

GAO

Report to Chairman, Subcommittee on
Civil Service, Committee on Government
Reform and Oversight, House of
Representatives

September 1998

OPM'S CENTRAL PERSONNEL DATA FILE

Data Appear Sufficiently Reliable to Meet Most Customer Needs





United States
General Accounting Office
Washington, D.C. 20548

General Government Division

B-276095

September 30, 1998

The Honorable John L. Mica
Chairman, Subcommittee on Civil Service
Committee on Government Reform and Oversight
House of Representatives

Dear Mr. Chairman:

This report presents the results of our review of the Office of Personnel Management's database of federal civilian employees, the Central Personnel Data File, which we undertook as part of our basic legislative authority. Because of your continuing interest in the accuracy of this database, you asked that we address this report to you. We know that you and other decisionmakers use the information contained in this database to track statistics on federal employees, and agencies' compliance with governmentwide policies. Because this database is the primary source of information on federal employees, the accuracy of its information is critical for analyses of the federal civilian workforce.

We are sending copies of this report to other appropriate congressional committees and executive branch agencies, including the Ranking Minority Member of the Subcommittee on Civil Service, House Committee on Government Reform and Oversight; the Chairman and Ranking Minority Member of the Subcommittee on International Security, Proliferation and Federal Services, Senate Committee on Governmental Affairs; and the Director of the Office of Personnel Management.

This report was prepared under the direction of Michael Brostek, Associate Director, Federal Management and Workforce Issues, who may be reached on (202) 512-8676 if you have any questions. Major contributors are listed in appendix VIII.

Sincerely yours,

A handwritten signature in cursive script that reads 'L. Nye Stevens'.

L. Nye Stevens
Director, Federal Management
and Workforce Issues

Executive Summary

Purpose

Each year, hundreds of federal personnel offices process millions of personnel actions, such as pay adjustments and promotions, that affect the working lives of about 1.9 million federal civilian employees. The Office of Personnel Management (OPM) collects data on these personnel actions and processes them through its Central Personnel Data System (the System) for entry into its personnel database, the Central Personnel Data File (CPDF). Policymakers use CPDF data for such things as obtaining statistics on federal employees, ensuring agencies' compliance with governmentwide policies, and making decisions on federal personnel policy. Researchers use them in studies of the federal workforce.

In spite of the important uses of CPDF data, no independent evaluation of the accuracy of the data has been done. Because GAO and others rely on CPDF data to do governmentwide and agency evaluations, GAO undertook a review of the CPDF and the Central Personnel Data System as part of its basic legislative authority. GAO's objectives were to determine (1) the extent to which selected CPDF data elements are accurate, including the data elements used by OPM's Office of the Actuaries for estimating the government's liability for future payments of federal retirement programs; (2) whether selected users of CPDF data believed that CPDF products met their needs, including whether the products were current, accurate, and complete and whether the cautions OPM provided to them on the limitations associated with using the data were sufficient for them to present the CPDF data correctly; and (3) whether OPM has documented changes to the System and verified the System's acceptance of those changes, as recommended in applicable federal guidance, and whether the System would implement CPDF edits as intended.

Background

The CPDF is a database that contains individual records for most federal employees and is the primary governmentwide source for information on federal employees. The records are made up of data elements, such as name, pay plan, and veterans status. As of February 1998, OPM organized data in the CPDF into 95 separate data elements. OPM provides guidelines, edit standards, and feedback to agencies on when and how to submit data to the CPDF and what they can do to improve the data they are submitting. CPDF edits are computer instructions that are designed to check the validity of individual data elements. For example, the edit for the sex data element checks that the character that is used to define the data element is either "M" for male or "F" for female; the edit identifies other characters as errors. After agencies submit their data, OPM uses the same CPDF edits it expects agencies to use and other analyses to check that data submissions

meet edit standards. When data submissions do not meet the CPDF edit standards, OPM staff are to contact agencies to ask them to correct problem data.

Results in Brief

OPM does not have an official standard for the desired accuracy of CPDF data elements. On a periodic basis, however, OPM measures CPDF data accuracy by comparing certain data found in a sample of former federal employees' official personnel folders to data in the CPDF for the same period. OPM generally makes the results of its measurements of CPDF data accuracy available to users of CPDF data within OPM but not to non-OPM users. Although the accuracy of the CPDF data GAO reviewed varied by data element, about two-thirds of the selected CPDF data elements GAO reviewed were 99-percent or more accurate. Among the data elements that were 99-percent accurate were those OPM's Office of the Actuaries uses for estimating the government's liability for future retirement payments to federal employees and their survivors, with the exception of one element—adjusted basic pay—that was about 94-percent accurate according to one GAO measurement method. Prior GAO work also showed that CPDF data accuracy also varied by agency.

GAO surveyed all the requesters of CPDF products that OPM identified as obtaining data directly from OPM for fiscal year 1996. Most of these CPDF users reported that CPDF products met their needs, including the data being current, accurate, and complete. The majority of surveyed users reported that they believed that the caution statements OPM provided were sufficient for them to use CPDF data correctly. However, OPM did not provide these users of CPDF data with all 28 cautions that explain how CPDF limitations could affect how they present or use CPDF data. For example, some CPDF data (e.g., education level) are collected at the time of the appointment and not routinely updated. Some users said that they would have presented or used CPDF data differently if they had known about all 28 caution statements.

Although applicable federal guidance recommended that agencies document the life cycle of an automated information system from its initiation through installation and operation, OPM did not document changes that it made to the System in 1986 when it did a major redesign of the System's software. OPM also did not have documentation to show that acceptance testing of those changes was done and, according to OPM, the testing was not done by an independent reviewer. However, OPM officials said that to their knowledge the System has not had problems processing

data reliably. GAO's review of the computer instructions for most CPDF edits used by the System showed that the System uses instructions that should implement the CPDF edits reviewed as intended. OPM officials acknowledged that for OPM to accomplish its future information technology goals it will have to follow an approach that includes documenting the development, modification, and management of its automated information systems and their software applications. OPM has committed to adopting this approach by no later than fiscal year 2002.

GAO is making recommendations to improve the availability of information about CPDF data accuracy and limitations and to ensure that future changes to the Central Personnel Data System are documented and independently verified.

Principal Findings

Selected Data Appear to Be Mostly Accurate in the CPDF, but OPM Does Not Report Results of Its Accuracy Measurements to Non-OPM Users

Although OPM screens data before accepting them, OPM relies on agencies to submit accurate data. Thus, the ultimate accuracy of CPDF data depends on the accuracy of the data that agencies submit. Errors in those data can occur at various stages of the personnel process, such as when agency personnel enter (1) data for newly hired employees; or (2) information on personnel actions (e.g., performance appraisals) for submission to OPM. OPM does not have an official accuracy standard for agencies' submissions. On a periodic basis, however, OPM measures CPDF data accuracy by comparing certain data found in a sample of former federal employees' official personnel folders to data in the CPDF for the same period. OPM generally makes the results of its measurements of CPDF data accuracy available to users of CPDF data within OPM but not to non-OPM users.

To measure the accuracy of CPDF data for its review, GAO asked a generalizable sample of current federal employees to verify CPDF data as of September 30, 1996 pertaining to them.¹ GAO also compared data in the official personnel folders and other records of a nongeneralizable sample of current federal employees selected from six of the largest personnel offices to fiscal year 1996 CPDF data.² Between the two methods GAO used

¹The complete results of GAO's survey appear in appendix V.

²Because they are among the eight largest personnel offices in the federal government, for its review GAO selected personnel offices at the Social Security Administration, Baltimore, MD; Department of the Army, Fort Benning, GA; U.S. Customs Service, Washington, D.C.; National Institutes of Health, Bethesda, MD; Department of State, Washington, D.C.; and Department of the Navy, Pensacola, FL.

to measure CPDF data accuracy, variation existed in the accuracy of some data elements, but at least 63 percent of CPDF data elements in both samples were 99 percent or more accurate. The least accurate data element, education level, was about 73 and 77 percent accurate according to GAO's two measurement methods. These results were broadly consistent with OPM's latest accuracy measurement.

Although both OPM's and GAO's reviews showed that most CPDF data elements reviewed were 99 percent or more accurate on a governmentwide basis, neither OPM nor GAO measured the accuracy of data for individual agencies. However, GAO's prior work has shown that specific data elements for individual agencies can be much less accurate. For example, in 1997 the House Committee on International Relations asked GAO to examine a discrepancy between the number of certain political appointees that the Agency for International Development (AID) reported to Congress (17) and the number that appeared in the CPDF for the period January 19, 1993, through November 14, 1995, (0). Through its analysis of the CPDF data, GAO determined that AID misidentified the legal authority that was used to appoint these individuals; as a result, the information in the CPDF did not correctly identify any of the 17 individuals as political appointees.

To estimate the government's liability for future retirement payments, OPM's Office of the Actuaries uses CPDF data on adjusted basic pay, sex, birth date, retirement plan, and service computation date. Except for adjusted base pay, which was about 94-percent accurate in GAO's nongeneralizable accuracy measurement, GAO found all of these data to be about 99-percent accurate. GAO shared these results with the actuary responsible for calculating the federal government's liability for future retirement payments to federal employees and their survivors, and he said that the CPDF data elements that the Office used were sufficiently accurate for making this liability estimate.

Despite the high governmentwide accuracy GAO and OPM found for selected CPDF data elements, the lower level of accuracy for some individual data elements could affect the validity of studies relying on such data. For instance, GAO's finding that data on federal employee education levels are about 23- and 27-percent inaccurate in its generalizable and nongeneralizable measurements, respectively, suggests that analysts using this data element would need to exercise caution in drawing conclusions about federal employees' education levels.

Most CPDF Users Said CPDF Products Met Their Needs, but Some Said Further Awareness of Cautions on CPDF Data Could Affect Use of Data

The questionnaire that GAO sent to 247 users of CPDF data showed that the majority of respondents believed that CPDF products met their needs. GAO's questionnaire asked individuals who requested CPDF products from OPM in fiscal year 1996 if they believed that (1) the CPDF products met their needs, including the data being current, accurate, and complete; and (2) the cautions OPM provided to them on the limitations associated with using CPDF data or products were sufficient for them to present the CPDF data correctly.

The majority of the 140 users who responded to our questionnaire said that to a great or very great extent, CPDF products met their needs (67 to 81 percent, depending on the product); and CPDF data were current (70 to 73 percent), accurate (65 to 87 percent), and complete (71 to 89 percent). The majority of respondents also reported that they received sufficient cautions about the limitations of the CPDF data or products they used. However, OPM officials said, and respondents' answers to GAO's questionnaire indicated, that the extent to which OPM provided users with all of the 28 known cautions on limitations associated with CPDF data varied. In this regard, 29 of the 71 CPDF users said knowing about cautions they were not made aware of would have affected the way they used or presented CPDF data. In discussions with GAO, OPM officials reported they were considering creating a CPDF web site on the Internet that would allow OPM to make CPDF data more widely available and allow OPM to "bundle" or link specific cautions about the limitations associated with specific data sets.

System Software Development Not Documented According to Applicable Federal Guidance, but Software Appears to Implement Edits as Intended

From 1976 to 1995, federal guidance recommended that agencies use a structured approach for operating and maintaining automated information systems, such as the Central Personnel Data System, which includes the computer applications that OPM uses to process data for the CPDF. According to the guidelines, as part of a structured approach, agencies were to document the life cycle of an automated information system from its initiation through its installation and operation. Among other things, such documentation helps agencies efficiently correct system problems even after the system designers have left an agency.

Although applicable guidelines recommended such documentation, OPM did not document changes that were made to the System in 1986 when it did a major redesign of the System's software or document that the testing of these changes was independently done to verify that they worked as intended. OPM officials said verification of the 1986 changes was done by

staff in the unit responsible for designing those changes. As described in GAO's guidance on Year 2000 computer conversions, such testing should be done by an independent reviewer. OPM officials said that although the 1986 changes were not documented, to their knowledge the System has not had problems processing data reliably. GAO's review of 718 of the 763 computer instructions used by the CPDF showed that those instructions should implement CPDF edits as intended.

OPM officials said that for OPM to accomplish its information technology goals it will have to follow a structured approach for future computer application development. The software development goal stated in OPM's Information Technology Architecture Vision would require that by 2002 newly developed, or newly modified, computer systems and programs would be developed under a systematic, well-documented approach. However, it is not clear how soon this requirement is to be implemented.

Recommendations

GAO recommends that the Director of OPM

- make the results of its historical measurements of the CPDF's accuracy available to all users,
- make the 28 caution statements associated with CPDF data available to all users, and
- document future CPDF computer system and software changes and independently verify that those changes are working as intended.

Agency Comments

OPM provided written comments on a draft of this report (see app. VII) that are discussed at the end of chapters 2, 3, and 4.

The OPM Director did not specifically refer to GAO's first two recommendations—that she make the results of OPM's historical measurements of the CPDF's accuracy available to all users and that she make the 28 caution statements associated with CPDF data available to all users. However, she said that OPM will make available appropriate explanatory material to all CPDF users. She said that she agreed with GAO's third recommendation—that OPM document future CPDF computer system and software changes and independently verify that those changes are working as intended. She also said that OPM will fully document all future computer system and software changes and perform independent verification that the changes function as intended.

Contents

Executive Summary		2
Chapter 1		12
Introduction	Background	12
	Objectives, Scope, and Methodology	16
Chapter 2		26
CPDF Data Reviewed	OPM Measures Historical Accuracy of CPDF, but Does Not Report Results of Its Accuracy Measurements to Non-OPM Users	26
Appear to Be Mostly	Most CPDF Data Tested Were Accurate and Agreed With Agencies' Personnel Records	27
Accurate in the	Conclusion	37
Aggregate	Recommendation to the Director of OPM	37
	Agency Comments and Our Evaluation	37
Chapter 3		39
USERS Generally	USERS Generally Reported That CPDF Data Met Their Needs Including Being Current, Accurate, and Complete	39
Reported CPDF	Most CPDF USERS Said Cautions OPM Provides on Data Limitations Were Sufficient, but Some Said Further Awareness of Cautions Could Affect Use of Data	41
Products Met Their	Most CPDF USERS Surveyed Rated the Overall Quality of CPDF Products as Excellent or Very Good	44
Needs, but Further	Conclusions	44
Awareness of	Recommendation to the Director of OPM	44
Cautions on CPDF	Agency Comments and Our Evaluation	44
Data Could Affect Use		
of the Data		

<p>Chapter 4 System Software Development Not Documented According to Applicable Federal Guidance, but Software Appears to Implement Edits as Intended</p>	<p>OPM Did Not Document an Upgrade of the System’s Software as Recommended in Federal Guidance 46</p> <p>The System Appears to Implement CPDF Data Edits Reliably 49</p> <p>OPM Has Implicitly Committed to Adopt an SDLC Approach 50</p> <p>Conclusions 52</p> <p>Recommendation to the Director of OPM 52</p> <p>Agency Comments and Our Evaluation 53</p>	<p>46</p> <p>46</p> <p>49</p> <p>50</p> <p>52</p> <p>52</p> <p>53</p>
<p>Appendixes</p>	<p>Appendix I: To Ensure CPDF Data Quality, OWI Provides Guidance to Agencies and Checks Data Before Entering Them in CPDF 54</p> <p>Appendix II: Results of GAO and OPM Measurements of CPDF Accuracy 63</p> <p>Appendix III: CPDF Data Elements by File Location, as of February 1998 64</p> <p>Appendix IV: Data Elements With the Highest Rate of Inaccuracy 68</p> <p>Appendix V: GAO’s CPDF Employee Verification of Personnel Information Questionnaire 70</p> <p>Appendix VI: GAO’s CPDF Customer Survey 76</p> <p>Appendix VII: Comments From the Office of Personnel Management 88</p> <p>Appendix VIII: Major Contributors 91</p>	<p>54</p> <p>63</p> <p>64</p> <p>68</p> <p>70</p> <p>76</p> <p>88</p> <p>91</p>
<p>Tables</p>	<p>Table 1.1: List of Data Elements Used in Employee Questionnaire and Comparison of CPDF Data With Official Personnel Folders 17</p> <p>Table 1.2: Breakdown of Sampled Employees Responding and Not Responding to the Questionnaire 20</p> <p>Table 2.1: Questionnaire Respondents Reported Most Data Elements Were Generally Accurate 29</p> <p>Table 2.2: Number of Differences Between 113 Employees’ Official Personnel Folders and Agency Records and the CPDF 30</p> <p>Table 2.3: Errors Identified by GAO’s and OPM’s Measurements of CPDF Data Accuracy Grouped by Percent of Errors 33</p>	<p>17</p> <p>20</p> <p>29</p> <p>30</p> <p>33</p>

Figure

Figure I.1: OWI's Acceptance Process

56

Abbreviations

AID	Agency for International Development
CMM®	Capability Maturity Model SM
CPDF	Central Personnel Data File
DOD	Department of Defense
FBI	Federal Bureau of Investigation
FIPS	Federal Information Processing Standards
HHS	Health and Human Services
ILDRS	Installation Level Data Retrieval System
IT	Information Technology
NBS	National Bureau of Standards
OIT	Office of Information Technology
OMB	Office of Management and Budget
OPF	Official Personnel Folder
OPM	Office of Personnel Management
OWI	Office of Workforce Information
SDLC	System Development Life Cycle
SSA	Social Security Administration
USER	User Simple and Efficient Retrieval

Introduction

According to the Office of Personnel Management's (OPM) Guide to the Central Personnel Data File (CPDF), the CPDF is the federal government's central personnel automated database that contains statistically accurate demographic information on about 1.9 million federal civilian employees. The CPDF's primary objective is to provide a readily accessible database for meeting the workforce information needs of the White House, Congress, OPM, other federal agencies, researchers, and the public. A second objective is to relieve agencies that submit personnel data to the CPDF of the need to provide separate data or reports to meet a variety of reporting requirements. Data that agencies submit to the CPDF represent their official workforce statistics.

OPM's Office of Workforce Information (OWI) is responsible for accepting and entering data into the CPDF and processes the data using the Central Personnel Data System. OWI also prepares reports using CPDF data and distributes CPDF data to both OPM and non-OPM users.

In order to safeguard the privacy of federal civilian employees as required under the Privacy Act of 1974, OPM must protect CPDF data from unauthorized disclosure. For example, at OPM access to agencies' CPDF submissions is limited to OPM staff responsible for determining if the data meet OPM's guidelines for acceptance into the CPDF. When disseminating CPDF data, OPM is to protect the privacy of individuals. For example, OPM is not to provide employees' names, Social Security numbers, or birth dates to requesters or to make this information available to the federal agencies³ that are allowed to access the CPDF via OPM's electronic User Simple and Efficient Retrieval (USER) system to retrieve personnel data to do their work.⁴

Background

The CPDF contains personnel data for most of the executive branch departments and agencies as well as a few agencies in the legislative branch. Included are all of the cabinet departments (e.g., State, Treasury, Justice); the independent agencies (e.g., Environmental Protection Agency, Small Business Administration, National Aeronautics and Space Administration); commissions, councils, and boards (e.g., National Council

³According to OPM, GAO, the Equal Employment Opportunity Commission, the Merit Systems Protection Board, the Department of Agriculture, Department of Labor, Environmental Protection Agency, National Guard, National Security Agency, Congressional Budget Office, and the Office of Management and Budget were trained and given access to this system by OPM.

⁴OPM will provide such data to GAO, the Merit Systems Protection Board, and the Equal Employment Opportunity Commission to survey federal employees about their opinions and experiences, but those agencies are to have their own procedures for protecting the privacy of survey respondents.

on the Handicapped); and selected legislative branch agencies, such as the Government Printing Office.

The CPDF does not contain employee data for the Central Intelligence Agency, Defense Intelligence Agency, the Board of Governors of the Federal Reserve System, National Security Agency, Office of the Vice President, Postal Rate Commission, Tennessee Valley Authority, U.S. Postal Service, or the White House Office. The CPDF also excludes from coverage non-U.S. citizens working for federal agencies in foreign countries; most nonappropriated fund personnel;⁵ commissioned officers in the Department of Commerce, Department of Health and Human Services (HHS), and the Environmental Protection Agency; and all employees of the judicial branch.⁶

The History of the CPDF

The Civil Service Commission, OPM's predecessor, decided it would install a type of central personnel database—the CPDF—in 1972 to provide a source that was capable of (1) satisfying minimum essential statistical data needs for central management agencies and the public; (2) meeting reporting requirements, such as periodic surveys of affirmative employment programs and semiannual turnover reports; and (3) alleviating the need for agencies to individually report similar information separately to requesters. The CPDF also expanded and replaced the Federal Personnel Statistics Program Sample File, which was established in 1962. The File contained a continuous work history on each federal employee whose Social Security account number ended in the digit “5,” a population that constituted a 10-percent sample of the federal workforce.

