United States Government Accountability Office

Report to Congressional Addressees

October 2016

FEDERAL HUMAN RESOURCES DATA

OPM Should Improve the Availability and Reliability of Payroll Data to Support Accountability and Workforce Analytics

October 2016

FEDERAL HUMAN RESOURCES DATA

OPM Should Improve the Availability and Reliability of Payroll Data to Support Accountability and Workforce Analytics

Why GAO Did This Study

OPM is tasked with supporting federal agencies' human capital management activities, which includes ensuring that agencies have the data needed to make staffing and resource decisions to support their missions. The EHRI system is OPM's primary data warehouse to support these efforts. The payroll database—one of the four databases in the EHRI system—became operational in 2009. Payroll data provide information on federal employees' pay and benefits and how they allocate their time, as reflected in hours charged to work activities and use of leave. EHRI data are essential to governmentwide human resource management and evaluation of federal employment policies, practices, and costs. The ability to capitalize on this information is dependent, in part, on the reliability of the collected data.

GAO undertook this review to examine the extent to which (1) EHRI payroll data have supported OPM's strategic and open data goals and (2) internal controls are in place to assure the reliability of the data. GAO reviewed literature, interviewed officials and reviewed documents from OPM and the payroll Service Centers, compared OPM's data quality processes to GAO's *Standards for Internal Control*, and performed electronic tests of the payroll data.

What GAO Recommends

GAO is making five recommendations, including that OPM improve the availability of its payroll data and implement additional internal control activities to better ensure data reliability.

OPM agreed with all of GAO's recommendations.

View GAO-17-127. For more information, contact Nancy Kingsbury at (202) 512-2700 or kingsburyn@gao.gov.

What GAO Found

The Enterprise Human Resources Integration (EHRI) payroll data are not fully supporting the Office of Personnel Management's (OPM) strategic and open data goals. This is because OPM has not taken the steps necessary to make the data widely available for use by other agencies and researchers. EHRI payroll data are intended to provide a centralized, standardized, and comprehensive source of pay and leave related data across the federal government. In this capacity, these data have the potential to provide a more efficient, cost effective, and precise data source for federal agencies and researchers who wish to assess human resources and policy decision making across the federal government. Because these data are not widely available, federal agencies and researchers must rely on other proxy sources for payroll data, which are more limited in the scope of analysis they can provide or the level of detail needed for data-driven human capital studies.

Although some elements of the data are sufficiently reliable for general use, weaknesses in OPM's internal controls for the EHRI payroll data will need to be addressed to enhance the reliability of other data elements. As shown in the table below, GAO's assessment of key internal control activities that are critical to ensuring the reliability of the EHRI payroll data found a number of areas where there is insufficient assurance that the control objective will be achieved. These weaknesses increase the risk of data errors, incomplete data fields, and ineffective monitoring of the EHRI payroll data. Unless OPM takes steps to correct these internal control weaknesses, it will be unable to fully leverage these data to meet its mission and allow others to make full use of these data for their research needs.

Assessment of Selected Internal Control Activities for Enterprise Human Resources Integration (EHRI) Payroll Data

Internal Control Attributes	Design	Implementation
Management designs control activities over the information technology infrastructure to support the completeness, accuracy, and validity of information processing by information technology.	○	○
Management designs control activities to limit user access to information technology through authorization control activities.	●	○
Management designs appropriate types of control activities, such as for appropriate documentation of transactions and internal controls.	○	○
Management performs ongoing monitoring of the design and operating effectiveness of the internal control system as part of the normal course of operations.	●	○
Management evaluates and documents the results of ongoing monitoring and separate evaluations to identify internal control issues.	○	○

Source: GAO Analysis of OPM information | GAO 17-127

●= Reasonable assurance control objective will be achieved

○= Insufficient assurance control objective will be achieved

United States Government Accountability Office

Contents

Figure

Abbreviations

ATAAPS	Automated Time, Attendance, and Production System
CSRS	Civil Service Retirement System
DCPS	Defense Civilian Pay System
DFAS	Defense Finance and Accounting Service
DHS	Department of Homeland Security
DOD IG	Department of Defense Inspector General
e-Biz	Business Management Redesign
EHRI	Enterprise Human Resources Integration
ETAMS	Electronic Time and Attendance Management System
FEGLI	Federal Employee Group Life Insurance
FERS	Federal Employees Retirement System
FEVS	Federal Employee Viewpoint Survey
FISCAM	Federal Information System Controls Audit Manual
GDS	Guide to Data Standards Part B
GHRR	Guide to Human Resources Reporting
GS	General Schedule
GSA	General Services Administration
IBC	Interior Business Center
NFC	National Finance Center
NPC	National Payroll Center
OPM	Office of Personnel Management
PP	pay period
SBA	Small Business Administration
SSA IG	Social Security Administration Office of Inspector General
STARWeb	Web-based System for Time and Attendance Reporting
TA	time and attendance
VA	Department of Veterans Affairs
WebTATEL	Web-based Time and Attendance Telecommunications Line
WinTA	Windows Time and Attendance
YTD	year-to-date

