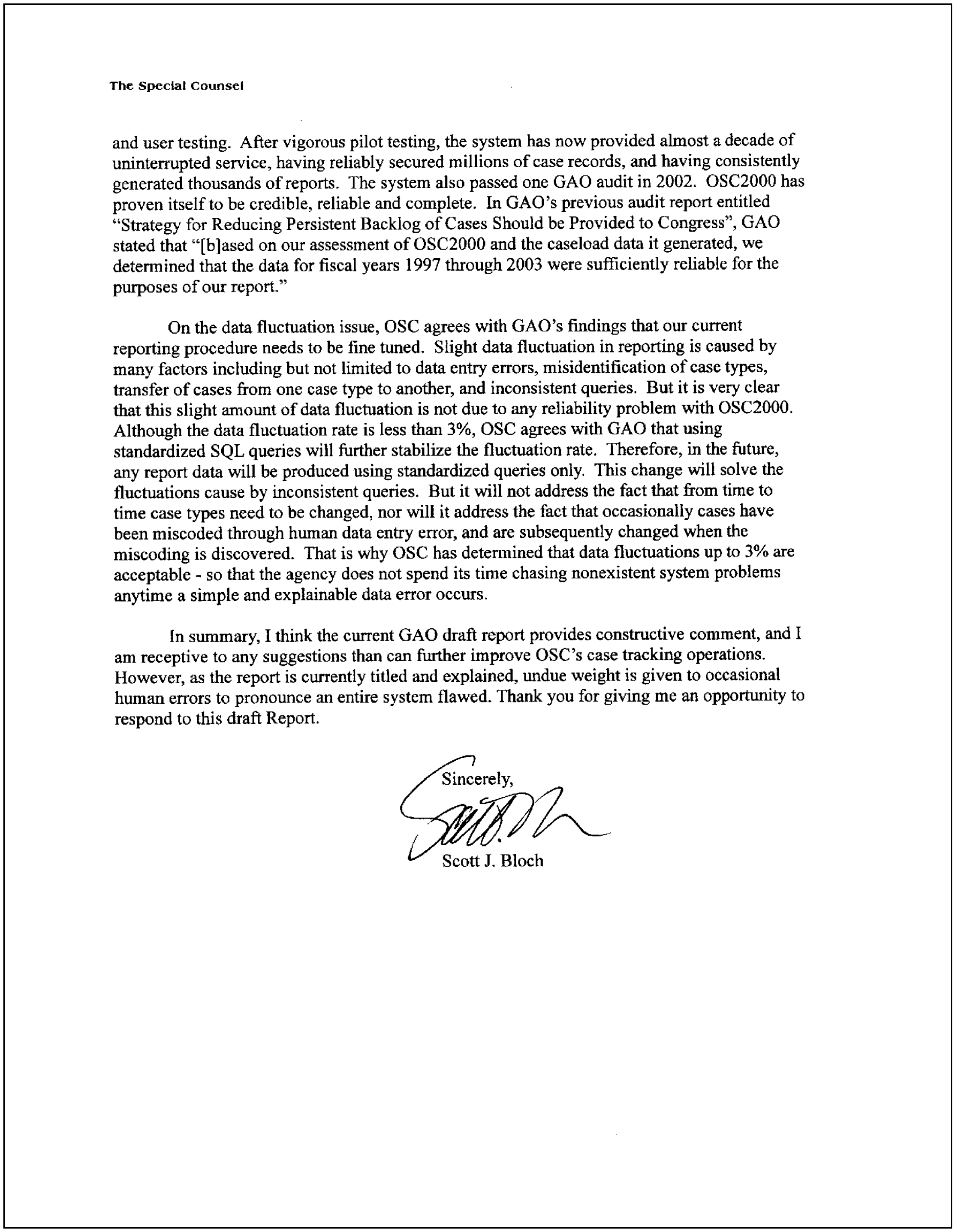
Dear Mr. Walker:

Thank you for the opportunity to formally respond in writing to the Government Accountability Office (GAO) draft Report (#GAO-07-318R), dated January 29, 2007, titled *Office of Special Counsel Did Not Follow Structured Life Cycle Management Practices for Its Case Tracking System.*

When I came on board in December of 2003, I inherited OSC's case tracking system-OSC 2000 and I was unaware of the "reliability" issue GAO references in this Report. At that time, the overriding concern I was confronted with, as GAO actually identified prior to me coming on board, was a chronic case backlog problem. The backlog problem then became my priority in 2004-2005.