How the CPDF Operates

OPM builds six files from agency-submitted data. These are the longitudinal history (a record of personnel actions arranged by date within Social Security number), organizational component (a listing of the codes used by each agency to identify its various work units, e.g., regions, divisions, branches); personnel office identifier (contains the mailing address and telephone number for personnel offices that report to the CPDF); name (a cross-reference listing of names, Social Security numbers, accession dates, and applicable separation dates of employees reported to the CPDF); status;

⁵Nonappropriated fund personnel are employees of activities that do not receive congressional appropriations, e.g., the Department of Defense's (DOD) Commissary Service.

⁶OPM has proposed that Congress grant it authority over any executive agency subject to the merit system principles set forth in 5 U.S.C. 2301, or their equivalent.

and dynamics files. Of the six, this report focuses on the status and dynamics files. They are the source of the demographic information used by OWI to write reports and to respond to data requests by users of CPDF data.

The status file consists of data elements describing each employee as of the date of the file. Agencies are required to submit these files on a quarterly basis, with the submissions due at OPM no later than the 22nd of the month following the end of the quarter (e.g., input for the quarter ending December 31 must be submitted by January 22). All of the employees covered by the CPDF are to be included in each file. The data elements include information on the type of work; the employee's pay; and personal information, such as gender and birth date.

The dynamics file consists of data elements describing each personnel action taken by an agency during the period covered by the file. Personnel actions are the official records of employees' careers, such as hires, promotions, reassignments, pay changes, resignations, and retirements. The file includes information about the action taken, the agency/subelement, the position, pay, and the individual employee. The normal reporting period is a calendar month but may end as of the last full biweekly pay period of the month. Submissions are due at OPM as soon as possible following completion of agency processing but no later than 22 days following the end of a monthly reporting period. As of February 1998, the CPDF consisted of 95 separate data elements. Of this number, 68 are to be reported by agencies in their monthly and quarterly dynamics and status file submissions.

OPM relies on agencies to ensure that the data they submit are timely, accurate, complete, and edited in accordance with OPM standards. OPM provides agencies with guidance, the Guide to the CPDF, which says agencies are to test the data they provide to the CPDF to ensure that the data are accurate and complete. To help agencies ensure the quality of their data, OPM provides them with the CPDF Edit Manual, which prescribes the data values to which agencies' data are to conform before they are submitted. To test the values of their data, agencies are to use OPM's CPDF edits. These edits are computer instructions that are to check the validity of individual data elements as well as the proper relationship of values among associated data elements. For example, the edit for the sex data element checks that the character used to define the data element is either "M" for male or "F" for female; the edit identifies other characters as errors. OPM expects agencies to incorporate these CPDF edits into their

internal personnel data systems. These edits constitute the minimum level of quality control OPM expects the agencies to employ. Agencies have the option of incorporating additional quality controls, such as testing a sample of the data for accuracy before submitting it, in addition to applying the CPDF edits.

The CPDF edits cannot detect all types of errors. For example, an edit for the sex data element would not be able to detect if the character “M” was incorrectly used to identify a female employee. According to OWI officials, although they provide agencies with the edits, errors still occur in submissions, which OPM strives to identify through OWI’s quality review process. The officials also said that errors in pay-related data elements often occur at the beginning of the year because agencies make their beginning-of-the-year submissions before they install edits that reflect annual cost-of-living pay increases. The Guide to the CPDF also informs agencies about what data elements should be included in their CPDF data submissions and the frequency and timing of the submissions. As mentioned earlier, frequency and timing requirements differ for the status and dynamics data files.

After agencies submit the personnel data, OPM puts the submissions through an acceptance process before the data can be entered into the CPDF. This process includes putting the data through the same CPDF edits the agencies were to use before submitting the data as well as other analyses. OWI manages the process. Its staff are to provide agencies with feedback on their submissions, requesting, as needed, corrections to submissions that fail edit checks or other analyses and preventing data that are not within the acceptable range of data values from being entered into the CPDF. OWI is to make the final decision about what data are entered into the CPDF. At the time of our review, the Central Personnel Data System was operated by OPM’s Office of Information Technology (OIT) for OWI. The CPDF Quality Control team that monitored agencies’ data submissions was part of OIT. However, operation of the System was transferred to OPM’s Retirement and Insurance Service in 1997, and the Quality Control team were reassigned to OWI.

OPM Has Authority to Request Agency Data for the CPDF

OPM may require agencies under 5 C.F.R. section 7.2 to report “in such manner and at such times as OPM may prescribe, such personnel information as it may request.” On the basis of this authority, OPM is able to direct agencies to submit selected personnel data to the CPDF. However, although the OPM Director can request data, she cannot ensure that

agencies provide accurate information in a timely manner. The responsibility for providing timely, accurate information remains with the head of the agency providing the information. OPM officials rely on federal agencies to voluntarily comply with CPDF guidelines and correct problem submissions.

Objectives, Scope, and Methodology

For this review, we had three objectives:

- (1) determine the extent to which selected CPDF data elements are accurate, including the data elements used by OPM's Office of the Actuaries for estimating the government's liability for future payments of federal retirement programs;
- (2) determine whether selected users of CPDF data believed CPDF products met their needs, including whether the products were current, accurate, and complete and whether the cautions OPM provided to them on the limitations associated with using the data were sufficient for them to present the CPDF data correctly; and
- (3) determine whether OPM has documented changes to the System and verified the System's acceptance of those changes, as recommended in applicable federal guidance, and whether the System would implement CPDF edits as intended.

Objective 1

To determine the extent to which selected CPDF data elements are accurate, including the data elements used by OPM's Office of the Actuaries for estimating the government's liability for future payments of federal retirement programs, we (1) designed and sent questionnaires to a random sample of federal employees to have them verify some of their CPDF data and (2) compared CPDF data with information in randomly selected official personnel folders and in other agency records at selected personnel offices. Table 1.1 presents a list of the CPDF data elements we used in our employee questionnaire and comparison of CPDF data with information in official personnel folders.

**Chapter 1
Introduction**

Table 1.1: List of Data Elements Used in Employee Questionnaire and Comparison of CPDF Data With Official Personnel Folders

Data element	Location in CPDF		GAO's approaches	
	Status file	Dynamics file	Questionnaire	Comparison
Adjusted basic pay	•	•	•	•
Agency/Subelement	•	•	•	•
Annuitant indicator	•		•	•
Birth date	•	•	•	•
Current appointment authority	•	•		•
Duty station	•	•	•	•
Education level	•	•	•	•
Effective date of action		•		•
Handicap	•	•	•	•
Legal authority code		•		•
Employee name ^a			•	•
Nature of action code		•		•
Occupation	•	•	•	•
Pay plan/grade ^b	•	•	•	•
Pay rate determinant	•	•		•
Personnel office identifier	•	•		•
Position occupied	•	•		•
Race or national origin	•	•	•	•
Rating of record	•	•	•	•
Retirement plan	•	•	•	•
Service computation date	•	•	•	•
Sex	•	•	•	•
Social Security number	•	•	•	•
Tenure	•	•		•
Veterans preference	•	•	•	•
Veterans status	•	•	•	•
Work schedule	•	•	•	•

^aTo protect the confidentiality of employee records, OPM stores employee names separately from the major CPDF databases in the CPDF name file.

^bPay plan and grade are separate data elements. We combined these two data elements into one on the questionnaire to make it easier for employees to respond.

Source: OPM's Office of Workforce Information and GAO.

We used two approaches, i.e., a questionnaire (see app. V) and a comparison of data in official personnel folders and agency records to CPDF data, to measure the accuracy of CPDF data. We sent the questionnaire to a sample of employees, because OPM studies show that official personnel files or agency records may be in error. We compared the results of both approaches to develop our findings.

We also reviewed past OPM accuracy measurements, examined CPDF data for missing and unusable information, and interviewed an official of OPM's Office of the Actuaries to discuss the accuracy of the data elements the Office uses for estimating the government's liability for future retirement payments. These steps are more fully described in the following sections.

Questionnaire

As part of our evaluation of the accuracy of CPDF data, we selected a stratified random sample of 565 federal employees and attempted to send each a questionnaire containing 20 data elements about themselves obtained from the CPDF (see app. V for a copy of our questionnaire). The data elements that we included in the questionnaire were among those we most frequently use to do our work, those OPM analysts use most frequently in preparing CPDF reports, and those used by OPM's Office of the Actuaries to estimate the government's liability for future payments of federal retirement programs. We selected those data elements that we believed employees would be able to verify. We included in each individual's questionnaire data elements from the September 1996 CPDF status file about that individual. The elements consisted of (1) Social Security number, (2) employing agency/subelement, (3) adjusted basic pay (including locality pay), (4) month and year of birth, (5) duty station, (6) pay plan, (7) grade, (8) handicap, (9) occupation, (10) race or national origin, (11) service computation date, (12) sex, (13) veterans preference, (14) veterans status, (15) work schedule, (16) education level, (17) rating of record, (18) retirement plan, (19) annuitant indicator, and (20) employee name. We asked the respondents to verify the accuracy of each data element, indicating whether it was correct or incorrect as of September 30, 1996. When a respondent indicated that a data element was incorrect, we asked the respondent to enter the correct information.

We pretested the questionnaire to assure ourselves that respondents could interpret the questions correctly and could provide the information requested. We modified question wording and questionnaire format on the basis of what we learned from five pretests.

The random sample of 565 was drawn from 7 strata to represent a study population of 1,905,787 non-Federal Bureau of Investigation (FBI) federal employees whose names were contained in the CPDF database as of September 30, 1996.⁷ Random samples of 30 selections each were drawn from 6 smaller strata, each of which comprised a single personnel office. These six personnel offices were among the eight largest personnel offices in the federal government.⁸ These offices were the Social Security Administration (SSA), Baltimore, MD; Department of the Army, Fort Benning, GA; U.S. Customs Service, Washington, D.C.; National Institutes of Health, Bethesda, MD; Department of State, Washington, D.C.; and Department of the Navy, Pensacola, FL. We selected these personnel offices because of their size and because our work at the offices could then be representative of a relatively large portion of records contained in the CPDF. The 6 personnel offices were among 1,425 in the government and served over 8 percent of the employees whose data were contained in the CPDF as of September 30, 1996. We used the selections from these six personnel offices for both the employee questionnaire and a review of official personnel folders. The remainder of the sample—385 selections—was randomly drawn to represent the remaining stratum of 1,746,592 employees from all other personnel offices. The total sample of 565 was designed to ensure that it approximately mirrored the population distribution with respect to type of appointment (career or noncareer), work schedule (full-time or non-full-time), type of service (competitive or excepted), and location (stationed in or outside the United States).

Because the CPDF does not contain mailing addresses for employees, we mailed most of our questionnaires to personnel officers whom we identified in the CPDF as serving the employees in our sample. In all, 562 of the 565 sampled employees were covered by 280 personnel officers. We were not able to identify the personnel officers for three of the sampled employees. We asked the personnel officers to whom we sent questionnaires to forward them to the sampled employees. In addition, we asked them to provide us with the direct mailing address of each sampled employee so that we would be able to mail follow-up questionnaires directly to sampled employees who did not return a questionnaire to us within 45 days. We also asked the personnel officers to furnish us with reasons why any of the questionnaires could not be forwarded to the

⁷Although the FBI submits data to the CPDF, it does not provide certain information, such as duty station, that we wanted to review.

⁸Although the Department of Veterans Affairs is the executive branch agency with the most employees after DOD, its personnel offices are not among the largest. We also did not include the Postal Service in our study because it does not submit data to the CPDF.

sampled employees. After an initial and a follow-up mailing, we received 407 usable questionnaires out of 565, for a 72 percent response rate. Table 1.2 presents a breakdown of the number of sampled federal employees responding to our questionnaire as well as the various reasons why some sampled employees did not respond.

Table 1.2: Breakdown of Sampled Employees Responding and Not Responding to the Questionnaire

Employees in initial sample		565
Respondents		407 ^a
Nonrespondents		
Refusal (questionnaire not returned or returned blank)		81
Employee resigned		16
Employee retired		15
Employee deceased		3
Employee transferred from agency		9
Employee on extended leave		4
Personnel officer could not locate employee		13
GAO could not locate personnel officer		3
POI ^b address unknown-returned by Postal Service		3
Questionnaires returned after the closeout date		3
Other miscellaneous reasons		8
Total nonrespondents		158

^aThe response rate was 72 percent.

^bPersonnel Office Identifier.

Source: GAO questionnaire.

We edited the questionnaires received from respondents to identify data elements marked as incorrect. In cases where a respondent indicated that a data element from the CPDF was incorrect, the editor then made an effort to determine if the correction entered onto the questionnaire by the respondent was logical. For example, a number of respondents indicated that the annual pay amount shown on the questionnaire was incorrect. However, in researching the “correct” amount entered by the respondent, it was determined that the amount entered was his or her current annual pay, not the annual pay as of September 30, 1996, as indicated in the question. In these cases, the response was changed from incorrect to correct by the editor.

The 407 returned questionnaires from the 7 strata were weighted to represent the population of 1,905,787 federal employees for all results presented in this report. Sampling errors have been calculated to take into account the different weights assigned to each stratum. Unless otherwise noted, the 95 percent confidence intervals around all reported results are plus or minus 5 percentage points or less.

In addition to sampling errors, the practical difficulties of administering any questionnaire may introduce other types of errors, commonly referred to as nonsampling errors. For example, differences in how a particular question is interpreted by the questionnaire respondents could introduce unwanted variability in the questionnaire's results. We took steps in the development of the questionnaire, the data collection, and the data editing and analysis to minimize nonsampling errors.

Comparison of Official Personnel Folders and Agency Records With CPDF Data

We also compared data contained in official personnel folders and other agency records with data in the CPDF for the same period at the six selected personnel offices. For each of the 6 personnel offices we selected, we chose 30 employees at random from the September 1996 CPDF status file. The employees were those who were reported by the CPDF as being served by the six respective personnel offices. At each of the 6 personnel offices, we asked for official personnel folders for the 30 employees. We also asked for information from the personnel offices' automated files on ratings, handicap, and race or national origin because such information is not necessarily contained in personnel folders. We then selected 20 employees at random from those whose official personnel folders were available. We over-sampled by 10 employees in our initial sample for each personnel office because we anticipated that some folders would be unavailable because of employee departures or other reasons. At SSA, we reviewed official personnel folders and other agency records for only 13 employees because the official personnel folders for 17 of the 30 employees we chose at random were located in offices throughout the country and not in a central location as we initially expected. In total, we reviewed folders and other agency records for 113 employees for the 6 personnel offices.

For each of the 113 employees in our sample, we obtained information from the September 1996 CPDF status and dynamics files. The information we obtained consisted of the 20 data elements we used for our questionnaire and the data elements that we most frequently use to do our work, including key status and dynamics data elements. The eight data elements that were in addition to the data elements used for the

questionnaire were current appointment authority, effective date of action, legal authority code, nature of action code, pay rate determinant, personnel office identifier, position occupied, and tenure. We reviewed a total of 28 data elements: 23 data elements common to both the status and dynamics files, 1 element found only in the status file, 3 elements found only in the dynamics file, and employee name (see table 1.1 for the CPDF data elements we reviewed and their file locations).

For each employee, we compared the CPDF data with relevant documents, such as Standard Forms 50 (notification of personnel action) and employment applications, in official personnel folders. We also compared the CPDF data with automated files on those employees' ratings, handicap, and race or national origin. We discussed any mismatches we found with personnel officials in an attempt to determine how differences can occur between the CPDF and agency documentation.

Past OPM Accuracy Measurements

OPM conducts periodic measurements of CPDF accuracy by comparing data in the official personnel folders of separated employees with data in the CPDF. We reviewed the six measurements of CPDF accuracy OPM did from April 1984 to July 1996 and compared the results of our evaluation of CPDF accuracy with the results of OPM's last two measurements, which were issued in January 1992 and July 1996.

Data Used by OPM's Office of the Actuaries

To determine if the CPDF data used by OPM's Office of the Actuaries to estimate the government's liability for future retirement payments are sufficiently accurate for use by the Office, we first met with the actuary responsible for calculating this liability to determine the CPDF data elements used in the estimate. After our analysis of the employee questionnaire and our comparison of personnel folders and other agency records to CPDF data, we again interviewed the actuary to discuss the results of our two approaches and the impact of errors on the estimate.

Additional Methodological Characteristics

The results of our employee questionnaire are generalizable to the universe of 1,905,787 employees included in the CPDF's September 1996 status file. Table 2.1 shows the generalized results as a percentage of records in the September 1996 status file. The results of our comparison of employees' official personnel folders and other agency records to CPDF data are not generalizable to the CPDF as a whole, although they may be indicative of the personnel offices at which we performed our work.

The CPDF data elements measured for accuracy generally were among those identified by OPM as key to the accuracy of its recurring reports. We

cannot determine from the work we did the accuracy of data elements we did not review. We did not independently verify educational levels reported by employees or any of the responses of employees.

Our accuracy findings are for CPDF data in the September 30, 1996, status file and the fiscal year 1996 dynamics file. The accuracy might differ for previous and future CPDF files, especially when agency procedures or information processing technology change.

Our accuracy measurement was not designed to evaluate the reliability of CPDF data from individual agencies or specific subsets of employees, such as those on leave without pay. OPM reports on the percentage of data elements in agency submissions that do not pass standard CPDF edits show considerable variation across agencies.

Objective 2

To determine whether selected users of CPDF data believed CPDF products met their needs including whether the products were current, accurate, and complete and whether the cautions OPM provided to them on the limitations associated with using the data were sufficient for them to present the CPDF data correctly, we designed, with advice from OPM, a CPDF customer questionnaire (see app. VI for a copy of our questionnaire). We mailed the questionnaires to 247 individuals identified by OPM's OWI as representing all the requesters of CPDF products in fiscal year 1996 who obtained data directly from OPM.

We mailed the customer questionnaires in May 1997 to the return addresses on letters in OWI's fiscal year 1996 correspondence files that had requested CPDF products and to recipients of recurring CPDF-based reports in 1996. We followed up our initial mailing with a second one in June and a third one in July. We did not include in our analysis any questionnaires received after August 6, 1997.

After August 6, 1997, we made follow-up telephone calls to all nonrespondents and determined that 40 of the original 247 individuals we sent the questionnaire to were either not CPDF users or had left their organizations. Of the remaining 207 individuals who were CPDF users, 140 (or 68 percent) responded to the mail questionnaire, and an additional 21 responded to an abbreviated version of the mail questionnaire we used in follow-up telephone calls to nonrespondents. The combined response rate for the mail-out questionnaire and the telephone follow-up was 78 percent.

After we received the questionnaires from the respondents, we edited them for completeness and consistency. All of the data from the questionnaires were double-keyed and verified during data entry. In addition, a random sample of these data was verified back to the source questionnaires.

**Additional Methodological
Characteristics**

The results of our customer questionnaire are not generalizable to the universe of users of CPDF data and products for 1996 because we could not define the universe of users necessary to draw a representative sample. The distribution of CPDF products, such as recurring reports, is not controlled. These products are available through various outlets, such as libraries, that do not track customers. Therefore, we relied on OWI to identify those customers who corresponded with it in 1996 to request CPDF data and sent our questionnaire to this defined but nonrepresentative subset of the 1996 universe of CPDF users.

Objective 3

To determine whether OPM has documented changes to the Central Personnel Data System and verified the System's acceptance of those changes, as recommended in applicable federal guidance, and whether the System would implement CPDF edits as intended, we first reviewed federal guidance on managing automated information systems. To determine the extent to which OPM's OIT followed the guidance in managing the development of the System, we conducted interviews at OIT, which was responsible for operating the System, and OWI, which is the System's owner, about their basis for determining the System's reliability. From these officials, we requested available documentation relating to modifications and upgrades of software used by the System to process CPDF data and documentation relating to verification that these modifications and upgrades worked as planned. We also reviewed available documentation on OPM's current Information Technology Strategy to determine whether it includes procedures for managing the System in the future. To determine whether the System would implement CPDF edits OPM uses to screen the 68 data elements reported by agencies to OPM as intended, we reviewed 18 of the 63 validity⁹ and all 700 of the call-relational¹⁰ edits the System uses to screen agencies' data submissions.