GAO

U.S. GOVERNMENT ACCOUNTABILITY OFFICE

441 G St. N.W.
Washington, DC 20548

October 7, 2016

The Honorable Jason Chaffetz
Chairman
The Honorable Elijah E. Cummings
Ranking Member
Committee on Oversight and Government Reform
House of Representatives

Data are essential for understanding the effectiveness and efficiency of government operations. Data can help determine whether federal agencies are accomplishing their missions, how efficiently, and at what cost.[1] Agency officials also rely on data to manage their programs and inform human capital decisions. If these data are unreliable—incomplete or inaccurate—agencies risk making costly and uninformed decisions that can affect both agency operations and the public. Recognizing the importance of available and reliable data to government operations and accountability, we have made several recommendations in prior reports for data quality improvements across a wide range of federal programs.

---

[1] Reliable performance data are necessary to track progress in achieving agencies' missions in compliance with GPRA Modernization Act of 2010 (GPRAMA) (GAO-15-788), reduce fragmentation, overlap, and duplication in federal government programs (GAO-15-404SP), and accurately report spending in compliance with the Digital Accountability and Transparency Act (DATA) Act (GAO-15-241T). GAO, *Managing for Results: Greater Transparency Needed in Public Reporting on the Quality of Performance Information for Selected Agencies' Priority Goals,* GAO-15-788 (Washington D.C.; Sep 2015); GAO, *2015 Annual Report: Additional Opportunities to Reduce Fragmentation, Overlap, and Duplication and Achieve Other Financial Benefits,* GAO-15-404SP (Washington D.C.; Apr 2015). GAO, *Federal Data Transparency: Effective Implementation of the DATA Act Would Help Address Government-wide Management Challenges and Improve Oversight,* GAO-15-241T (Washington D.C.; Dec 2014).

These reports have highlighted the costs and risks of failing to collect accurate data to manage and evaluate agency programs.[2]

One particularly important type of data for ensuring effective and efficient government operations, as well as government accountability, is human resources data, or information on the federal workforce itself. As we have previously noted, agencies must monitor workforce data to avoid serious human capital shortfalls, such as critical skills gaps, which can erode their ability to carry out their missions.[3] In addition, Congress, audit organizations, policy groups, and others regularly use federal workforce data to evaluate federal employment policies, practices, and costs. Audit organizations use workforce data to detect and prevent fraud, waste, and abuse in government operations.

The Office of Personnel Management (OPM) is tasked with supporting agencies' human capital management activities, which includes ensuring that agencies have the data needed to make staffing and resource decisions to support their missions. One of OPM's strategic goals is to be the federal leader in data-driven human resources analytics. To that end, OPM collects a variety of human capital data from federal agencies. The Enterprise Human Resources Integration (EHRI) system is the primary data warehouse to support these efforts. OPM designed EHRI to leverage

---

[2] We have recommended data quality improvements in areas such as hospital quality data, airline safety, federal property management, trends in greenhouse gas emissions, and costs of federal activities such as telework and use of official time, among others. GAO, *Hospital Quality Data: Issues and Challenges Related to How Hospitals Submit Data and How CMS Ensures Data Reliability*, GAO-08-555T (Washington, D.C.: Mar 2008). GAO, *Aviation Safety: FAA Is Taking Steps to Improve Data, but Challenges for Managing Safety Risks Remain,* GAO-12-660T (Washington, D.C.: Apr 2012). GAO, *Federal Real Property: Better Guidance and More Reliable Data Needed to Improve Management*, GAO-14-757T (Washington, D.C.: Jul 2014). GAO, *Climate Change Science: High Quality Greenhouse Gas Emissions Data are a Cornerstone of Programs to Address Climate Change,* GAO-09-423T (Washington, D.C.: Feb 2009). GAO, *Human Capital: Telework Programs Need Clear Goals and Reliable Data* GAO-08-261T (Washington, D.C.: Nov 2007). GAO, *Labor Relations Activities: Actions Needed to Improve Tracking and Reporting of the Use and Cost of Official Time,* GAO-15-9 (Washington, D.C.: Oct 2014).