Although your report identifies some case number discrepancies, the overall project concerning the life cycle management practices has major flaws and we suggest the title of this Report be changed to something more reflective of the actual situation, such as "**Office of Special Counsel Adopts Standardized SQL Queries to Avoid Slight Data Fluctuations in Reporting.**"

Regarding the specifics of the Report, OSC fully concurs with GAO's latest recommendations: 1) that formal system documentation needs to be updated before redesigning and redeveloping OSC2000 to be web enabled, and 2) to eliminate ad hoc queries to reduce data discrepancies. On the system documentation issue, OSC disagrees with GAO's latest observation that OSC did not follow structured life cycle management practices for the development of its current case tracking system, OSC2000, and GAO's implicit suggestion that OSC2000 may be unreliable. OSC2000 was first conceptualized in 1992. The planning and analysis phase lasted for almost three years and was well controlled to ensure the integrity of the system and IT infrastructure. Standard testing procedures were followed to ensure consistency and data integrity during the testing phase. Various standard tests were run during the testing phase to identify and isolate potential weaknesses. These tests included hardware testing, performance testing, database recovery testing, data security testing, parallel testing,



The Special Counsel

and user testing. After vigorous pilot testing, the system has now provided almost a decade of uninterrupted service, having reliably secured millions of case records, and having consistently generated thousands of reports. The system also passed one GAO audit in 2002. OSC2000 has proven itself to be credible, reliable and complete. In GAO's previous audit report entitled "Strategy for Reducing Persistent Backlog of Cases Should be Provided to Congress", GAO stated that "[b]ased on our assessment of OSC2000 and the caseload data it generated, we determined that the data for fiscal years 1997 through 2003 were sufficiently reliable for the purposes of our report."

On the data fluctuation issue, OSC agrees with GAO's findings that our current reporting procedure needs to be fine tuned. Slight data fluctuation in reporting is caused by many factors including but not limited to data entry errors, misidentification of case types, transfer of cases from one case type to another, and inconsistent queries. But it is very clear that this slight amount of data fluctuation is not due to any reliability problem with OSC2000. Although the data fluctuation rate is less than 3%, OSC agrees with GAO that using standardized SQL queries will further stabilize the fluctuation rate. Therefore, in the future, any report data will be produced using standardized queries only. This change will solve the fluctuations cause by inconsistent queries. But it will not address the fact that from time to time case types need to be changed, nor will it address the fact that occasionally cases have been miscoded through human data entry error, and are subsequently changed when the miscoding is discovered. That is why OSC has determined that data fluctuations up to 3% are acceptable - so that the agency does not spend its time chasing nonexistent system problems anytime a simple and explainable data error occurs.

In summary, I think the current GAO draft report provides constructive comment, and I am receptive to any suggestions than can further improve OSC's case tracking operations. However, as the report is currently titled and explained, undue weight is given to occasional human errors to pronounce an entire system flawed. Thank you for giving me an opportunity to respond to this draft Report.

Sincerely,

Scott J. Bloch

Enclosure IV: GAO Contact and Staff Acknowledgements

GAO Contact

George H Stalcup, (202) 512-9490 or stalcupg@gao.gov..

Staff Acknowledgements

In addition to the individual named above, Belva M. Martin, Assistant Director; Karin Fangman, David Fox, Randy Hite, Kiki Theodoropoulos, Jason Vassilicos, and Greg Wilmoth made key contributions to this report.

(450484)

GAO's Mission

The Government Accountability Office, the audit, evaluation and investigative arm of Congress, exists to support Congress in meeting its constitutional responsibilities and to help improve the performance and accountability of the federal government for the American people. GAO examines the use of public funds; evaluates federal programs and policies; and provides analyses, recommendations, and other assistance to help Congress make informed oversight, policy, and funding decisions. GAO's commitment to good government is reflected in its core values of accountability, integrity, and reliability.

Obtaining Copies of GAO Reports and Testimony

The fastest and easiest way to obtain copies of GAO documents at no cost is through GAO's Web site (www.gao.gov). Each weekday, GAO posts newly released reports, testimony, and correspondence on its Web site. To have GAO e-mail you a list of newly posted products every afternoon, go to www.gao.gov and select "Subscribe to Updates."

Order by Mail or Phone

The first copy of each printed report is free. Additional copies are $2 each. A check or money order should be made out to the Superintendent of Documents. GAO also accepts VISA and Mastercard. Orders for 100 or more copies mailed to a single address are discounted 25 percent. Orders should be sent to:

U.S. Government Accountability Office
441 G Street NW, Room LM
Washington, D.C. 20548

To order by Phone:	Voice:	(202) 512-6000
	TDD:	(202) 512-2537
	Fax:	(202) 512-6061

To Report Fraud, Waste, and Abuse in Federal Programs

Contact:

Web site: www.gao.gov/fraudnet/fraudnet.htm
E-mail: fraudnet@gao.gov
Automated answering system: (800) 424-5454 or (202) 512-7470

Congressional Relations

Gloria Jarmon, Managing Director, JarmonG@gao.gov (202) 512-4400
U.S. Government Accountability Office, 441 G Street NW, Room 7125
Washington, D.C. 20548

Public Affairs

Paul Anderson, Managing Director, AndersonP1@gao.gov (202) 512-4800
U.S. Government Accountability Office, 441 G Street NW, Room 7149
Washington, D.C. 20548

PRINTED ON RECYCLED PAPER