⁹Validity edits check data against a defined range of acceptable values to identify data that fall outside the range.

¹⁰The call-relational edits are a series of subroutines or programs within the "Dynamics Main Edit Module" and the "Status Main Edit Module" that control the editing of an agency's dynamics and status submission files. These edits do not make corrections to any of the data elements. They produce reports that show which fields or data elements are incorrect or failed validity checks.

**Additional Methodological
Characteristics**

We judgmentally selected only the 18 validity edits OWI uses to screen the data elements it considers critical; therefore, the findings of our review of these 18 edits cannot be generalized to all 63 validity edits. Because we did not actually put test data through the System or otherwise test the reliability of the System's hardware and software under operating conditions, we cannot verify the reliability of the System. We did not assess the likelihood that the CPDF would be Year 2000 compliant by December 31, 1999.

We conducted our work between November 1996 and June 1998, in accordance with generally accepted government auditing standards. The employee CPDF data verification questionnaire and CPDF customer survey were administered between May 1997 and September 1997; thus, the data are as of those dates.

We requested comments on a draft of this report from the Director of OPM. OPM provided written comments on a draft of this report (see app. VII) that are discussed at the end of chapters 2, 3, and 4.

CPDF Data Reviewed Appear to Be Mostly Accurate in the Aggregate

The accuracy of the data the CPDF contains depends on the accuracy of the data that agencies submit. Errors in those data can occur at various stages of the personnel process, such as when agency personnel clerks enter data for newly hired employees or when they code information on personnel actions (e.g., performance appraisals). OPM does not have an official accuracy standard for agencies' submissions. On a periodic basis, however, OPM draws a governmentwide sample of CPDF records and measures CPDF data accuracy by comparing selected data in former federal employees' official personnel folders to data in the CPDF for the same period. OPM generally makes the results of its measurements of CPDF accuracy available to OPM users of CPDF data but not to non-OPM users. In spite of the important uses of CPDF data, no independent evaluation of the accuracy of the data has been done. Our work showed that most of the CPDF data elements we reviewed were 99 percent or more accurate on a governmentwide basis. The rating of record and education level data elements had the highest error rates, at about 5 and 16 percent for rating of record, and 23 and 27 percent for education level based on our questionnaire and comparison, respectively. Our overall findings are broadly similar to what OPM found when it measured historical accuracy¹¹ in 1996 by comparing 1994 data in former employees' official personnel folders with the data in the CPDF. We shared the results of our work with the actuary responsible for calculating the federal government's liability for future retirement payments to retired federal employees and their survivors, and he said that the CPDF data elements were sufficiently accurate for making this estimate.

OPM Measures Historical Accuracy of CPDF, but Does Not Report Results of Its Accuracy Measurements to Non-OPM Users

OPM periodically measures the accuracy of selected data that are in the CPDF. As we said earlier, OPM relies on agency data passing CPDF edits to eliminate errors that would result in inaccurate data being entered in the CPDF. For example, the edits are to identify a salary amount that is too high for a particular pay plan or grade. However, the edits are not able to identify an error in salary that is within the range of that pay plan or grade. Thus, inaccurate data can get into the CPDF.

To measure the historical accuracy of CPDF data, OPM periodically compares certain data found in a sample of former federal employees' official personnel folders to data in the CPDF for the same period. From April 1984 to July 1996, OPM conducted six such measurements. OPM analysts used a sample of former employees and compared certain data

¹¹The CPDF/OPF Accuracy Survey, which has been done every few years (the most recent is for fiscal year 1994 CPDF data and was issued in July 1996), identifies historical error rates.

elements in their official personnel folders to information in the CPDF's status and dynamics files. For example, the latest measurement, which was released in 1996 for fiscal year 1994 data, used a sample of 135 former employees and compared 35 status file and 40 dynamics file data elements to information in the official personnel folders. An error was defined as a value found in the CPDF that was not the same as that found in the employee's official personnel folder.¹² OPM officials told us—and OPM's accuracy surveys state—that the surveys were designed to measure the accuracy of governmentwide data only and not the accuracy of data from individual agencies. OPM generally makes the results of its measurements of CPDF data accuracy available to CPDF data users within OPM but not to non-OPM users.

In five of the historical accuracy measurements, OPM found that most CPDF data were generally accurate, and in most cases the selected data elements matched the corresponding official personnel folder entries 99 percent or more of the time. However, OPM did not make that statement for its December 1990 measurement of 1988 CPDF data. Instead, it advised OPM users of CPDF data to review the results of the accuracy measurement and determine for themselves whether the data were sufficiently accurate for their use. OPM officials said that OPM does not routinely inform non-OPM users of the results of its measurements of historical accuracy.

OPM has not promulgated a standard for the accuracy of CPDF data. To our knowledge, no federal agency has promulgated accuracy standards that are generally applicable to federal databases. In general, the level of accuracy for data must be balanced against what the data are to be used for and the cost of obtaining a greater level of accuracy.

Most CPDF Data Tested Were Accurate and Agreed With Agencies' Personnel Records

To measure the accuracy of the CPDF, we (1) sent a questionnaire to a random sample of federal employees to gather information about the accuracy of 20 of the 68 CPDF data elements reported by agencies and (2) compared data for 28 data elements in the CPDF with the data contained in the official personnel folders and other agency records for 113 randomly selected employees at 6 of the largest federal personnel offices.¹³ We found that most CPDF data elements we tested were accurate and agreed

¹²OPM does not count asterisks or missing data as errors. OPM analysts insert asterisks into the CPDF when submitted data elements do not pass edit checks and agencies do not provide corrected data.

¹³These offices were SSA, Baltimore, MD; Department of the Army, Fort Benning, GA; U.S. Customs Service, Washington, D.C.; National Institutes of Health, Bethesda, MD; Department of State, Washington, D.C.; and Department of the Navy, Pensacola, FL. Although the Department of Veterans' Affairs is the executive branch agency with the most employees after DOD, its personnel offices are not among the largest. And, we did not include the Post Office in our study because it does not submit data to the CPDF.

with information in employees' official personnel folders and other agency personnel records. Although our methodology differed from the one OPM uses in its measurements of historical accuracy, the results of our review were broadly similar to OPM's results.

Questionnaire Results and Comparison of Selected CPDF Data to Employee Records Showed Most Data Were Accurate and Agreed With Agencies' Personnel Records

To determine the accuracy of 20 selected CPDF data elements, we sent a questionnaire to a random sample of federal employees that was representative of federal employees governmentwide (see ch. 1 for a description of our sampling methodology). We asked them to review information about themselves that we obtained from the September 1996 CPDF. The data elements we asked about were those about which we believed employees would be most familiar, including employee name,¹⁴ birth date, and Social Security number. The results of our questionnaire showed that 14 of the 20 data elements, or 70 percent, matched data in the CPDF in 99 percent or more of the cases (see table 2.1). There were no inaccuracies for seven of these data elements and the other seven data elements had error rates of less than 1 percent. The remaining six data elements had error rates greater than 1 percent (see table 2.1).

The two most error-prone data elements were education level and rating of record. Education level had a 26.7 percent error rate and rating of record had a 4.7 percent error rate. The education level data element is intended to reflect the highest education level that a federal employee achieved. The rating of record data element indicates an employee's most recent rating or performance appraisal. The results of our employee questionnaire are generalizable to the universe of 1,905,787 employees included in the CPDF's September 1996 status file. Table 2.1 shows the generalized results as a percentage of records in the September 1996 status file.

¹⁴OPM provided us with employee names for the purposes of this review.

Chapter 2
CPDF Data Reviewed Appear to Be Mostly
Accurate in the Aggregate

Table 2.1: Questionnaire Respondents Reported Most Data Elements Were Generally Accurate

Data element	Percentage of errors in our sample	95% Confidence interval for federal civilian workforce	
		Lower boundary	Upper boundary
Annuitant indicator	0.0%	0.00%	0.9%
Birth date (month and year)	0.0	0.00	0.9
Agency/Subelement	0.0	0.00	0.9
Occupation	0.0	0.00	0.9
Retirement plan	0.0	0.00	0.9
Social Security number	0.0	0.00	0.9
Work schedule	0.0	0.00	0.9
Sex	0.1	0.00	1.34
Duty station	0.3	0.01	1.95
Service computation date (month and year)	0.4	0.01	1.95
Pay plan/grade	0.7	0.09	2.48
Veterans preference	0.7	0.09	2.48
Employee name ^a	0.9	0.09	2.48
Adjusted basic pay	1.2	0.24	2.98
Race or national origin	2.0	0.64	3.91
Veterans status	2.2	0.88	4.36
Handicap	2.7	2.16	6.51
Rating of record	4.7	2.37	6.77
Education level	26.7	21.50	31.14

Note 1: The returned questionnaires were weighted to represent the population of 1,905,787 federal employees for the results presented in this table. The percentages are generalizable to the universe of federal employees in the CPDF as of September 1996, excluding employees of the FBI.

Note 2: Because the questionnaire results come from a sample of employees, all questionnaire results are subject to sampling error. We are 95-percent confident that the percentage of error for the federal civilian workforce as a whole falls between the lower and upper boundaries listed for each data element. The percentages of errors for the questionnaire results in our sample are reported in the table.

Note 3: Pay plan and grade data elements are combined; therefore, although we checked the accuracy of 20 data elements, the table shows 19 data elements.

^aTo protect the confidentiality of employee records, OPM stores employees' names separately from the major CPDF databases.

Source: GAO questionnaire.

Chapter 2
CPDF Data Reviewed Appear to Be Mostly
Accurate in the Aggregate

We also compared data in employees' personnel folders or other agency records with data in the CPDF for 113 randomly selected employees at 6 of the largest federal personnel offices (see the Objectives, Scope, and Methodology section in ch. 1 for a discussion of our selection process). For this comparison, we reviewed a total of 28 data elements: 23 data elements common to both the status and dynamics files, 1 element found only in the status file, 3 elements found only in the dynamics file, and the employee name data element found in the CPDF name file. (See table 1.1 in the Objectives, Scope, and Methodology section in ch. 1 for the CPDF data elements we reviewed and their file locations.)

In our review of official personnel folders and agency records, we found no inconsistencies among the 23 data elements we included in our comparison that were common to both the status and dynamics files. For example, if the status file data element showed an erroneous education level for a given employee, the dynamics file element showed the same erroneous code. Our review of official personnel folders showed that personnel actions reflected in the CPDF dynamics file appeared to be generally complete.¹⁵ There were no inaccuracies for 12 of the data elements. For another five data elements, our comparison showed error rates of less than 1 percent. The remaining nine data elements had error rates greater than 1 percent. For the legal authority code data element, we could not determine the error rate because some employees had no transactions for fiscal year 1996. Table 2.2 shows the results of our comparison.

Table 2.2: Number of Differences Between 113 Employees' Official Personnel Folders and Agency Records and the CPDF

Data element	Number of differences	Percentage of errors^a
Annuitant indicator	0	0.0%
Effective date of action	0	0.0
Duty station	0	0.0
Agency/Subelement	0	0.0
Handicap	0	0.0
Nature of action code	0	0.0
Position occupied	0	0.0
Race or national origin	0	0.0
Retirement plan	0	0.0
Social Security number	0	0.0
Tenure	0	0.0

(continued)

¹⁵We found two instances of personnel actions in official personnel folders not being recorded in the CPDF dynamics file. These two missing personnel actions are not reflected in the table.

Chapter 2
CPDF Data Reviewed Appear to Be Mostly
Accurate in the Aggregate

Data element	Number of differences	Percentage of errors^a
Work schedule	0	0.0
Birth date (month and year)	1	0.9
Occupation	1	0.9
Personnel office identifier	1	0.9
Service computation date (month and year)	1	0.9
Sex	1	0.9
Employee name ^b	2	1.8
Pay rate determinant	2	1.8
Legal authority code	3	^c
Pay plan/ grade	3	2.7
Veterans preference	4	3.5
Adjusted basic pay	7	6.2
Veterans status	8	7.1
Current appointment authority	11	9.7
Rating of record	18	15.9
Education	26	23.0

Note 1: The percentages reported in this table are based on a random sample of official personnel folders at six of the largest personnel offices in the federal government and cannot be generalized governmentwide.

Note 2: Pay plan and grade data elements are combined; therefore, the table shows 27 data elements rather than 28.

^aThe percentage of errors is based on the number of folders reviewed rather than on the number of personnel transactions documented in the folders.

^bTo protect the confidentiality of employee records, OPM stores employee name separately from the major CPDF databases. In the two cases where the name was incorrect, the employees' names had changed due to a change in their marital status.

^cFor this data element, we could not determine the percentage of errors using the universe of 113 employees because some employees had no transactions for fiscal year 1996.

Source: GAO analysis of agency records and the CPDF.

Concerning the most error-prone data elements, our review of employees' official personnel folders and agency records showed results similar to those of our questionnaire—education level and rating of record were the most error-prone data elements. (See app. III for a more detailed discussion of the data elements that contained the highest rates of error.) However, the results of our comparison between the data in the official personnel folders and the CPDF differ somewhat from those of our questionnaire. For example, the results of the questionnaire showed

education level to have a 26.7 percent error rate and rating of record to have a 4.7 percent error rate. The results of the comparison showed education level to have a 23.0 percent error rate and rating of record to have a 15.9 percent error rate. Although we did not try to determine the reason for these differences, two reasons appear most likely. First, the results of the questionnaire are generalizable governmentwide, although the results of the comparison are not because the sample of the comparison is not generalizable. Second, the information in the employees' official personnel folders might not be current. In particular, employees may not have informed their personnel offices of additional education completed, so this information may not be in the official personnel folder. Thus, the information in the official personnel folder might match the CPDF, but neither would be current.

The Results of Our Review
Were Broadly Consistent
With Those of OPM's
Historical Accuracy
Measurements

In its measurements of historical accuracy of CPDF data, OPM has reported results broadly consistent with ours. That is, OPM has found that most data elements it reviewed were 99 percent or more accurate but has found high error rates for rating of record and education level. Table 2.3 groups by percent of errors the error rates identified by our two methods for measuring CPDF status and dynamics file data accuracy and OPM's measurement of the historical accuracy of fiscal year 1994 CPDF status file data. The table shows that between the two methods we used to measure CPDF data accuracy (although variation existed in the accuracy of some data elements) at least 63 percent of CPDF data elements were 99 percent or more accurate.

Chapter 2
CPDF Data Reviewed Appear to Be Mostly
Accurate in the Aggregate

Table 2.3: Errors Identified by GAO’s and OPM’s Measurements of CPDF Data Accuracy Grouped by Percent of Errors

Percentage of errors	GAO approaches ^a				OPM’s accuracy measurement for fiscal year 1994	
	Questionnaire		Comparison		Percent	Number of data elements
	Percent	Number of data elements	Percent	Number of data elements ^b		
0	35%	7	44%	12	44%	8
<1	35	7	19	5	28	5
1 and above	30	6	37	10	28	5

Note: GAO’s questionnaire and comparison approaches included status and dynamics file data. The OPM accuracy measurement for fiscal year 1994 results used in the table are for status file data only.

^aAlthough we combined pay plan and grade when we surveyed employees, we counted these two data elements separately when determining how many data elements were 99 percent or more accurate.

^bAlthough we compared 28 data elements, this table includes 27. These numbers do not include the legal authority code data element because we were not able to calculate an error rate for it.

Source: GAO analysis of CPDF data.

Although OPM’s and our results were broadly consistent, there are important differences between OPM’s methodology and ours. First, we sent our questionnaire to a generalizable sample of current federal employees and reviewed a random sample of official personnel files of current federal employees. In contrast, OPM reviewed centrally located records of former employees. Second, OPM’s methodology in comparing CPDF data with those in employees’ official personnel folders differed from ours. We often relied on agency records (e.g., records maintained separately from official personnel folders for race, national origin, and handicap) in cases where data were not in official personnel folders, but OPM generally limited its review to documents that were in personnel folders. Third, the way we determined errors differed in part from OPM’s. OPM did not determine if the official personnel folder data element itself was correct, but we did so by researching available agency personnel records.

The Accuracy of CPDF Data Varied by Agency

Our review of employees’ official personnel folders and other agency records was intended to evaluate CPDF accuracy in general, not to compare CPDF data accuracy among individual agencies. Such a review would have required a much larger sample to represent each agency. But during our review, we did find circumstances that demonstrated how accuracy varied by agency and why. For example, although the five other agencies we

reviewed were routinely providing information on employee performance ratings to the CPDF, SSA had not updated rating information in the CPDF for over 2 years at the time of our review. SSA officials told us this lapse occurred because temporary procedures that had been established to correct SSA's difficulty in providing appraisal data to HHS proved to be cumbersome; as a result, SSA did not provide its appraisal data to HHS for HHS to submit the data to the CPDF for 1995. According to SSA officials, SSA continued to capture these data in its human resource management information system, but HHS did not ask for the data, and SSA was not aware that it was to report them to the CPDF.

The importance of a data element to an agency can affect the level of effort that the agency gives to ensuring the data element's accuracy. For example, some personnelists in the offices we visited said that the accuracy of education level information was "of little concern" to them. In contrast, two other personnel offices reported taking steps to improve the accuracy of this information. Officials in one of these offices (the Pensacola Naval Air Station) told us they had updated the education level information on their employees as part of an overall records review. An official in the other office (State) told us that promotions for certain of their employees are based, in part, on education levels. Therefore, the official said that employees are asked to review such information maintained by the agency and report needed changes.

We also observed in previous work that agency-specific CPDF data could be inaccurate. For example, in 1997 the House Committee on International Relations asked us to examine a discrepancy between the number (17) of Schedule C political appointees¹⁶ reported to Congress by the Agency for International Development (AID) and the number (0) that appeared in the CPDF for the period January 19, 1993, through November 14, 1995. Through our analysis of the CPDF data, we determined that AID used the wrong legal authority when coding the appointing authority for these individuals. As a result, the information in the CPDF (0) did not correctly identify any of the 17 individuals as political appointees.

Inaccuracies in specific agencies' CPDF data, such as SSA not submitting current rating of record data for 2 years and AID using the wrong legal authority code for Schedule C political appointees, can distort users' analyses, findings, and conclusions and result in OPM's reporting on federal agencies that misinforms policymakers and the public. These examples

¹⁶Upon specific authorization by OPM, agencies may make Schedule C appointments to positions excepted from the competitive service that are policy-determining or that involve a close and confidential working relationship with the head of an agency or other key appointed officials.

also show that errors in agency-specific data may go unnoticed for several years and that the accuracy of a particular data element can vary from year to year for a particular agency.

OPM officials told us that they believe that the periodic accuracy measurements that OPM does are a good indicator of problematic data elements governmentwide. For example, OPM's measurement of historical accuracy for fiscal year 1994 discusses why errors occurred and gives error rates for status and dynamics file data elements governmentwide. However, as we said earlier, OPM does not provide the results of these measurements to non-OPM users of CPDF data. Therefore, non-OPM users of CPDF data are most likely not aware of the findings of OPM's accuracy measurements. In addition, OPM officials said that their periodic accuracy measurements are not useful for identifying errors in CPDF data elements at individual agencies. OPM officials said they sometimes become aware of agency-specific inaccuracies in the CPDF when non-OPM users of the data, such as us or the agencies affected, contact OPM about the inaccuracies. For example, OPM said that after it discovered that AID Schedule C appointees were not identified in the CPDF, it began working with AID to improve the future reporting on political appointees.

Awareness of inaccuracies in specific data elements and variation in data accuracy among agencies is important because OPM and non-OPM users rely on CPDF data to monitor and report on individual agencies' demographics, compliance with government policies, or other characteristics. For example:

- OPM's Office of Merit Systems Oversight and Effectiveness uses CPDF data to monitor and report on individual agencies' compliance with selected Merit Systems Principles set out in title 5 of the United States Code;¹⁷
- the National Performance Review used CPDF data in a 1993 report on Transforming Organizational Structures to compare the numbers of federal personnel by occupation;
- the Equal Employment Opportunity Commission used CPDF data in its fiscal year 1991 report to the President and Congress on affirmative employment programs for minorities and women and for hiring, placement, and advancement of people with disabilities in the federal government; and
- we use the data in some of our reports to Congress.

¹⁷5 U.S.C. 2301.

According to these officials, OPM's current approach for measuring CPDF data accuracy is not designed to include representative samples for individual agencies, and such a sample would be significantly larger than the 135 official personnel folders OPM examined to do its latest measurement for fiscal year 1994 data. OPM officials recognize that the results of rigorous measurements of CPDF data accuracy, i.e., measurements designed to test the accuracy of individual agencies' data, could help users of CPDF data determine if the data are sufficiently accurate for their purposes. However, OPM officials believe the cost of doing such measurements would be prohibitive and would not guarantee that users would consider the measurements when working with CPDF data or that agencies would use the results of the measurements to improve the accuracy of their CPDF data submissions.

OPM's Office of the Actuaries Reported That CPDF Data Are Sufficiently Accurate for Estimating the Government's Liability for Future Retirement Payments

OPM's Office of the Actuaries uses CPDF data to help estimate the federal government's liability for future payments of federal retirement programs. According to the actuary responsible for calculating the federal government's liability for future retirement payments to federal employees and their survivors, the office uses CPDF data on adjusted basic pay, sex, birth date, retirement plan, and service computation date in calculating the estimate of this liability.

We discussed with the actuary the error rates we found for these data elements both as measured in our employee questionnaire and in comparison to official personnel folders and records. Except for adjusted base pay, which was about 94-percent accurate in our nongeneralizable comparison of official personnel folders and CPDF data, we found all of these data to be 99 percent or more accurate. We shared these results with the actuary, and he told us that the CPDF data elements were sufficiently accurate for making the liability estimate. The actuary also told us that erroneous national economic assumptions were much more likely to affect his estimate than inaccuracies in the CPDF data. For instance, the actuary said that slight variances in estimated future interest rates or rates of return on investment could have a significant impact on the government's estimated liability for future payments. Furthermore, the actuary said that the CPDF is not the only source of information for certain information the office uses for its estimate.¹⁸ For example, the actuary told us that he makes independent calculations of salaries by using data on contributions

¹⁸The CPDF is only one of three major databases the office uses for liability calculation. The other two databases are the Postal Data File and the Annuitant File.

to pension plans. In addition, OPM received an unqualified opinion on its retirement program financial statements for fiscal year 1997.

Conclusion

Most of the 28 data elements we reviewed were 99 percent or more accurate in the aggregate. A minority of data elements we reviewed, especially education level and rating of record, was much less accurate. OPM has found broadly similar results in its accuracy measurements but has not informed non-OPM users of CPDF data of these results even though the lower level of accuracy for some data elements could affect the validity of analyses relying on those data elements. Further, the accuracy levels that both OPM and we have found are generalizable only governmentwide. Anecdotal evidence from this review and our prior work illustrates that the accuracy of CPDF data elements can vary significantly among agencies. Nevertheless, OPM and non-OPM analysts rely on CPDF data to monitor and report on individual agencies' demographics, compliance with government policies, and other characteristics. OPM officials said that gauging the accuracy of individual data elements by agency would require a significantly larger measurement sample and thus increase its measurement costs. Informing users of CPDF data of the governmentwide accuracy results and a specific caution that individual agencies' results may vary significantly could nevertheless be useful. This would allow analysts and those using CPDF products to make better informed judgments before using agency-specific CPDF data and perhaps to seek information to corroborate the CPDF data.

Recommendation to the Director of OPM

We recommend that the Director of OPM make the results of OPM's measurements of historical accuracy available to all users. To make this information available OPM could post the results of its accuracy measurements on its Internet web site including cautionary language indicating that the accuracy of CPDF data elements may vary by agency. OPM could also inform users of the availability of this information whenever it distributes CPDF data or reports.

Agency Comments and Our Evaluation

In a letter dated September 11, 1998, (see app. VII), the OPM Director said our findings are consistent with OPM's internal quality measures. In particular, the OPM Director cited our draft report's findings that CPDF data, including the data used by OPM's Office of the Actuaries to estimate the government's liability for future retirement payments, were accurate.

Chapter 2
CPDF Data Reviewed Appear to Be Mostly
Accurate in the Aggregate

The OPM Director also said that although our findings were positive, she believed many of the report's headings tended to obscure rather than clarify the findings. In addition, she said that the Results in Brief discussion of CPDF accuracy standards and error rates in education level data is so limited that it presents only our view of CPDF limitations. According to the OPM Director, for "complete and accurate information that provides a more balanced rationale for CPDF specifications, one must look beyond the Results in Brief" to the body of the report.

We believe the view presented in the Results in Brief is balanced. For example, in the first paragraph, we report that about two-thirds of the selected CPDF data elements it reviewed were at least 99-percent accurate. We also disagree that the report's headings tend to obscure rather than clarify the findings. The report's title, chapter titles, and main captions note the positive findings of our review. We believe, as the OPM Director acknowledged, that our report clearly states that most of the CPDF data we reviewed were accurate.

The OPM Director did not specifically refer to our recommendation that she make the results of OPM's historical measurements of the CPDF's accuracy available to all users. However, she said that OPM will make available appropriate explanatory material to all CPDF users. As stated in this chapter, we believe that this explanatory material should include the accuracy measurements.

USERS Generally Reported CPDF Products Met Their Needs, but Further Awareness of Cautions on CPDF Data Could Affect Use of the Data

We used a questionnaire¹⁹ to determine the extent to which selected CPDF users believed (1) the CPDF data they used met their needs, including whether the products were current, accurate, and complete; and (2) they received sufficient cautions about the limitations of CPDF data to use or present the CPDF data correctly. OPM officials identified 247 CPDF users as representing all of the requesters of CPDF data products who corresponded directly with OPM in 1996. We surveyed those 247, and 40 said they did not use CPDF products. Of the remaining 207, 161 responded to our questionnaire as users of the CPDF. The results of our CPDF customer questionnaire showed that the majority of CPDF users responding believed that CPDF products met their needs, including being sufficiently current, accurate, and complete. However, 29 of the 71 CPDF users said knowing about cautions they were not made aware of would have affected the way they used or presented CPDF data. OPM officials said, and respondents' answers to our questionnaire indicated, that the extent to which OPM provided users cautions about the general limitations of the CPDF varied. OPM officials said they were considering creating a CPDF web site that would allow OPM to make CPDF data more widely available and to "bundle" or link specific cautions on limitations associated with specific sets of data.

USERS Generally Reported That CPDF Data Met Their Needs Including Being Current, Accurate, and Complete

OPM distributes a variety of CPDF-based products, including data extracts that consist of selected data elements, e.g., "service computation date" or "duty station," which are provided on tape or diskette to users; recurring reports, such as the Demographic Profile of the Federal Workforce;²⁰ ad hoc reports containing specific information from the CPDF, such as results of matching CPDF data with other data; and the User Simple and Efficient Retrieval (USER) system, which is an information retrieval system that provides electronic access to the CPDF's status and dynamics files. The majority of the respondents to our questionnaire reported that the data in the CPDF products they used met their needs, including being current, accurate, and complete.

For example, when asked about the extent to which CPDF products that they used over the past 2 years (i.e., data extracts, recurring reports, ad hoc reports, and the USER system) met their needs, depending on the type of product, 67 to 81 percent of respondents rated CPDF products as meeting

¹⁹The complete results of GAO's questionnaire appear in appendix V.

²⁰The Demographic Profile of the Federal Workforce report is published biennially by OPM. It replaces the Equal Employment Opportunity Statistics series (1963-1980), and the biennial Affirmative Employment Statistics report last published for September 1990.

Chapter 3
USERS Generally Reported CPDF Products
Met Their Needs, but Further Awareness of
Cautions on CPDF Data Could Affect Use of
the Data

their needs to a great or very great extent. When asked about the extent that these products were current enough to meet their needs, the majority of CPDF users responding to this question reported that the CPDF was, to a great or very great extent, current enough to meet their needs. Seventy to 73 percent of the users who answered this question rated the data products we asked about as current enough for their needs to a great or very great extent.

When asked about the extent to which they believed the CPDF products that they used over the past 2 years were accurate, the majority (65 to 87 percent) of users responding to this question rated the products we asked about as accurate to a great or very great extent. Similarly, the majority (71 to 89 percent) of the users responding to our question about the completeness of CPDF data said they believe the products listed were complete to a great or very great extent. Of those users of CPDF products who reported that specific products met their needs to a great or very great extent, a large majority also reported that those products were accurate and complete.

In addition to the data products that we asked about, 15 respondents to our questionnaire reported they used the Installation Level Data Retrieval System (ILDRS)—a database system that uses CPDF data to provide a “snapshot” of a federal agency’s personnel.²¹ When asked about the extent to which ILDRS was current enough to meet their needs over the past 2 years, unlike the response we got from most users about the currency of CPDF products, only 4 of these 15 respondents rated it as being current enough to meet their needs to a great or very great extent. Eight of the 15 respondents rated ILDRS as being accurate to a great or very great extent, and 9 of the 15 rated ILDRS as being complete to a great or very great extent.

²¹ILDRS is used by evaluators in OPM’s Office of Merit Systems Oversight and Effectiveness in preparing for on-site evaluation activities; preparing for the analysis of installation personnel activity off-site; and producing a variety of statistical indicators to measure the performance of human resource management systems at the bureau, agency, and governmentwide levels.

Most CPDF USERS
Said Cautions OPM
Provides on Data
Limitations Were
Sufficient, but Some
Said Further
Awareness of
Cautions Could Affect
Use of Data

OWI does not provide users of CPDF products with a uniform set of cautions about the limitations of the data elements contained in the CPDF. The extent of the cautions OPM provides about the limitations of CPDF data to users of CPDF-based products varies because, according to OPM officials, the cautions are tailored to the CPDF product being requested. Users responding to our questionnaire demonstrated a wide range of awareness of caution statements about the CPDF data's limitations. The majority of users responding to our questionnaire reported that they were aware of the limitations of the data they received and that the caution statements on limitations provided by OPM were sufficient for them to correctly use the data.

OPM Does Not Disclose to
Users All the Cautions
About the CPDF's
Limitations

Although OPM's CPDF-based governmentwide and ad hoc reports contained some cautions on limitations, none of the reports we reviewed disclosed all of the cautions on the CPDF. We observed that CPDF products, such as ad hoc reports, that OPM prepares to respond to requests for specific information do not fully disclose all 28 cautions about the limitations of the CPDF that OPM officials identified for us.²² For example, OPM's response to a state's request for CPDF data that were to be used in a data match to identify federal employees by selected data elements, such as pay grade, who graduated from state education and training programs cautioned the requester that the CPDF contains records of personnel only in executive branch agencies. OPM did not warn the requester that OPM's quality assurance procedures cannot detect agency miscoding of certain data elements, such as pay grade (e.g., submission of grade 11 when the grade is actually 12). In contrast, the recurring reports that are widely distributed and that contain governmentwide statistics, such as OPM's Biennial Report of Employment by Geographic Area, contained quality measurements of the data in the reports and error rates (i.e., estimated percentage of data elements that failed edit checks) for each of the data elements reported.

OWI analysts routinely monitor and report to agencies submitting data about the quality of their own submissions, that is, the degree to which their data submissions fall within OWI's acceptable range of data values, or edit standards. This information is also made available within OPM and to certain non-OPM users. For example, information on the percentage of data elements not passing CPDF edits and the quality of the CPDF status and dynamics files is currently available through OPM's USER system. According to OPM, we, the Equal Employment Opportunity Commission, the Merit

²²A copy of our questionnaire containing the results from respondents is in appendix VI. For a complete list of the 28 caution statements about the limitations of the CPDF, see question 6.

Chapter 3
USERS Generally Reported CPDF Products
Met Their Needs, but Further Awareness of
Cautions on CPDF Data Could Affect Use of
the Data

Systems Protection Board, the Department of Agriculture, Department of Labor, Environmental Protection Agency, National Guard, National Security Agency, Congressional Budget Office, and the Office of Management and Budget were trained and given access to this system by them. OPM officials reported that they do not know to what extent these agencies use the quality reports available through the USER system.

Although OPM does not make information about the quality of individual agencies' CPDF submissions directly available to nonfederal and most federal users, it bases some caution statements to users about the limitations of CPDF data on this information. For example, in its Demographic Profile of the Federal Workforce as of September 30, 1996, OPM informed users that about 0.4 percent of the total CPDF records available for the report were rejected because they failed edits on key data elements. OPM also cautions users in correspondence responding to requests for information and in its recurring CPDF-based reports, such as OPM's Biennial Report of Employment by Geographic Area, about certain general limitations of the data, such as the exclusion of certain agencies' employees from the CPDF's population coverage. However, OPM does not caution users about other limitations, such as that OPM may change submitted values that are missing or known to be in error.

Most CPDF Users Said
CPDF Products Met Their
Needs, but Some Said
Further Awareness of
Cautions on CPDF Data
Could Affect Use of Data

In our questionnaire to CPDF customers, we asked them to indicate how many of the 28 cautions about the CPDF OPM made them aware of.²³ The CPDF users responding to our questionnaire showed a wide range of awareness of the cautions. For example, more than 95 percent of those answering our question about CPDF cautions said they were cautioned by OPM that certain agencies are exempt from reporting to the CPDF. However, only about 34 percent of those answering the question said they were made aware that OPM may change submitted values that are missing or known to be in error by matching records to older files or making values consistent with statistical assumptions. According to OPM officials, these changes rarely happen; and, when they do, they affect only one or two agencies once every four quarterly status files. Overall, from 72 to 86 percent of the users reported that the caution statements on limitations provided by OPM were sufficient for them to correctly use or present the data contained in the various CPDF products they used to a great or very great extent. However, 29 of the 71 CPDF users said knowing about

²³A copy of our questionnaire containing the results from respondents is in appendix VI. For a complete list of the 28 caution statements about the limitations of the CPDF, see question 6.

Chapter 3
USERS Generally Reported CPDF Products
Met Their Needs, but Further Awareness of
Cautions on CPDF Data Could Affect Use of
the Data

cautions they were not made aware of would have affected the way they used or presented CPDF data.

Of the 28 caution statements about limitations of the CPDF listed on our questionnaire, the 5 that respondents were least aware of were the following: (1) a small number (0.2 percent) of employees have more than 1 record in a CPDF status file; (2) the FBI does not report duty station location for employees outside of the District of Columbia; (3) OPM may change submitted values that are missing or known to be in error by matching records to older files or making values consistent with statistical assumptions; (4) there is no CPDF standard format for submitting employee names; and (5) CPDF status files are generally considered to reflect employment at the end of the quarter, but they might actually reflect employment at the end of the pay period just prior to the end of the quarter.

OWI officials reported that OPM provides information about the specific limitations of a data product to requesters but does not provide information about other limitations, such as the list of 28 caution statements about CPDF data, to all requesters. OPM officials said that making caution statements about CPDF data limitations more widely available might be useful to some users of the data. However, OWI officials believe this alone would not prevent the possible misinterpretation of a specific set of data by a third-party user, i.e., someone who does not receive CPDF data directly from OPM reports or OPM. Because OPM officials are not always aware of the intended use of data requested by users, these officials may not be aware of which of the 28 caution statements would be most beneficial to those users. For example, if a user intended to derive the average education level of the employees of a particular agency but only requested status file data as of a particular date from OPM, OPM officials might not provide the user with the caution statement that some data were collected at the time of appointment, e.g., education level data, but not routinely updated. Therefore, the average education level derived for the agency would not be current and most likely be understated. OWI officials reported they have been considering creating a CPDF web site that would allow OPM to make CPDF data more widely available and allow OPM to bundle specific caution statements on limitations with the sets of data.

Most CPDF USERS Surveyed Rated the Overall Quality of CPDF Products as Excellent or Very Good

OPM sends customer feedback questionnaires to its CPDF users to determine if it is meeting their needs and to solicit suggestions for improvement. We reviewed 149 OPM customer feedback questionnaires for the period covering March 27, 1990, through February 28, 1994, and determined that 140 of the 149 (94 percent) of the CPDF users responding rated the overall quality of the CPDF products they received as very good or excellent. The majority (from about 72 to 84 percent) of the CPDF users responding to our questionnaire also rated the overall quality of the specific CPDF products they used as very good or excellent.

Conclusions

Most of the users of CPDF data we surveyed reported that they believed that the data in those CPDF-based products they used met their needs, including being current, accurate, and complete. The majority of users we sent questionnaires to reported they had received sufficient cautions about the CPDF's limitations to use or present the data correctly. However, although OPM highlighted cautions about CPDF data that are most likely to be applicable to the interests of a particular requester of those data, it did not make all 28 caution statements available to each of those requesters. Some users reported that knowing about cautions they were not made aware of would have affected the way they used or presented CPDF data. In addition, users who obtain CPDF data regularly without a specific request to OPM may not be cautioned about the limitations associated with using the data.

Recommendation to the Director of OPM

We recommend that the Director of OPM ensure that OPM make all 28 caution statements about limitations associated with CPDF data available to all users. In addition, it may be useful for OPM to continue its practice of highlighting cautions on the data limitations of the CPDF that are most likely to be applicable to the interests of a particular requester of CPDF data. To make this information available to all users, OPM could (1) post, on its Internet web site, a complete listing of the 28 caution statements about limitations associated with CPDF data, (2) apprise all recipients of CPDF data of the availability of the caution statements, and (3) implement its proposal to bundle specific cautions on limitations associated with specific sets of data.

Agency Comments and Our Evaluation

In a letter dated September 11, 1998, (see app. VII), the OPM Director said our findings were consistent with OPM's internal quality measures. The OPM Director cited our draft report's findings that most of the users of CPDF

Chapter 3
USERS Generally Reported CPDF Products
Met Their Needs, but Further Awareness of
Cautions on CPDF Data Could Affect Use of
the Data

data we surveyed rated the overall quality of the data excellent to very good and believed they received explanatory material that enabled them to use the data correctly.

The OPM Director also said that although our findings were positive, she believed the Results in Brief section was too skimpy and that many of the report's headings tended to obscure rather than clarify the findings. We believe the view presented in the Results in Brief is balanced. We also disagree that the report's headings tend to obscure rather than clarify the findings. The report's title, chapter titles, and main captions note the positive findings of our review. We believe, as the OPM Director acknowledged, that our report clearly states that most CPDF users' needs were met.

The OPM Director did not specifically refer to our recommendation that she make all 28 caution statements about limitations associated with CPDF data available to all users. However, she said that OPM will make available appropriate explanatory material to all CPDF users. As stated in this chapter, we believe that this explanatory material should include all 28 caution statements about limitations associated with CPDF data. In addition, the OPM Director identified additional agencies that have access to OPM's USER system, which we added to the report where appropriate.

System Software Development Not Documented According to Applicable Federal Guidance, but Software Appears to Implement Edits as Intended

From 1976 to 1995 applicable federal guidance recommended that agencies use a structured approach for operating and maintaining automated information systems, such as the Central Personnel Data System. The guidance suggested that agencies document the life cycle of an automated information system from its initiation through installation and operation. Although the guidance was issued before OPM's major redesign of the System software in 1986, OPM's OIT did not document changes that were made to the System or have independent testing done to ensure that changes to the software would perform as intended. OIT officials said that to their knowledge the System has not had problems processing data reliably and that the System's owner, OPM's OWI, concurred. Our review of 718 of the 763 computer instructions used by the CPDF showed that the System uses instructions that should implement CPDF edits as intended. OIT officials said that for OPM to accomplish its future information technology (IT) goals it will have to follow a structured approach for computer application development. Toward this end, OPM has adopted a software development goal that would require such an approach no later than fiscal year 2002.

OPM Did Not Document an Upgrade of the System's Software as Recommended in Federal Guidance

From 1976 to 1995, federal guidance issued by the National Bureau of Standards²⁴ and other federal agencies said that sufficient planning and documentation are needed for cost-effective operation and maintenance of information systems. This guidance described the need for organizations to adopt a structured, or System Development Life Cycle (SDLC), approach. 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. Although federal guidance recommending that agencies follow best practices for automated information systems' best practices were issued before OPM's major redesign of the System's software in 1986, OIT did not document changes that were made to the System. OIT officials said that to their knowledge there was no effect on the System from their not having used the SDLC approach because they believe the System was still reliable without it.

²⁴The National Bureau of Standards has been replaced by the National Institute of Standards and Technology.

Federal Guidance
Recommended Using a
Structured Approach to
System's Software
Development

From 1976 to 1995, federal guidance existed to assist agencies as they developed computer software applications and made changes in their automated information systems from initiation through operation. For example, on February 15, 1976, the Department of Commerce's National Bureau of Standards issued the Federal Information Processing Standards (FIPS) Publication 38, which provided basic guidance for the preparation of 10 document types that agencies were to use in the development of computer software. FIPS Publication 64, which was issued on August 1, 1979, provided guidance for determining the content and extent of documentation needed for the initiation phase of the software life cycle. In 1995 the Secretary of Commerce approved the withdrawal of nine such guidelines, including FIPS Publications 38 and 64. However, agencies that find these guidelines useful may continue to use them.

The National Bureau of Standards' 1988 Guide to Auditing for Controls and Security: A System Development Life Cycle Approach, was to be used as an audit program for auditing automated information systems under development. It included many guidelines that were published from 1976 through 1984 that described the SDLC approach and its requirements, including documentation. This guide also referenced other federal sources that required documentation, including federal information resource management reports and OMB Circular A-130.²⁵

The federal government does not follow a single SDLC approach, but 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).

SDLC documentation is important because it provides a basis for (1) systematically making decisions while moving through a system's life-cycle phases and establishing a baseline for future changes to the system and (2) auditing systems that are under development. According to federal guidance, software acceptance testing, like other testing of the

²⁵OMB Circular No. A-130 provides uniform governmentwide information resources management policies as required by the Paperwork Reduction Act of 1980, as amended by the Paperwork Reduction Act of 1995, 44 U.S.C. Chapter 35.

Chapter 4
System Software Development Not
Documented According to Applicable
Federal Guidance, but Software Appears to
Implement Edits as Intended

automated information system, must be documented carefully, with traceability of test cases to the system requirements and the acceptance criteria. Without acceptance testing, changes to an automated information system cannot be verified as working as intended.

Ensuring an information system's reliability is not the only reason for following an SDLC approach. The National Bureau of Standards' Guide to Auditing for Controls and Security: A System Development Life Cycle Approach states that if agencies use a structured approach to systems development, the probability increases for a well-defined life cycle and compliance to such a cycle. According to the Guide, an unstructured approach leads to free-form system development that may result in serious omissions. Without a structured approach to software applications development, no assurance exists that adequate testing, verification, validation, and certification will be done; resources will be appropriately expended; the anticipated return on investment will be achieved; or user requirements will be met. In addition, without documentation, 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.

During the evaluation and acceptance phase, the computer instructions that have been written or modified undergo testing to verify that they will perform according to user specifications. Although federal guidance said that some changes to the SDLC may be appropriate "if the subject to be addressed is a major modification to a system rather than the development of a new one," it also said that "the need to continually assess the user's needs (validation) and to ensure the conceptual integrity of the design (verification) are not arguable." Thus, evaluation and acceptance testing is a phase that no agency should leave out of an SDLC. As we have described in guidance for the Year 2000 computing challenge, acceptance testing should be done by an independent reviewer.²⁶ An independent review helps to ensure that internal controls and security are adequate to produce consistently reliable results.

OPM Did Not Document
Upgrade of the System's
Software as Recommended
in Federal Guidance

According to OPM officials, since the System's development in 1972, it has gone through only one major software upgrade, which was done in conjunction with the replacement of the System's hardware. According to an OIT official, in 1985, OPM replaced its existing Honeywell computer with an IBM computer and converted CPDF application programs to run on the

²⁶Year 2000 Computing Crisis: A Testing Guide, Exposure draft, (GAO/AIMD-10.1.21, June 1998).

Chapter 4
System Software Development Not
Documented According to Applicable
Federal Guidance, but Software Appears to
Implement Edits as Intended

new hardware. He also reported that at about the same time, OPM decided to upgrade CPDF capabilities by procuring several commercial software packages as well as designing customized software. According to OIT managers, the software upgrade was done in 1986 to improve the timeliness and accuracy of the CPDF because it was not working efficiently. The OIT managers who were responsible for the System at that time told us that OPM did not document the phases of this major system software modification as recommended in applicable federal guidance under an SDLC approach. Other OIT officials also told us that OPM did not follow an SDLC approach for these 1986 CPDF changes or have documentation that would show that acceptance testing was done. In addition, the testing that was done was not done by an independent reviewer. OPM officials said that because of time constraints, OIT staff who designed the software modifications also did the acceptance testing and did not document it.

Although OIT did not follow an SDLC approach and did not have documentation to show that the 1986 software upgrade passed acceptance tests or that subsequent modifications to the System's software applications worked as intended, its managers said that they believe the System is reliable. They said that they base their beliefs on the fact that OPM's OWI, the System's owner, has not complained that the System is not meeting its needs.

The System Appears
to Implement CPDF
Data Edits Reliably

Because OIT did not document software upgrades and modifications to the System, we could not review this type of documentation as a basis for independently evaluating the extent to which the System is operating as intended. As an alternative, of the 763 total edits (700 call-relational²⁷ and 63 validity²⁸) that the system used at the time we did our work, we reviewed the computer instructions written to implement the 700 (470 dynamic file and 230 status file) call-relational edits and 18 of the 63 validity edits that together check the key status and dynamics data fields. This approach allowed us to indirectly determine if the System would reliably implement CPDF data edits, the computer instructions that are to check the validity of individual data elements.

²⁷The call-relational edits are a series of subroutines or programs within the "Dynamics Main Edit Module" and the "Status Main Edit Module" that control the editing of an agency's dynamics and status submission files. These edits do not "edit" or make corrections to any of the data elements. They produce reports that show which fields or data elements are incorrect or failed validity checks.

²⁸Validity edits check data against a defined range of acceptable values to identify data that fall outside the range.

Chapter 4
System Software Development Not
Documented According to Applicable
Federal Guidance, but Software Appears to
Implement Edits as Intended

Putting test data through the System or otherwise testing the reliability of the System's hardware and software under operating conditions would have allowed us to directly test the reliability of the System. However, we did not attempt to directly test the System's reliability. OPM officials raised a concern about the possible adverse effects of putting test data through the System. They were concerned that putting test data through the System could disrupt its production schedule and introduce "bad" data that could have unforeseeable consequences on the System's operations. Because of the lack of any indications that routine System operations to process agencies' data submissions had caused data errors and the concern raised by OPM, we decided to limit our test of the System's reliability to a review of the computer instructions the System uses to implement edits.

Through our review, we determined that the computer instructions the System uses would implement as intended the selected CPDF call-relational edits and the validity edits used to identify data inconsistencies in the data elements submitted by agencies. We found only one true error. The computer instructions for a dynamics file call-relational edit that is 1 of 20 subprograms used to edit the prior basic pay data element was written in 1995 but was not applied to agencies' dynamics file submissions. CPDF programmers attributed this error to a mistake and oversight on their part and not to a lack of documentation.

OPM Has Implicitly Committed to Adopt an SDLC Approach

In January 1997, OPM initiated a project to develop and implement an Information Technology Architecture Vision, which describes the hardware, software, network, and systems management components of the technical infrastructure required to support OPM business applications and data. This project was initiated in response to various federal government initiatives intended to help ensure that government agencies achieve their missions by changing management practices concerning IT investment and operational decisions.

The first phase of this project was the development of an OPM IT architecture vision, which is intended to provide the framework within which OPM can make IT decisions. OPM published its IT architecture vision in December 1997 and has as one of its components a description of the technology infrastructure that will be needed to support OPM's data and application needs. Under this technology infrastructure component, OPM is to adopt standards for application development and plans to provide training to staff with the goal of reaching a specified software

Chapter 4
System Software Development Not
Documented According to Applicable
Federal Guidance, but Software Appears to
Implement Edits as Intended

development level of process maturity as described in the Capability Maturity ModelSM (CMM®).²⁹

CMM® was developed by the Software Engineering Institute, which is a federally funded research and development center operated by Carnegie Mellon University. It has as a major purpose guiding process improvement efforts in a software organization. CMM® uses five maturity levels—(1) initial, (2) repeatable, (3) defined, (4) managed, and (5) optimizing—to represent evolutionary plateaus on the road to a high level of software process capability. Each maturity level after the first defines several key process areas—groups of related software practices—all of which must be satisfied for an organization to attain that level. An OIT official reported that OPM's IT is at level 1 and has a goal under its IT architecture vision of reaching level 2 or higher by fiscal year 2002. CMM® recommends that an organization use specific software development practices, tools, and methodologies. It does not stipulate how the organization must perform software development or management activities.

For level 2 and higher, CMM® requires an agency to define and document an SDLC approach that is to be used in the development, modification, and management of automated information systems and their software applications. Therefore, by adopting level 2 as a goal, OPM also is committing to follow an SDLC approach by fiscal year 2002. Because an SDLC approach under CMM® applies to the development, modification, and management of all significant systems, once OPM has adopted an SDLC approach, it would need to make changes to the CPDF that would conform to an SDLC approach.

Successfully adopting an SDLC approach would be a significant change for OPM because it said in its IT architecture vision that OPM's application development style has been situational, with few common approaches to system development. The lack of an SDLC was a repeat material weakness reported in independent audits of the financial statements for fiscal years 1996 and 1997 of the retirement program that was administered by OPM's Retirement and Insurance Service.

OIT officials told us that they recognize the importance of having an SDLC approach for accomplishing the applications development goals OPM's IT architecture vision and in its strategic plan for fiscal years 1997 to 2002. In

²⁹Capability Maturity ModelSM is a service mark of Carnegie Mellon University, and CMM® is registered in the U.S. Patent and Trademark Office.

**Chapter 4
System Software Development Not
Documented According to Applicable
Federal Guidance, but Software Appears to
Implement Edits as Intended**

the strategic plan, OPM includes a strategy for ensuring that OPM's mission-critical computer systems, of which the CPDF is one, are Year 2000 compliant in time to ensure that services to customers are not interrupted. This strategy includes detailed tracking of progress on renovation and testing of each IT system and validating and testing that software changes are working as intended. These steps generally conform to SDLC requirements.

Other than efforts for making its information systems Year 2000 compliant, it is not clear whether OPM would follow an SDLC approach when modifying any other systems, including the CPDF, before fiscal year 2002. Neither the IT architecture vision nor the strategic plan specifically identifies when OPM plans to adopt an agencywide SDLC approach.

Conclusions

OPM has not followed an SDLC approach to software development that includes documenting the phases of such development as recommended in applicable federal guidance. OPM also has not documented the testing of changes to software to verify that those changes worked as intended or had such changes tested by an independent reviewer. Nevertheless, although we did not directly test the System's hardware and software under operating conditions, our review of the computer instructions the System uses to implement CPDF call-relational and validity edits shows that the System should implement these edits reliably.

OPM has adopted a goal of achieving at least a CMM® level 2 by 2002, and doing so would require OPM to define and document an agencywide SDLC approach. OPM's current significant modification to CPDF and other mission-critical systems to be Year 2000 compliant is following a structured approach like an SDLC, but it is unclear when OPM might adopt an SDLC approach for other future system changes.

Documentation of system changes in part helps agencies make any future system modifications more quickly and cost effectively, and independent review of system or software changes helps ensure that they will work as intended. Therefore, following these procedures for any changes to the CPDF before OPM adopts an agencywide SDLC could be beneficial.

Recommendation to the Director of OPM

We recommend that the Director of OPM document any changes to the CPDF before OPM adopts an agencywide SDLC approach as specified in CMM®

**Chapter 4
System Software Development Not
Documented According to Applicable
Federal Guidance, but Software Appears to
Implement Edits as Intended**

guidelines and that such changes be independently verified to ensure that they will work as intended.

Agency Comments and Our Evaluation

In a letter dated September 11, 1998, (see app. VII), the OPM Director said our findings are consistent with OPM's internal quality measures. The OPM Director cited our draft report's findings that the CPDF edit programs should function well.

The OPM Director also said that although our findings were positive, she believed many of the report's headings tended to obscure rather than clarify the findings. According to the OPM Director, for "complete and accurate information that provides a more balanced rationale for CPDF specifications, one must look beyond the Results in Brief" to the body of the report.

We believe the view presented in the Results in Brief is balanced. We also disagree that the report's headings tend to obscure rather than clarify the findings. The report's title, chapter titles, and main captions note the positive findings of our review. We believe, as the OPM Director acknowledged, that our report clearly states that the System's edit programs should operate as intended.

The OPM Director agrees with our recommendation that OPM document all future computer system and software changes and perform independent verification that the changes function as intended. She said that OPM is committed to adopting a formal SDLC methodology and is currently in the process of implementing interim measures to ensure that the System is fully documented and continues to function reliably. As an enclosure to her comments, the Director provided OPM's plans for implementing an SDLC methodology.

To Ensure CPDF Data Quality, OWI Provides Guidance to Agencies and Checks Data Before Entering Them in CPDF

This appendix contains an explanation of the process that the Office of Personnel Management's (OPM) Office of Workforce Information (OWI) reported it follows to ensure the quality of the data it enters into the Central Personnel Data File (CPDF).

OWI Says It Provides Agencies With Guidelines on How to Submit Data to the CPDF

According to OWI officials, OWI provides agencies with guidelines on which data elements and personnel transactions are to be reported, when data submissions are to be made, and how the data must be formatted and edited. These guidelines are mostly contained in the following OPM operating manuals: (1) the Guide to the Central Personnel Data File, which lists the data elements, such as current appointment authority and Social Security number that OPM expects agencies to submit from their personnel systems to the CPDF; (2) the CPDF Edit Manual, which provides the edit standards (i.e., specifications and logic for computer instructions) needed for software that checks data quality; (3) the Guide to Personnel Data Standards, which lists data elements and the meanings of their values; and (4) the Guide to Processing Personnel Actions, which provides guidance for the processing of individual personnel actions. Under the guidelines, agencies are responsible for collecting personnel data; editing them for validity, accuracy, and completeness; and furnishing them to the CPDF.

According to the Guide to the CPDF, agencies are to test the data they provide to the CPDF to ensure that the data are accurate and complete. To help agencies ensure the quality of their data, OWI officials told us they provide agencies with the CPDF Edit Manual, which prescribes the data values to which agencies' data are to conform before they are submitted. To test their data's values, agencies are to use the CPDF edits. Agencies are responsible for installing the edits on their automated personnel systems.

These CPDF edits are to check the validity of individual data elements as well as the proper relationship of values among associated data elements. For example, the edit for the sex data element checks that the character used to define the data element is either "M" for male or "F" for female; the edit identifies other characters as errors. OWI expects agencies to incorporate these CPDF edits into their internal personnel data systems. According to OWI officials, these edits constitute the minimum level of quality control; i.e., agencies have the option of incorporating additional quality controls, such as testing the data for accuracy before submitting them, in addition to applying the CPDF edits.

**Appendix I
To Ensure CPDF Data Quality, OWI
Provides Guidance to Agencies and Checks
Data Before Entering Them in CPDF**

However, the CPDF edits cannot detect all types of errors. For example, an edit for the sex data element would not be able to detect if the character “M” was incorrectly used to identify a female employee. According to OWI officials, even though they provide agencies with the edits, errors still occur in submissions and are identified by OWI’s quality review process. According to OWI officials, errors in pay-related data elements often occur at the beginning of the year because agencies make their beginning-of-the-year submissions before they install edits that reflect annual cost-of-living pay increases. The Guide to the CPDF also informs agencies about what data elements should be included in their CPDF data submissions and the frequency and timing of the submissions. Coverage, frequency, and timing differ for two of the CPDF’s databases, or data files—status and dynamics.

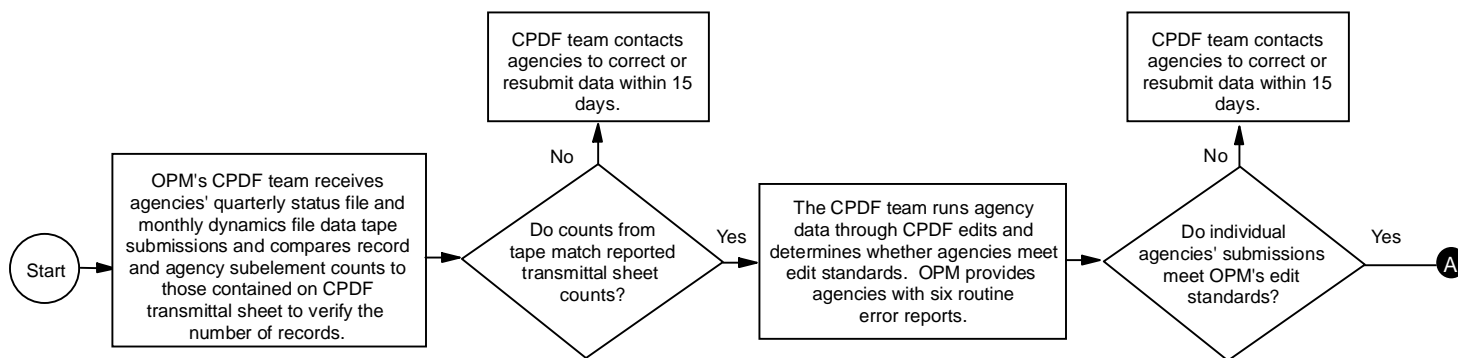
**Purpose of OWI’S
Acceptance Process Is
to Ensure CPDF Data
Quality**

According to OWI officials, when OWI receives agencies’ data submissions, its CPDF team initiates the data acceptance process. OWI reported that the process consists of (1) verifying the number of records³⁰ and agency/subelements, (2) doing edit checks of the data, (3) assessing the aggregate consistency of the data, (4) doing final acceptance reviews of the data, and (5) making the decision to accept the data. Figure I.1 illustrates OWI’s acceptance process.

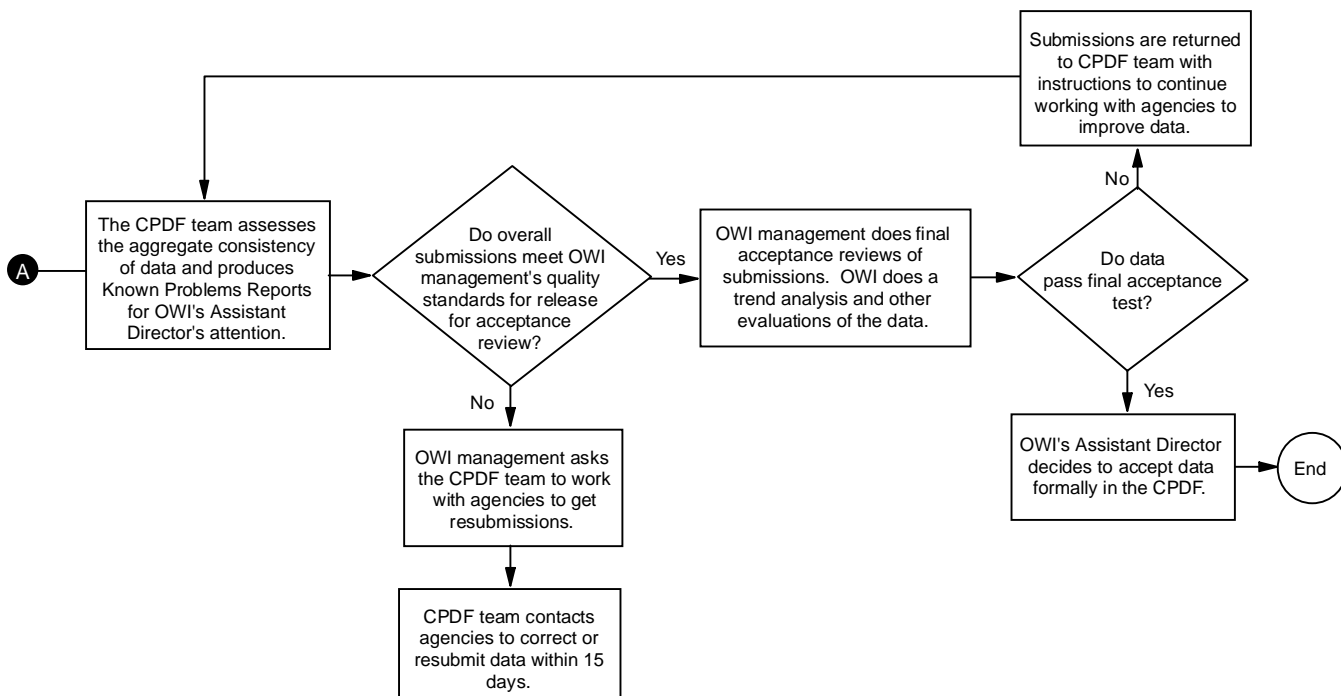
³⁰Each status record submitted represents one employee. Each dynamics record submitted represents one personnel action. In dynamics submissions, one employee may have no records or have several records.

Appendix I
To Ensure CPDF Data Quality, OWI
Provides Guidance to Agencies and Checks
Data Before Entering Them in CPDF

Figure I.1: OWI's Acceptance Process



**Appendix I
To Ensure CPDF Data Quality, OWI
Provides Guidance to Agencies and Checks
Data Before Entering Them in CPDF**



Source: OPM.

Verifying Number of Records

The CPDF team told us that they verify the number of records the agencies send by comparing the number of records and agency/subelements on the agencies' tape submissions to the number on the transmittal sheets (prepared by the agencies) accompanying the submissions. According to the CPDF team members, if there is a discrepancy, a member of the team usually informally contacts the agency to determine why the discrepancy exists. If the discrepancy is not satisfactorily resolved, the team told us they ask the agency to resubmit its data within 15 days.

Doing Edit Checks

According to the CPDF team, after they verify the number of records and agency/subelements, they analyze individual agencies' submissions using the same set of CPDF edits the agencies were expected to use to prepare their submissions. In addition, OWI also produces a Quality Control Action Report that it uses for each status file submission that displays employee population changes, key fields with error rates of more than 1 percent, and changes in data element codes and values. According to OWI officials, OWI produces such a report to evaluate the quality of submissions.

According to OWI, it also provides the results of the edit checks to agencies in reports on their status and dynamics file submissions. These reports are:

- the Status Submission Quality Control Report, which displays the overall quality of the submission, the data elements with errors of more than 1.5 percent, and the 10 most frequently occurring types of errors;
- the Status File Overview Report/Unreleased Status File, which shows the count and percentage of records containing invalid data for those data elements for which a specific count by data value is not provided;
- the CPDF Error File, which shows each incorrect record with codes that identify the errors;
- the Dynamics Submission Quality Control Report, which shows the data elements with errors of more than 1 percent, the 20 most frequently occurring error codes, and the 5 most frequently occurring error codes by nature of action³¹ category;
- the Dynamics Volume and Currency Report, which shows the volume and currency of transactions, cancellations, and corrections; and
- the CPDF Quality Control Report (Unreleased Dynamics Overview), which shows, by data element (or data element group within Nature of Action group) and for all Nature of Actions, the count of records that contain valid data values, the count of records that contain invalid data values, and

³¹The specific personnel action used to create or change a civilian personnel record.

Appendix I
To Ensure CPDF Data Quality, OWI
Provides Guidance to Agencies and Checks
Data Before Entering Them in CPDF

the percentage of error for all records for which data were required or submitted. OWI officials told us these reports are also produced when an agency's resubmission is received.

According to OWI officials, agencies' status files and dynamics files submissions are both reported on, and the CPDF team members told us that they may refer to the reports when they contact agencies about correcting problems with their submissions. Although all the agency personnelists we spoke with said their agencies regularly receive reports from OWI on the quality of their status and dynamics files submissions, only Department of Defense (DOD) officials said they use them as a basis for improving future submissions.

When problems are found with agencies' submissions during edit checks, the CPDF team members told us they ask agencies to fix the problems. Team members and OWI managers and analysts reported that in practice agencies do not always respond to OWI's initial request for corrections, and trying to get agencies to improve their submissions is an ongoing process that includes informal and formal contacts with agencies. CPDF team members said they examine the results of edit checks for individual agency submissions to determine if (1) key status file data elements exceed error thresholds spelled out in the Guide to the CPDF or (2) key dynamics file data elements and non-key data elements exceed the team's judgmental thresholds for allowable error rates.³² Key data elements are the data elements that OPM analysts use most frequently in preparing CPDF reports. The Guide to the CPDF specifically provides that agencies' status file submissions may not be accepted if they contain records with errors, in any of the key status file data fields, that exceed the percentage allowable based on the size of the agency. If the agency's population is 1,000 or greater, no more than 3 percent of the records may have errors or unusable data in any of the 23 key fields identified by OPM. For agencies with populations between 50 and 1,000, no more than 5 percent of the records may have errors or unusable data in any of the key fields (see app. III).

According to the Guide to the CPDF, OWI may reject a dynamics file submission if, in its judgment, the file contains significant errors.

³²The key status file fields are agency/subelement; basic pay; current appointment authority; date of birth; duty station; grade, level, class, rank, or pay band; handicap; health plan; locality adjustment; occupation; pay basis; pay plan; pay status; position occupied; race or national origin; service computation date; sex; special pay table identifier; supervisory status; tenure; veterans preference; veterans status (active military service); and work schedule. The key dynamics file fields are nature of action, effective date of action, Social Security number, and agency/subelement code.

Appendix I
To Ensure CPDF Data Quality, OWI
Provides Guidance to Agencies and Checks
Data Before Entering Them in CPDF

Examples of such errors are the total absence of certain categories of actions (such as accessions or separations) or the total absence of a key data element, such as agency/subelement. CPDF team members reported that in practice, the judgmental threshold for dynamics file submissions is that no more than 50 percent of records can have errors in any of the four key data fields.

OWI officials also reported that the acceptance requirements contained in the Guide to the CPDF are minimum requirements. That is, according to OWI officials, OWI can and does ask for corrections and resubmissions for submissions that meet these requirements but have significant errors in non key fields. CPDF team members told us that dynamics submissions may also be rejected if they exceed the team's judgmental thresholds for allowable error rates. Examples of such errors are failure of any edit by over 20 percent of the records when the number of records is greater than 20, failure of any 2 or more related edits by 10 to 20 percent, or a significant number or percent of fatal errors (those that result in rejection of a record instead of the placement of asterisks in data fields within the record). According to OWI officials, they may also ask agencies to resubmit dynamics data if the dynamics submission does not contain what they believe is a reasonable volume of current records.

If a status or a dynamic file is rejected, OWI officials are to notify the agency's Director of Personnel. Agencies are to correct and resubmit rejected files within 15 calendar days following receipt of the OWI notice. However, the agency officials we spoke with told us they sometimes may not respond to the CPDF team's inquiries about problem submissions because sufficient staff may not be available, or the agency may have higher priorities. However, according to OWI officials the computer software instructions that do the CPDF edit checks place asterisks in those status and dynamics data fields that contain erroneous data to ensure that incorrect data are not knowingly entered into the CPDF.

**Assessing the Aggregate
Consistency of the Data**

While waiting for agencies to respond to CPDF team members' requests for data corrections or resubmissions, OWI officials told us it places all agencies' submissions in a holding status on the system. According to these officials, the amount of time the data are kept in the holding status varies, depending on how long it takes agencies to fix initially identified problems and any additional problems that may be identified as a result of further team analyses. However, OWI officials told us, OWI's internal guidelines establish processing goals of 49 calendar days for status file

Appendix I
To Ensure CPDF Data Quality, OWI
Provides Guidance to Agencies and Checks
Data Before Entering Them in CPDF

submissions and 101 days for dynamics file submissions, from the time they are received by OWI to the time they are entered into the CPDF. OWI's Assistant Director told us he considers these guidelines along with (1) the CPDF team's recommendation; and (2) the Known Problems reports, which are produced by the CPDF team members. These reports summarize the results of agency file submissions and may include a description of significant improvements, a listing of agencies that have not yet submitted data, the status of OWI's requests for those submissions, a summary of any problems with the data identified in earlier reports, and the results of efforts to resolve these problems. According to the Assistant Director, he and his team leaders review the Known Problems reports and decide whether or not to proceed to the next step in creating the file, i.e., combining the submissions into a single file. According to OWI officials, the Known Problems reports may be revised if the CPDF team has planned to continue to work with one or more agencies or if the Assistant Director asks the CPDF team to continue to work with the agencies to get file resubmissions.

**Doing Final Acceptance
Reviews**

According to OWI's Assistant Director, he decides when the data are ready to be released from the holding area on the system for acceptance review. The Assistant Director told us the data are released in the aggregate along with several reports. According to him, the reports include the latest version of the Known Problems reports, which has been updated to show the extent to which status or dynamics file submissions' problems have been solved. In addition, he said these status file reports are released: the CPDF Overview-Released Status File, Status Change, and SF 113-A Benchmarking reports. Three reports are released for the dynamics file as well: the Released Dynamics Volume and Currency, Released Dynamic Overview,³³ and Quarterly Status/Dynamics Compare reports. According to OWI officials, OWI statisticians use these reports along with their own reviews of the files to determine if the file should be made available for general use.

According to OWI officials, OWI statisticians do a trend analysis of the data, which looks for variances. For example, does the number of status file records for agency employees change unexplainably from one reporting period to the next? OWI officials told us that OWI also compares the data to the agencies' Monthly Report of Federal Civilian Employment (SF 113-A) it receives. The SF 113-A report covers all federal civilian hires. In addition,

³³The Released Dynamics Overview report is an updated version of the Unreleased Dynamics Overview report.

**Appendix I
To Ensure CPDF Data Quality, OWI
Provides Guidance to Agencies and Checks
Data Before Entering Them in CPDF**

according to OWI, the data are compared to the agency profiles to identify variances, and data in status files may be compared to interim dynamics file submissions in the status release process. For example, when an employee's pay grade changes between an agency's December and March status file submissions, analysts review the interim dynamics file submissions to determine if the change is reflected. According to OWI officials, the comparison between status and dynamics files is done routinely, producing a standard report, as part of the dynamics release process. They also said that variances or inconsistencies identified through trend analysis, SF 113-A comparison, or status and dynamics file comparison may result in OWI statisticians asking the CPDF team to go back and work with the agencies to improve the files.

**Making the Decision to
Accept**

According to OWI officials, after considering the results of the acceptance review; consulting with internal OPM users of CPDF data (e.g., the Office of Merit Systems Oversight and Effectiveness) and OWI subject area experts and statisticians; and weighing the possible benefit of giving agencies more time to correct data problems versus the cost of delaying the data's release into the CPDF, OWI's Assistant Director makes the final decision about whether or when to allow the data to be added to the CPDF. The Assistant Director said the key consideration in delaying the entry of the data into the CPDF is whether possible improvements in the quality of the data submission warrant the delay. That decision, according to the Assistant Director, is not based on well-defined or well-documented criteria; it is a judgment call he makes.

Although OWI's Assistant Director said that meeting timeliness standards plays a part in making his decision about how much effort and time OWI will spend to improve the data, he told us that data quality is not sacrificed for the sake of timeliness. However, according to the Assistant Director, if the CPDF team and OWI are unable to get an agency to correct the data in a timely manner, the release of file data may be delayed; OWI may substitute status file data from a previous submission instead of the incorrect or problem status file data; or, on rare occasions, in order to maintain a file's useability, OWI may change submitted values that are missing or known to be in error. OWI officials reported that they do not knowingly accept wrong or miscoded data into the CPDF because the edit check process replaces agency data that do not pass the edits with asterisks.

Results of GAO and OPM Measurements of CPDF Accuracy

Data element	Percentage of errors found by GAO approaches		Percentage of errors found by OPM's accuracy measurement for fiscal year 1994
	Results of comparison	Results of employee verification questionnaire	
Annuitant indicator	0.0%	0.0%	0.0%
Effective date of action	0.0	a	a
Duty station	0.0	0.3	0.0
Agency/Subelement	0.0	0.0	0.7
Handicap	0.0	2.7	a
Nature of action code	0.0	a	a
Position occupied	0.0	a	0.0
Race or national origin	0.0	2.0	a
Retirement plan	0.0	0.0	0.0
Social Security number	0.0	0.0	a
Tenure	0.0	a	0.7
Work schedule	0.0	0.0	0.7
Birth date (month and year)	0.9	0.0	0.0
Occupation	0.9	0.0	0.0
Personnel office identifier	0.9	a	0.7
Service computation date (month and year)	0.9	0.4	0.7
Sex	0.9	0.1	1.5
Employee name ^b	1.8	0.9	a
Pay rate determinant	1.8	a	0.0
Legal authority code	b	a	a
Pay plan/grade	2.7	0.7	a
Veterans preference	3.5	0.7	0.0
Adjusted basic pay	6.2	1.2	a
Veterans status	7.1	2.2	5.2
Current appointment authority	9.7	a	4.5
Rating of record	15.9	4.7	5.1
Education	23.0	26.7	8.2

Note: GAO's questionnaire and comparison included status and dynamics file data. The OPM accuracy measurement for fiscal year 1994 results in the table are for status file data only.

^aThis data element was not included in GAO's comparison of the CPDF and agency personnel records.

^bFor this data element, we could not determine the percentage of errors using the universe of 113 employees because some employees had no transactions for fiscal year 1996.

Source: GAO's analysis and OPM's accuracy measurement of 1994 CPDF data.

CPDF Data Elements by File Location, as of February 1998

As of February 1998, the CPDF consisted of 95 separate data elements. Of this number, 68 are reported by agencies in their monthly and quarterly dynamics and status file submissions. Of the remaining 27 data elements, 23 are computer generated and are used by OPM and others in longitudinal surveys and other analyses of federal employees. Agencies use three data elements (effective date of personnel action being corrected, nature of action being corrected, and Social Security number being corrected) only when there are corrections to these key data elements in the dynamics file. Agencies report a final data element, organizational title, in the CPDF's Organizational Component Translation database.

Data element	Reported by agency	Status	File where located	
			Dynamics	
Adjusted basic pay	^a	•	•	
Agency/subelement ^b	•	K		K
Annuitant indicator	•	•		
As of date	^a	•		
Award amount	•		•	
Bargaining unit	•	•		
Basic pay	•	K	•	
Benefit amount	•		•	
Consolidated metropolitan statistical area	^a	•	•	
Cost of living allowance	^a	•	•	
Creditable military service	•	•	•	
Current appointment authority	•	K	•	
Date of birth	•	K	•	
Duty station	•	K	•	
Dynamics category	^a		•	
Education level	•	•	•	
Effective date of personnel action	•			K
Effective date of personnel action being corrected	^c			
Employee name ^d	•			
Fair Labor Standards Act Category	•	•		
Federal employees' group life insurance	•	•		
Federal employees' retirement system coverage	•	•		
Frozen service	•	•	•	

(continued)

**Appendix III
CPDF Data Elements by File Location, as of
February 1998**

Data element	Reported by agency	File where located	
		Status	Dynamics
Functional classification	•	•	
Grade, level, class, rank, or pay band	•	K	•
General Schedule (GS) related grade	^a	•	•
Handicap	•	K	•
Health plan ^e	•	K	
Individual/group award	•		•
Instructional program	•	•	•
Legal authority	•		•
Law enforcement officer geographic pay area	^a	•	•
Locality adjustment	•	K	•
Locality adjustment indicator	^a		
Locality pay area	^a	•	•
Metropolitan statistical area	^a	•	•
Nature of action	•		K
Nature of action being corrected	^b		
Occupation	•	K	•
Occupational category	^a	•	•
OPM oversight office	^a	•	•
OPM service center	^a	•	•
Organizational component	•	•	•
Organizational title	^f		
Pay basis	•	K	•
Pay plan	•	K	•
Pay rate determinant	•	•	•
Pay status	•	K	
Personnel office identifier	•	•	•
Position occupied	•	K	•
Previous retirement coverage	•		•
Prior adjusted basic pay	^a		•
Prior basic pay	•		•
Prior duty station	•		•
Prior grade, level, class, rank, or pay band	•		•
Prior law enforcement officer geographic pay area	^a		•
Prior locality adjustment	•		•

(continued)

**Appendix III
CPDF Data Elements by File Location, as of
February 1998**

Data element	Reported by agency	File where located	
		Status	Dynamics
Prior locality pay area	a		•
Prior occupation	•		•
Prior pay basis	•		•
Prior pay plan	•		•
Prior pay rate determinant	•		•
Prior step or rate	•		•
Prior work schedule	•		•
Processing flag	a		•
Race or national origin	•	K	•
Rating of record (level)	•	•	•
Rating of record (pattern)	•	•	•
Rating of record (period)	•	•	•
Retained grade	•	•	
Retained pay plan	•	•	
Retained step	•	•	
Retention allowance	•	•	•
Retention allowance indicator	a, g		
Retirement plan	•	•	•
Senior pay levels indicator	a	•	•
Service computation date (leave)	•	K	•
Sex	•	K	•
Social security number	•	•	K
Social security number being corrected	b		
Special pay table identifier	•	K	
Staffing differential	•	•	•
Staffing differential indicator	a, g		
Step or rate	•	•	•
Supervisory differential	•	•	•
Supervisory differential indicator	a, g		
Supervisory status	•	K	•
Tenure	•	K	•
Total pay	a	•	•
Type of appointment	a	•	•
U.S. citizenship	•	•	
Veterans preference	•	K	•
Veterans status (active military service)	•	K	•

(continued)

**Appendix III
CPDF Data Elements by File Location, as of
February 1998**

Data element	Reported by agency	File where located	
		Status	Dynamics
Work schedule	•	K	•
Year degree or certificate attained	•	•	•
Total 95	68	23 key	4 key

Note: Key data elements are indicated by the letter "K." The data elements of the status and dynamics files are not mutually exclusive. As the table indicates, in many cases the same data element applies to both files.

^aSome data elements are computer-generated and are used by OPM and others for longitudinal studies or other analysis of federal employees and are not reported by agencies.

^bOnly the two coded positions of this data element that designate the agency are considered key.

^cAgencies use this data element to correct entries made to the preceding data element.

^dAlthough reported in agencies' dynamics submission, to protect the confidentiality of employee records OPM stores names separately from the major CPDF data bases.

^eHealth Plan is required to be submitted only in March and September and is considered key for those submissions.

^fThis data element is reported in a separate CPDF data file called the "Organizational Component Translation" database.

^gThis data element is reported in a separate CPDF data file called the "Longitudinal History File."

Data Elements With the Highest Rate of Inaccuracy

According to our review, the following data elements had the highest level of inaccuracy.

Rating of Record

The rating of record indicates an employee's most recent rating or performance appraisal. We found 18 rating of record mismatches in our review, which compared the 113 employees' official personnel folders or agency records with the CPDF. In one case, the CPDF status file showed that an employee had not been rated when agency records showed a rating. In another example, the CPDF did not reflect the most current rating for seven employees at one personnel office. A human resources official from that office told us that it takes about 2 months from the end of a rating period for a rating to be prepared and entered into the CPDF. Ten Social Security Administration employees had obsolete ratings information in the CPDF. We were informed that the agency had inadvertently failed to update ratings information in the CPDF since fiscal year 1995, but the agency was in the process of correcting the problem.

OPM found a similar situation in a 1992 CPDF accuracy survey. In that survey, OPM noted that "agencies commonly submit . . . rating of record actions to CPDF several months after the ending date of the rating period to which they apply. We believe that most of the errors for rating of record in the CPDF are due to the presence of obsolete or superseded ratings resulting from agency inattention to timely processing of rating of record data."

Education Level

The education level data element is intended to reflect the highest education level achieved by a federal employee. In our review of official personnel folders, we found education level to be inaccurate for 26 of the 113 employees whose records we reviewed. The inaccuracies varied from one personnel office to another. We found inaccuracies ranging from 9 of 20 employees at 1 personnel office to 2 of 20 at another. In 24 of the 26 cases, education levels were understated in the CPDF.

Human resource officials at the personnel offices we visited attributed the relatively high number of errors to several reasons. For example, personnel offices do not always update education level coding for employees who obtain additional education after being hired by the government. Although the higher level of education may be reflected in the employees' personnel folders (e.g., on updated applications for federal employment when employees apply for promotions), the coding is not

Appendix IV
Data Elements With the Highest Rate of
Inaccuracy

necessarily changed in the agencies' automated records or in the CPDF. This may occur because human resource staff are initially concerned about employees meeting the minimum education requirements for their jobs. Any additional education gained by the employees after being hired is of less importance and may not always be reflected when education levels are being coded for agency files and the CPDF. In addition, the CPDF contains 22 education level codes, many of which describe levels of education between formal degree programs. For example, one code is to be used for employees who have had some college courses but less than 30 semester hours; another code is to be used for those who have done some work at a level higher than a 6-year degree but have no additional higher degree. According to some personnel officials, when coding education levels, some personnelists may be more concerned about the employee having a high school diploma or a Bachelor's, Master's, or Doctorate degree and not the levels between degrees. According to the officials, the reason for this may be because education level codes in automated agency files and in the CPDF do not affect pay or any other personnel matter. Therefore, due care is not always exercised when education levels are coded.

In a 1996 CPDF accuracy survey, OPM also noted a relatively high error rate in the CPDF for education level. OPM noted that "education level values appear reliable for determining general educational groupings (e.g., less than high school, high school graduate, some college) but less reliable when used to determine the precise education level."

We also found that data were missing in over 1.0 percent of the records in the September 1996 CPDF status file for education level. Under certain circumstances, agencies may leave some blanks in their data submissions if they do not have complete information on an employee. Also, OPM may delete data that do not pass its edits and replace the data with asterisks. We reviewed the frequency of blank or asterisk-filled data fields in the September 1996 CPDF status file. The highest occurrence of blank or asterisk-filled fields—1.07 percent—was in education level.

GAO's CPDF Employee Verification of Personnel Information Questionnaire

GAO

U.S. General Accounting Office

Central Personnel Data File - Employee Verification of Personnel Information

The United States General Accounting Office, a Congressional agency, is verifying human resource information contained in the Central Personnel Data File (CPDF). The CPDF is a personnel data base maintained by the Office of Personnel Management and contains information on virtually all federal employees in most agencies. Information in the data base is used by the Office of Personnel Management, the Congress, and other oversight agencies to review and modify human resource policies. Therefore, the accuracy of information in the data base is of utmost importance.

In order to evaluate the accuracy of information in the CPDF, we selected a random sample of employees and are asking them to review the accuracy of certain information contained in the data base. Your name was selected in our sample.

We would appreciate your reviewing the information on the attached questionnaire and letting us know if the information is accurate. If you find any inaccuracies, please check the appropriate column and enter the correct information in the space provided.

We would greatly appreciate it if you could return the questionnaire to us within 10 days. You can be assured that we will not use the information you provide us for any purpose other than verifying information in the CPDF. The number on this questionnaire is included only to aid us in our follow-up efforts. Because we are mailing this questionnaire to only a small sample of federal employees, your response is essential to our review.

This questionnaire should take no more than 5 to 10 minutes to complete. After you have finished this questionnaire, please return it in the enclosed pre-addressed, postage paid envelope. If the envelope is missing, please mail the completed questionnaire to:

Mr. Joe Buschy
U. S. General Accounting Office
Denver Regional Office
1244 Speer Boulevard, Suite 800
Denver, CO 80204-3581

If you have any questions, please call Mr. Joe Buschy, collect at (303) 572-7351 or Mr. Dom Nieves, collect at (202) 512-5884.

Thank you for your assistance.

**Appendix V
GAO's CPDF Employee Verification of
Personnel Information Questionnaire**

The following information about you was taken from the Office of Personnel Management's Central Personnel Data File as of **September 30, 1996**. Please look at the information and let us know if it is correct or incorrect by checking the appropriate column. If you are not completely sure about any item, please refer to any records you may have available. For those items that are incorrect, please enter the correct information in the last column. If any of the information printed below was correct as of September 30, 1996 but changed after that date, please indicate that the information is correct. Thank you.

	<u>Correct</u>	<u>Incorrect</u>	<u>If incorrect, enter correction</u>
1. Name	_____	_____	_____
2. Social Security number	_____	_____	_____
3. Employing agency	_____	_____	_____
4. Annual pay, including any locality pay as of <u>September 30, 1996</u>	_____	_____	_____
5. Month and year of birth	_____	_____	_____
6. Duty station where your position was located	_____	_____	_____
7. Pay plan and grade as of September 30, 1996	_____	_____	_____
8. Handicap	_____	_____	_____
9. Occupation as of September 30, 1996	_____	_____	_____
10. Race/national origin	_____	_____	_____

(Continued on next page.)

**Appendix V
GAO's CPDF Employee Verification of
Personnel Information Questionnaire**

	<u>Correct</u>	<u>Incorrect</u>	<u>If incorrect, enter correction</u>
11. Service computation date (month & year)	_____	_____	_____
12. Sex	_____	_____	_____
13. Veterans' preference	_____	_____	_____
14. Veterans' status	_____	_____	_____
15. Work schedule as of September 30, 1996	_____	_____	_____
16. Highest education level as of September 30, 1996	_____	_____	_____
<p>Please refer to Appendix A (gold color sheet) of this questionnaire to find out the exact meaning of the education code entered above. The code for all reported education levels are contained in the attachment. Please review the descriptions of the codes and see if any other code better describes your highest education level.</p>			
17. Your last official performance rating <u>on or prior to September 30, 1996</u>	_____	_____	_____
18. Retirement plan	_____	_____	_____
19. Annuitant indicator	_____	_____	_____
<i>(Continued on next page.)</i>			

**Appendix V
GAO's CPDF Employee Verification of
Personnel Information Questionnaire**

We would appreciate it if you would answer the following two brief questions.

1. For your last official performance rating on or prior to September 30, 1996, which of the following performance rating scales was used to rate your work? *(Check one.)*

- a. 5 level scale
- b. 4 level scale
- c. 3 level scale
- d. 2 level scale (pass/fail)
- e. Do not know
- f. Not rated
- g. Other - Please specify: _____

2. Were you promoted during the 12 month period from October 1, 1995 to September 30, 1996? *(Check one.)*

- a. No → This completes the survey. Thank you.
- b. Yes → Please complete the following statement:

I was promoted **from** pay plan (e.g., GS) _____ grade _____
to pay plan _____ grade _____. The effective date of this promotion
was *(month, day, year)* _____.

Thank you.

Please return your completed questionnaire in the enclosed envelope.

Appendix V
GAO's CPDF Employee Verification of
Personnel Information Questionnaire

APPENDIX A

EDUCATION LEVEL

Code Explanation

- 01 No formal education or some elementary school--did not complete. Elementary school means grades 1 through 8, or equivalent, not completed.
- 02 Elementary school completed--no high school. Grade 8 or equivalent completed.
- 03 Some high school--did not graduate. High school means grades 9 through 12, or equivalent.
- 04 High school graduate or certificate of equivalency.
- 05 Terminal occupational program--did not complete. Program extending beyond grade 12, usually no more than three years; designed to prepare students for immediate employment in an occupation or cluster of occupations; not designed as the equivalent of the first two or three years of a baccalaureate degree program. Includes cooperative training or apprenticeship consisting of formal classroom instruction coupled with on-the-job training.
- 06 Terminal occupational program--certificate of completion, diploma or equivalent. See code 05 above for definition of terminal occupational program. Two levels are recognized: (1) The technical and/or semi-professional level preparing technicians or semiprofessional personnel in engineering and non-engineering fields; and (2) the craftsman/clerical level training artisans, skilled operators, and clerical workers.
- 07 Some college--less than one year. Less than 30 semester hours completed.
- 08 One year college. 30-59 semester hours or 45-89 quarter hours completed.
- 09 Two years college. 60-89 semester hours or 90-134 quarter hours completed.
- 10 Associate Degree. 2-year college degree program completed.
- 11 Three year college. 90-119 semester hours or 135-179 quarter hours completed.
- 12 Four years college. 120 or more semester hours or 180 or more quarter hours completed--no baccalaureate (Bachelor's) degree.
- 13 Bachelor's degree. Requires completion of at least four, but no more than five, years of academic work; includes Bachelor's degree conferred in a cooperative plan or program that provides for alternate class attendance and employment in business, industry, or government to allow student to combine actual work experience with college studies.

Appendix V
GAO's CPDF Employee Verification of
Personnel Information Questionnaire

APPENDIX A

EDUCATION LEVEL

Code Explanation

- 14 Post-Bachelor's. Some work beyond (at a higher level than) the Bachelor's degree but no additional higher degree.
- 15 First professional. Signifies the completion of academic requirements for selected professions that are based on programs requiring at least two academic years of previous college work for entrance and a total of at least six academic years of college work for completion, e.g., Dentistry (D.D.S. or D.M.D.), Law (LL. B. or J.D.), Medicine (M.D.), Theology (B.D.), Veterinary Medicine (D.V.M.), Chiropody or Podiatry (D.S.C. or D.P.), Optometry (O.D.), and Osteopathy (D.O.).
- 16 Post-first professional. Some work beyond (at a higher level than) the first professional degree but no additional higher degree.
- 17 Master's degree. For liberal arts and sciences customarily granted upon successful completion of one (sometimes two) academic years beyond the Bachelor's degree. In professional fields, an advanced degree beyond the first professional but below the Ph.D., e.g., the LL.M.; M.S. in surgery following the M.D.; M.S.D., Master of Science in Dentistry; M.S.W., Master of Social Work.
- 18 Post-Master's. Some work beyond (at a higher level than) the Master's degree but no additional higher degree.
- 19 Sixth-year degree. Includes such degrees as Advanced Certificate in Education, Advanced Master of Education, Advanced Graduate Certificate, Advanced Specialist in Education Certificate, Certificate of Advanced Graduate Study, Certificate of Advanced Study, Advanced Degree in Education, Specialist in Education, Licentiate in Philosophy, Specialist in Guidance and Counseling, Specialist in Art, Specialist in Science, Specialist in School Administration, Specialist in School Psychology, Licentiate in Sacred Theology.
- 20 Post-sixth year. Some work beyond (at a higher level than) the sixth-year degree but no additional higher degree.
- 21 Doctorate degree. Includes such degrees as Doctor of Education, Doctor of Juridical Science, Doctor of Public Health, and the Ph.D. (or equivalent) in any field. Does not include a Doctor's degree that is a first professional degree, per code 15.
- 22 Post-Doctorate. Work beyond the Doctorate.

GAO's CPDF Customer Survey



United States General Accounting Office

Central Personnel Data File (CPDF) Customer Survey

Introduction

The U.S. General Accounting Office (GAO), an independent agency of Congress, is studying the reliability of the Office of Personnel Management's (OPM) Central Personnel Data File (CPDF). As part of our study, we are sending a questionnaire to CPDF customers inside and outside of the Federal government.

We are conducting this survey to obtain information from CPDF customers on their experience with CPDF data extracts and CPDF based reports as well as from subscribers to the CPDF USER System.

CPDF data extracts are selected data elements taken from the CPDF, such as, service computation date, duty station, agency-subelement, etc. These are provided by OPM on tape or diskette.

CPDF reports include recurring reports such as "Pay Structure of the Federal Civil Service" or "ad-hoc" reports which contain specific information from the CPDF, such as, the results of matching CPDF data with other data.

The CPDF USER System (User Simple & Efficient Retrieval) is a computerized information retrieval system based on the CPDF.

Most of the questions in this questionnaire can be answered by checking boxes or filling in blanks. Space has been provided at the end of the questionnaire for any additional comments. The questionnaire should take about 30 minutes to complete.

GAO will take steps to safeguard the privacy of your responses. The number on this questionnaire and the name asked for below are included only to aid us in our follow-up efforts. Before releasing our report, we will remove the number and name so that your questionnaire can no longer be matched. Survey results will be reported in summary form only.

Please return your completed questionnaire in the enclosed pre-addressed, postage-paid envelope within 10 working days of receipt. In the event the envelope is misplaced, the return address is:

U.S. General Accounting Office
Mr. Dom Nieves
441 G Street, N.W., Room 2908
Washington, DC 20548

If you have any questions, please call Dom Nieves at (202) 512-5884.

We urge you to complete this questionnaire. We cannot develop meaningful information without your frank and honest answers. Thank you very much for your time.

This questionnaire should be answered by the person (or in consultation with the person) who is most familiar with the data. Please enter this person's name, title, organizational affiliation, and telephone number below in case we need to contact him or her in order to clarify a response.

Name: _____

Title: _____

Organization: _____

Telephone: (____) _____

Appendix VI
GAO's CPDF Customer Survey

1. Which of the following CPDF products have you used over the past 2 years? *(Check all that apply.)*

- N=76 CPDF data extracts - i.e., selected data elements taken from CPDF provided on tape or diskette
- N=73 CPDF recurring reports - i.e., scheduled reports such as "Pay Structure of the Federal Civil Service"
- N=79 CPDF "ad-hoc" reports - i.e., reports containing specific information from the CPDF
- N=97 CPDF USER System - i.e., User Simple & Efficient Retrieval System
- N=39 Other CPDF based product - Please specify: _____

2. Over the past 2 years, how often have you used each of the following CPDF products?
(Check one box in each row.)

CPDF Products		Very often (At least once a month) (1)	Often (About 6-11 times a year) (2)	Occasionally (About 1-5 times a year) (3)	Rarely (About 3-4 times in the past 5 years) (4)	Never (5)
a. CPDF data extracts	N=135	17.8%	11.9%	18.5%	5.2%	46.7%
b. CPDF recurring reports	N=136	9.6%	16.2%	22.1%	2.9%	49.3%
c. CPDF "ad-hoc" reports	N=135	14.8%	17.0%	15.6%	7.4%	45.2%
d. CPDF USER System	N=136	24.3%	16.9%	14.0%	4.4%	40.4%
e. Other CPDF based product - Specify: _____	N=135	8.9%	5.9%	6.7%	0%	78.5%

Appendix VI
GAO's CPDF Customer Survey

3. Over the past 2 years, for which of the following general purposes did you use the following CPDF products?
(Check all boxes that apply.)

CPDF Products	Government- wide research (1)	Research on individual federal agencies (2)	Research on individual types of employees (e.g., Hispanic, Schedule C, etc.) (3)	Other purposes- <i>Please check box and specify general purpose below.</i> (4)
a. CPDF data extracts	N=41	N=41	N=33	<input type="checkbox"/> N=19
b. CPDF recurring reports	N=56	N=34	N=31	<input type="checkbox"/> N=10
c. CPDF "ad-hoc" reports	N=54	N=51	N=35	<input type="checkbox"/> N=14
d. CPDF USER System	N=54	N=65	N=43	<input type="checkbox"/> N=12
e. Other CPDF based product - Specify: _____	N=12	N=19	N=11	<input type="checkbox"/> N=15

4. Over the past 2 years, for what specific purposes did you use CPDF products (e.g., study of federal pay structure)?

CPDF Product Name:
(e.g., "Pay Structure of the Federal
Civil Service" report, USER system, etc.)

Specific purpose:

Written entry - N=128
No written entry - N=9

Appendix VI
GAO's CPDF Customer Survey

5. The listing below contains data elements in the CPDF. Please indicate those data elements you made use of in your work over the past 2 years. *(Check all that apply.)*

- | | | | |
|-------|---------------------------------------------|-------|---------------------------------------|
| N=58 | Adjusted basic pay | N=59 | Organizational component |
| N=117 | Agency | N=46 | Pay basis |
| N=102 | Agency-subelement | N=100 | Pay plan |
| N=13 | Annuitant indicator | N=33 | Pay rate determinant (PRD) |
| N=49 | Award amount | N=41 | Pay status |
| N=26 | Bargaining unit status | N=62 | Personnel office identifier (POI) |
| N=60 | Basic pay | N=48 | Position occupied |
| N=50 | Consolidated MSA (CMSA) | N=68 | Race or national origin (RNO) |
| N=18 | Cost of living adjustment | N=56 | Rating (performance level) |
| N=21 | Creditable military service | N=29 | Retained grade |
| N=58 | Current appt. authority (CAA1) | N=23 | Retained pay plan |
| N=52 | Current appt. authority (CAA2) | N=23 | Retained step or rate |
| N=41 | Date of birth | N=25 | Retention allowance |
| N=81 | Duty station (state, city, county);location | N=26 | Retirement plan |
| N=40 | Educational level | N=24 | Senior pay levels indicator |
| N=47 | Employee name | N=47 | Service computation date (SCD) |
| N=27 | Fair Labor Standards Act (FLSA) category | N=74 | Sex |
| N=42 | Functional classification | N=42 | Social security number |
| N=102 | Grade | N=18 | Special pay table ID |
| N=77 | GS related grade | N=16 | Staffing differential |
| N=36 | Handicap | N=47 | Step or rate |
| N=13 | Health plan | N=19 | Supervisory differential |
| N=45 | Individual/group award | N=61 | Supervisory status |
| N=7 | Instructional program | N=56 | Tenure |
| N=25 | Locality adjustment | N=36 | Total pay |
| N=23 | Locality adjustment area | N=74 | Type of appointment |
| N=58 | Metro. Statistical Area (MSA) | N=18 | U.S. citizenship status |
| N=81 | Nature of personnel action | N=58 | Veterans preference |
| N=103 | Occupation | N=40 | Veterans status (not Vietnam era vet) |
| N=86 | Occupational category (PATCO) | N=31 | Vietnam era vet |
| N=18 | OPM area office | N=73 | Work schedule |
| N=17 | OPM regional office | N=16 | Year degree attained |
| | | N=12 | Other - Specify: _____ |

**Appendix VI
GAO's CPDF Customer Survey**

6. Please answer the questions in column A and column B for each caution statement listed below.

*(Check **yes**, **no**, or **not applicable [N/A]** in **both** columns for **each** caution statement. If you don't remember if you were made aware of a caution by OPM, check "DR" [Don't remember] in column A. If you check "DR" you should still respond in column B.)*

Caution statements	Column A		Column B	
	Were you made aware of caution by OPM?		Did (Would) knowledge of caution help you use data?	
1. CPDF is not a personnel accounting system. Its intended use is to support statistical analyses for governmentwide personnel management programs.	N=124 Yes=83.1% DR=10.5%	No=6.5% N/A=9	N=114 Yes=73.7% N/A=14	No=26.3%
2. Certain agencies (e.g., Postal Service, Congress, CIA, etc.) are exempt from reporting to CPDF.	N=127 Yes=95.3% DR=2.4%	No=2.4% N/A=6	N=120 Yes=86.7% N/A=11	No=13.3%
3. Agencies do not report data for certain employees (e.g., Public Health Service's Commissioned Officer Corps, nonappropriated fund employees, foreign nationals employed overseas).	N=127 Yes=83.5% DR=4.7%	No=11.8% N/A=6	N=118 Yes=85.6% N/A=13	No=14.4%
4. A small number (0.2%) of employees have more than one record in a CPDF Status file. Most are either transferees between agencies reported by both their old and new agencies, or experts and consultants.	N=124 Yes=43.5% DR=16.9%	No=39.5% N/A=9	N=104 Yes=57.7% N/A=25	No=42.3%
5. All CPDF data are submitted by agencies from their own separate personnel systems.	N=127 Yes=93.7% DR=2.4%	No=3.9% N/A=6	N=119 Yes=81.5% N/A=13	No=18.5%
6. Some CPDF data are collected at the time of the appointment and not routinely updated (e.g., education level may reflect a bachelor's degree at the time of the employee's appointment but not the master's degree the employee subsequently earned).	N=123 Yes=67.5% DR=9.8%	No=22.8% N/A=10	N=109 Yes=75.2% N/A=21	No=24.8%
7. FBI does not report duty location for employees outside of the District of Columbia.	N=117 Yes=52.1% DR=12.0%	No=35.9% N/A=16	N=94 Yes=58.5% N/A=36	No=41.5%
8. FBI does not report Dynamics data (e.g., promotions, etc.) for any of its employees.	N=115 Yes=57.4% DR=13.0%	No=29.6% N/A=18	N=91 Yes=61.5% N/A=40	No=38.5%
9. OPM runs validity and relationship edits on CPDF data elements to ensure codes are valid and consistent with other related data elements (e.g., if pay plan is GS then grade must be 01-15), but these edits cannot detect miscoding by the agency (e.g., submission of grade 11 when grade is actually 12).	N=123 Yes=78.0% DR=8.9%	No=13.0% N/A=10	N=114 Yes=81.6% N/A=17	No=18.4%
10. OPM places asterisks in data fields that fail the edits to prevent invalid data from reaching CPDF.	N=121 Yes=66.9% DR=11.6%	No=21.5% N/A=10	N=114 Yes=78.9% N/A=15	No=21.1%
11. Error rates vary from agency to agency.	N=126 Yes=77.0% DR=10.3%	No=12.7% N/A=7	N=115 Yes=73.9% N/A=15	No=26.1%
12. Error rates vary from data element to data element.	N=124 Yes=76.6% DR=11.3%	No=12.1% N/A=9	N=113 Yes=74.3% N/A=16	No=25.7%

Continued on next page.

Appendix VI
GAO's CPDF Customer Survey

6. (Continued.) Please answer the questions in column A and column B for each caution statement listed below.

(Check yes, no, or not applicable [N/A] in both columns for each caution statement. If you don't remember if you were made aware of a caution by OPM, check "DR" [Don't remember] in column A. If you check "DR" you should still respond in column B.)

Caution statements	Column A		Column B	
	Were you made aware of caution by OPM?		Did (Would) knowledge of caution help you use data?	
13. OPM performs periodic accuracy surveys and has estimates of error rates for many CPDF data elements.	N=125 Yes=57.6% DR=17.6%	No=24.8% N/A=8	N=116 Yes=75.0% N/A=15	No=25.0%
14. In order to maintain a file's useability OPM may, in rare circumstances (one or two agencies once every four quarterly files) change submitted values that are missing or known to be in error by matching records to older files or making values consistent with statistical assumptions.	N=122 Yes=34.4% DR=15.6%	No=50.0% N/A=11	N=106 Yes=54.7% N/A=23	No=45.3%
15. Once a CPDF file is released, it is no longer subject to correction.	N=124 Yes=51.6% DR=21.8%	No=26.6% N/A=10	N=106 Yes=62.3% N/A=25	No=37.7%
16. CPDF Status files are generally considered to reflect employment at the end of the quarter but might actually reflect employment at the end of the pay period just prior to the end of the quarter.	N=126 Yes=50.8% DR=15.1%	No=34.1% N/A=7	N=108 Yes=59.3% N/A=23	No=40.7%
17. CPDF Dynamics files are generally considered to contain all personnel actions occurring in a quarter, but because files cannot remain open for agency submission for an indefinite period of time, they are closed after a sufficient period of time and late actions may end up in the file for a subsequent quarter.	N=118 Yes=60.2% DR=18.6%	No=21.2% N/A=14	N=104 Yes=67.3% N/A=26	No=32.7%
18. Changes in agency organizations and personnel classifications over time require changes in CPDF codes. Sometimes the definition of an existing CPDF code changes (e.g., the definition for DD04 changed from Defense Communications Agency to Defense Information Systems Agency).	N=122 Yes=78.7% DR=9.8%	No=11.5% N/A=11	N=118 Yes=85.6% N/A=14	No=14.4%
19. Sometimes the item represented by a CPDF code remains the same but the code itself changes (e.g., the code for the Secret Service changed from TR12 to TRAC during an overall restructuring of Treasury Codes in 1994).	N=121 Yes=59.5% DR=14.0%	No=26.4% N/A=12	N=111 Yes=77.5% N/A=20	No=22.5%
20. Sometimes CPDF codes are deleted (e.g., the deletion of the separate codes for supervisors and managers, codes 1 and 3 respectively).	N=120 Yes=63.3% DR=12.5%	No=24.2% N/A=14	N=112 Yes=83.9% N/A=19	No=16.1%
21. Sometimes CPDF codes are added (e.g., the creation of the combined code for supervisors and managers, code 2).	N=119 Yes=64.7% DR=12.6%	No=22.7% N/A=14	N=111 Yes=82.9% N/A=19	No=17.1%
22. All CPDF salaries, except those for fee basis and piece work, are annualized based on a full-time work schedule and therefore overstate the salaries of seasonal, intermittent, and part-time employees.	N=123 Yes=56.1% DR=16.3%	No=27.6% N/A=11	N=108 Yes=73.1% N/A=24	No=26.9%
23. The salary fields do not represent earnings so they exclude any night or overtime pay.	N=121 Yes=59.5% DR=17.4%	No=23.1% N/A=13	N=100 Yes=73.0% N/A=31	No=27.0%

Continued on next page.

**Appendix VI
GAO's CPDF Customer Survey**

6. (Continued.) Please answer the questions in column A and column B for each caution statement listed below.

(Check *yes*, *no*, or *not applicable [N/A]* in *both* columns for *each* caution statement. If you don't remember if you were made aware of a caution by OPM, check "DR" [Don't remember] in column A. If you check "DR" you should still respond in column B.)

Caution statements	Column A		Column B	
	Were you made aware of caution by OPM?		Did (Would) knowledge of caution help you use data?	
24. Total pay (basic pay plus any locality adjustment, staffing or supervisory differential, and retention allowance) is intended to represent the total pay to which an individual is entitled. Unlike other salary fields, it is not capped to reflect any applicable pay caps.	N=116 Yes=61.2% DR=12.1%	No=26.7% N/A=17	N=101 Yes=73.3% N/A=28	No=26.7%
25. Where the salary is blank, the agency did not report the salary, or the salary was blanked out due to incorrect values.	N=119 Yes=57.1% DR=21.0%	No=21.8% N/A=15	N=101 Yes=70.3% N/A=31	No=29.7%
26. Although the vast majority of names in CPDF conform to a last name, first name format, there is no CPDF standard format for submitting names.	N=111 Yes=24.3% DR=27.0%	No=48.6% N/A=23	N=93 Yes=51.6% N/A=40	No=48.4%
27. Dynamics actions placing seasonal employees in or out of pay status may not be reported in a timely manner or may, in some cases, be missing from CPDF Dynamics files.	N=115 Yes=41.7% DR=24.3%	No=33.9% N/A=18	N=101 Yes=74.3% N/A=30	No=25.7%
28. CPDF counts may differ from other sources of OPM data due to differences in coverage and reporting periods as well as errors inherent in each data source.	N=124 Yes=76.6% DR=9.7%	No=13.7% N/A=9	N=114 Yes=87.7% N/A=17	No=12.3%

7. Overall, to what extent, if at all, do you believe the information on cautions provided by OPM was sufficient for you to correctly use/present the data contained in the following CPDF products?
(Check one box in each row.)

CPDF Products		To a very great extent (1)	To a great extent (2)	To a moderate extent (3)	To little extent (4)	To no extent (5)	Not applicable, did not use product (6)
		a. CPDF data extracts	N=72	40.3%	40.3%	13.9%	
b. CPDF recurring reports	N=69	47.8%	37.7%	11.6%	2.9%	0%	N=67
c. CPDF "ad-hoc" reports	N=75	50.7%	33.3%	12.0%	2.7%	1.3%	N=61
d. CPDF USER System	N=79	46.8%	35.4%	15.2%	1.3%	1.3%	N=57
e. Other CPDF based product - Specify:	N=28	53.6%	17.9%	21.1%	0%	7.1%	N=106

**Appendix VI
GAO's CPDF Customer Survey**

8. For those cautions that you may not have been made aware of by OPM, would knowledge of these cautions have affected the way you used/presented the data from CPDF? *(Check one and explain your response.)*

N=71
40.8% Yes
59.2% No

N=59 Not applicable, I was made aware of all cautions that were related to my work by OPM

If you responded "Yes" or "No", please explain your response: _____

Written entry - N=52
No written entry - N=85

9. To what extent, if at all, was (were) the CPDF product(s) you used over the past 2 years current enough to meet your needs? *(Check one box in each row.)*

CPDF Products	To a very great extent (1)	To a great extent (2)	To a moderate extent (3)	To little extent (4)	To no extent (5)	Not applicable, did not use product (6)
a. CPDF data extracts N=71	40.8%	32.4%	23.9%	2.8%	0%	N=64
b. CPDF recurring reports N=69	36.2%	33.3%	29.0%	1.4%	0%	N=67
c. CPDF "ad-hoc" reports N=71	46.5%	25.4%	19.7%	8.5%	0%	N=61
d. CPDF USER System N=78	41.0%	30.8%	20.5%	7.7%	0%	N=56
e. Other CPDF based product - Specify: _____ N=27	25.9%	18.5%	37.0%	18.5	0%	N=107

If you answered "To little extent" or "To no extent" (columns 4 or 5) for any product in question 9 above, please explain your responses.

CPDF Product Name: **Explanation:**
(e.g., "Pay Structure of the Federal Civil Service" report, USER system, etc.)

Written entry - N=15
No written entry - N=122

Appendix VI
GAO's CPDF Customer Survey

10. Did you or anyone else in your organization independently validate the accuracy of the data contained in the CPDF product(s) that you used over the past 2 years? *(Check one.)*

Note: By "accuracy" we mean the extent to which the data does not contain erroneous information about federal employees.

N=128

64.1% No, accuracy of the data was not validated

35.9% Yes, accuracy of the data was validated → *Please identify the product(s) and briefly describe how the data were verified.*

CPDF Product Name:

(e.g., "Pay Structure of the Federal Civil Service" report, USER system, etc.)

Description of how data were verified for accuracy:

Written entry - N=53

No written entry - N=84

.....

.....

.....

11. Did you or anyone else in your organization independently validate the completeness of the data contained in the CPDF product(s) that you used over the past 2 years? *(Check one.)*

Note: By "completeness" we mean the extent to which the data includes information for all federal employees in all data fields for those agencies supplying data to the CPDF.

N=128

79.7% No, completeness of the data was not validated

20.3% Yes, completeness of the data was validated → *Please identify the product(s) and briefly describe how the data were verified.*

CPDF Product Name:

(e.g., "Pay Structure of the Federal Civil Service" report, USER system, etc.)

Description of how data were verified for completeness:

Written entry - N=34

No written entry - N=103

.....

.....

.....

Appendix VI
GAO's CPDF Customer Survey

12. Overall, to what extent do you believe the data contained in the CPDF products you have used over the past 2 years were accurate and complete? *(Check one box in each row.)*

		To a very great extent (1)	To a great extent (2)	To a moderate extent (3)	To little extent (4)	To no extent (5)	Not applicable, did not use product (6)
Accuracy of CPDF Product							
a. CPDF data extracts	N=71	25.4%	57.7%	16.9%	0%	0%	N=63
b. CPDF recurring reports	N=67	29.9%	56.7%	13.4%	0%	0%	N=67
c. CPDF "ad-hoc" reports	N=74	29.7%	47.3%	20.3%	2.7	0%	N=61
d. CPDF USER System	N=79	25.3%	48.1%	25.3%	1.3%	0%	N=56
e. Other CPDF based product - Specify:	N=26	26.9%	38.5%	34.6%	0%	0%	N=106
Completeness of CPDF Product							
a. CPDF data extracts	N=69	30.4%	52.2%	15.9%	1.4%	0%	N=63
b. CPDF recurring reports	N=64	23.4%	65.6%	10.9%	0%	0%	N=64
c. CPDF "ad-hoc" reports	N=70	27.1%	52.9%	15.7%	4.3	0%	N=61
d. CPDF USER System	N=75	28.0%	46.7%	22.7%	2.7%	0%	N=55
e. Other CPDF based product - Specify:	N=24	33.3%	37.5%	29.2%	0%	0%	N=106

Overview

13. Thinking of all the CPDF products that you used over the past 2 years, how would you rate them in terms of overall quality? *(Check one box in each row.)*

CPDF Products		Excellent (1)	Very good (2)	Adequate (3)	Poor (4)	Very poor (5)	Not applicable, did not use product (6)
a. CPDF data extracts	N=73	39.7%	43.8%	13.7%	2.7%	0%	N=82
b. CPDF recurring reports	N=71	28.2%	54.9%	16.9%	0%	0%	N=85
c. CPDF "ad-hoc" reports	N=77	31.2%	41.6%	23.4%	3.9%	0%	N=79
d. CPDF USER System	N=92	28.3%	46.7%	20.7%	4.3%	0%	N=62
e. Other CPDF based product - Specify:	N=35	42.9%	28.6%	20.0%	8.6%	0%	N=119

Appendix VI
GAO's CPDF Customer Survey

If you answered "Poor" or "Very poor" (columns 4 or 5) for any product in question 13 on the previous page, please explain your responses.

CPDF Product Name: **Explanation:**

(e.g., "Pay Structure of the Federal Civil Service" report, USER system, etc.)

Written entry - N=10
No written entry - N=127

.....

.....

.....

14. Thinking of all the CPDF products that you used over the past 2 years, to what extent, if at all, did they meet your needs? *(Check one box in each row.)*

CPDF Products		To a very great extent (1)	To a great extent (2)	To a moderate extent (3)	To little extent (4)	To no extent (5)	Not applicable, did not use product (6)
a. CPDF data extracts	N=71	45.1%	35.2%	18.3%	1.4%	0%	N=63
b. CPDF recurring reports	N=67	32.8%	43.3%	23.9%	0%	0%	N=68
c. CPDF "ad-hoc" reports	N=73	38.4%	42.5%	15.1%	4.1%	0%	N=62
d. CPDF USER System	N=79	35.4%	41.8%	17.7%	5.1%	0%	N=56
e. Other CPDF based product - Specify: _____	N=27	37.0%	29.6%	33.3%	0%	0%	N=106

If you answered "To little extent" or "To no extent" (columns 4 or 5) for any product in question 14 above, please explain your responses.

CPDF Product Name:

Explanation:

(e.g., "Pay Structure of the Federal Civil Service" report, USER system, etc.)

Written entry - N=8
No written entry - N=129

.....

.....

.....

.....

Appendix VI
GAO's CPDF Customer Survey

15. Have you used or considered using **non-CPDF based information sources** instead of CPDF-based products for information on the Federal civilian workforce? *(Check one.)*

N=134

46.3% Yes

53.7% No

Written entry - N=94

No written entry - N=43

Please explain your answer: _____

16. Please use the space below to describe any problems you had (either documented or suspected) with the CPDF product(s) concerning their accuracy, completeness, or any other quality-related factor.

CPDF Product Name:
(e.g., "Pay Structure of the Federal Civil Service" report, USER system, etc.)

Problems:

Written entry - N=62

No written entry - N=75

17. If you have any other comments concerning the CPDF, such as the advantages or benefits you have experienced using CPDF-based products, please use the space below. If necessary, you may add additional sheets.

Written entry - N=63

No written entry - N=74

Thank you for your response.
Please return your completed questionnaire in the enclosed preaddressed envelope.

Comments From the Office of Personnel Management



OFFICE OF THE DIRECTOR

UNITED STATES
OFFICE OF PERSONNEL MANAGEMENT

WASHINGTON, D.C. 20415

SEP 11 1998

Mr. Michael Brostek
Associate Director, Federal Management
and Workforce Issues
U.S. General Accounting Office
Washington, DC 20548

Dear Mr. Brostek:

Thank you for sharing your draft report on our Central Personnel Data File (CPDF) with us. We are pleased that after a comprehensive two year study of the CPDF, the General Accounting Office has confirmed that the system is operating well.

The report's findings are consistent with our own internal quality measures, as you mentioned. In summary, your report found that:

- CPDF data were very accurate. As you noted, the majority of the data elements were 99 percent accurate. Depending on the methodology, 63 percent or 70 percent of the data elements were at least 99 percent accurate and 35 percent to 44 percent were 100 percent accurate.
- All data used by the Office of Personnel Management's (OPM) Office of the Actuaries to help estimate the Federal Government's liability for future payments of federal retirement funds were 99 percent or more accurate, except adjusted basic pay which was about 94 percent accurate.
- CPDF reports met the information needs of those who used them. In your survey of recipients of CPDF reports, from 67 percent to 81 percent, said the reports met their needs to a very great extent or to a great extent.
- Users were impressed with the quality of CPDF's information. In your survey, from 72 percent to 84 percent respondents rated the overall quality of CPDF products as excellent to very good.
- Users believed they received explanatory material that enabled them to use CPDF data correctly. In your survey, 72 percent to 86 percent of the respondents said this was true to a very great extent or to a great extent.
- The CPDF edit programs function well. In your review of 718 edits, you found one error that affected prior pay in one pay plan, representing 0.02 percent of CPDF employment. That error was corrected in January 1998.

**Appendix VII
Comments From the Office of Personnel
Management**

Mr. Michael Brostek

2

While we believe the findings themselves are very positive, many of your report's headings tend to obscure rather than clarify the findings. In addition, the Results in Brief discussion of CPDF accuracy standards and error rates in educational level data are so limited as to present only GAO's views of system limitations. For complete and accurate information that provides a more balanced rationale for CPDF system specifications, one must look beyond the Results in Brief. The result leaves the reader with an erroneous and misleading impression of data accuracy.

Your statement that only the General Accounting Office, Equal Employment Opportunity Commission, and the Merit Systems Protection Board have been approved for access to OPM's USER system should be changed. In addition to these three agencies, the Department of Agriculture, Department of Labor, Environmental Protection Agency, National Guard, National Security Agency, Congressional Budget Office, and the Office of Management and Budget were trained and given access to this system.

As you noted, the CPDF data are used extensively not only within the Office of Personnel Management but by other central management agencies and the public. We are committed to continuing to provide quality information services to all of those customers. As part of our ongoing efforts to improve our services, we will ensure that appropriate explanatory material is available to our customers.

We agree with your recommendation that we fully document all future computer system and software changes, and perform independent verification that the changes function as intended. As noted in your report, we are committed to adopting a formal Systems Development Life Cycle methodology and are currently in the process of implementing interim measures to ensure that the CPDF is fully documented and continues to function reliably. Additional information on this topic is enclosed.

We appreciate the opportunity to comment on your draft report.

Sincerely,


Janice Lachance
Director

Enclosure

**Appendix VII
Comments From the Office of Personnel
Management**

Enclosure

Systems Development Life Cycle Methodology Implementation

The Retirement and Insurance Service's Systems Modernization and Maintenance Division (SMMD), which maintains the CPDF, is an active participant in the OPM Chief Information Officer's (CIO) initiative to develop and implement SDLC processes in the OPM. SMMD is a working member of the CIO's SDLC work group. The work group has identified a top down strategy for documenting and implementing SDLC methodologies and tools to be used by OPM information technology customers and professionals. Using Intranet technologies, OPM will provide information on the software development life cycle for managers and employees who plan and implement software system changes. Linked to the documentation will be tools to assist in the various phases of the life cycle, from requirements analysis and cost estimation, through project planning, management and verification testing. The work group is in the process of identifying best practices and tools to implement the SDLC framework.

SMMD is also implementing short term measures to manage, control, and document changes to software code in systems, including the CPDF. Some of the tools are being implemented in support of OPM's Year 2000 project. These code management tools and techniques will be implemented into the SDLC processes as they are introduced throughout the agency. CPDF applications will be included in the code managed through the tools to insure the documentation of changes to applications.

We are also in the process of creating the Quality Assurance, Testing, and Documentation Branch within the SMMD. This new Branch will be responsible for ensuring that all systems, including the CPDF, are fully documented, and that valid test and acceptance procedures are followed for all system changes.

Major Contributors

General Government
Division, Washington,
D.C.

Steven J. Wozny, Assistant Director (202) 512-5767
Domingo Nieves, Evaluator-in-Charge
Jeffrey W. Dawson, Evaluator
Michael J. O'Donnell, Advisor
Gregory H. Wilmoth, Supervisory Social Science Analyst
Stuart M. Kaufman, Senior Social Science Analyst
Kiki Theodoropoulos, Senior Evaluator
George H. Quinn, Jr., Computer Analyst

Accounting and
Information
Management Division,
Washington, D.C.

Brian C. Spencer, Technical Assistant Director

Denver Regional
Office

Joseph J. Buschy, Senior Evaluator

Ordering Information

The first copy of each GAO report and testimony is free. Additional copies are \$2 each. Orders should be sent to the following address, accompanied by a check or money order made out to the Superintendent of Documents, when necessary. VISA and MasterCard credit cards are accepted, also. Orders for 100 or more copies to be mailed to a single address are discounted 25 percent.

Orders by mail:

U.S. General Accounting Office
P.O. Box 37050
Washington, DC 20013

or visit:

Room 1100
700 4th St. NW (corner of 4th and G Sts. NW)
U.S. General Accounting Office
Washington, DC

Orders may also be placed by calling (202) 512-6000 or by using fax number (202) 512-6061, or TDD (202) 512-2537.

Each day, GAO issues a list of newly available reports and testimony. To receive facsimile copies of the daily list or any list from the past 30 days, please call (202) 512-6000 using a touchtone phone. A recorded menu will provide information on how to obtain these lists.

For information on how to access GAO reports on the INTERNET, send an e-mail message with "info" in the body to:

info@www.gao.gov

or visit GAO's World Wide Web Home Page at:

<http://www.gao.gov>

**United States
General Accounting Office
Washington, D.C. 20548-0001**

**Bulk Rate
Postage & Fees Paid
GAO
Permit No. G100**

**Official Business
Penalty for Private Use \$300**

Address Correction Requested