[3] GAO, *Human Capital: Update on Strategic Management Challenges for the 21st Century* GAO-15-619T (Washington, D.C.: May 2015).

its existing IT infrastructure, staff, and tools to integrate federal personnel, payroll, training, and retirement databases.[4]

The payroll database—one of the four databases of the EHRI system—became operational in 2009. Payroll data provide information on how federal employees allocate their time—reflected in hours charged to specific categories of work and use of leave. Every two weeks, information for over two million federal civilian employees is transmitted to OPM from four payroll service centers and other, individual payroll providers responsible for processing payroll transactions for federal agencies. Service centers standardize and collate certain fields in agencies' time and attendance data in order to report specified fields to the EHRI payroll database. These fields include unique identifiers, such as an EHRI ID number, to link individual records to other personnel data in EHRI, as well as data elements such as agency, time charge categories, and pay rates. As of the last documented update in 2013, the database contained nearly 300 data elements, with roughly 260 fields agencies are required to report and roughly 30 fields that are optional for reporting.

While such data collection and warehousing involves considerable resources, it also provides a range of potential benefits, including improved cost effectiveness, consistent and comprehensive data, and other analytic efficiencies. EHRI data are designed to be centralized, standardized, and comprehensive—essential to governmentwide human resource management and evaluation of federal employment policies, practices, and costs. However, fully capitalizing on these benefits is not possible if the data collected are unreliable. Reliability is central to leveraging the intended efficiencies of EHRI and utilizing the payroll data for resource management and policy decision making. Although OPM has collected these data since 2009, the data's reliability has been in question. While OPM has systematically assessed the reliability of key fields in the personnel database—another component of the EHRI system—no such assessment has been conducted of the payroll data.

---

[4] EHRI was developed following the electronic government (e-Government) initiative, which was aimed at using information technology to enhance the access to and delivery of government information and service to citizens, business partners, and employees, and to improve the internal efficiency and effectiveness of the federal government.

We performed our work under the authority of the Comptroller General to conduct work on federal data systems to assist Congress with its oversight responsibilities. In this report we assess the extent to which (1) EHRI payroll data have supported OPM's strategic and open data goals and (2) internal controls are in place to ensure the reliability of the data.

To address these objectives, we reviewed relevant federal and OPM policies and guidance on personnel data collection, information systems, payroll systems, and time and attendance; interviewed agency officials; reviewed documentation; and analyzed data. To assess the extent to which EHRI payroll data have supported OPM's strategic and open data goals, which involve making human resources data available to support government accountability and workforce analytics, we conducted a literature search to identify (1) studies specifically relying on EHRI payroll data and (2) studies using pay, payroll, and other time and attendance related data from sources other than the EHRI payroll system. We included studies with variables corresponding to fields available in the EHRI payroll database, such as telework or annual leave, to identify potential uses of EHRI payroll data. We limited our search to articles and reports published after 2008 because the EHRI payroll database became operational in 2009. We reviewed literature from peer-reviewed journal articles, agency publications, including GAO reports, and other publications from policy organizations.[5] Lastly, we reviewed OPM documentation on its strategic plans and information technology initiatives, interviewed OPM officials about their actual and potential uses of EHRI payroll data, and conducted searches of OPM and other federal websites that support government data transparency initiatives, such as Data.gov and itdashboard.gov, to determine whether the EHRI payroll database was included in published documentation of existing databases.

To determine the extent to which EHRI payroll system controls ensure the reliability of the data for its intended purpose, we compared OPM's documentation and process for EHRI quality assurance to Standards for

---

[5] We used 22 databases (academic, government, and non-governmental organization literature) to conduct our initial search, using terms such as "EHRI," and "federal pay or payroll." We reviewed these initial 6,966 results for explicit references to federal pay related topics and identified 62 publications for further review. In addition, we searched Inspector General websites for each of the major Departments to identify audit reports that relied on payroll data. Lastly, we reviewed articles that OPM officials had identified as having utilized OPM data.

Internal Control in the Federal Government.[6] According to these standards, internal control systems consist of five components: (1) the control environment, (2) risk assessment, (3) control activities, (4) information and communication, and (5) monitoring.[7] Of these five elements, we determined that control activities and monitoring were the most relevant for ensuring reliable payroll data, and focused on these in our assessment of the EHRI payroll data system. To provide a more detailed methodology for examining the broad controls outlined in Standards for Internal Control in the Federal Government, we relied on specific "look for" indicators for Information System controls outlined in the Federal Information System Controls Audit Manual (FISCAM).[8] The FISCAM contains guidance for reviewing information system controls that affect the confidentiality, reliability, and availability of computerized information, which are aligned with the broader controls outlined in Standards for Internal Control in the Federal Government. We evaluated controls by (1) reviewing OPM's policies, procedures, practices, and standards for EHRI payroll data quality to determine whether guidance was documented and up-to-date; (2) analyzing OPM and service center documentation to determine if the effectiveness of quality controls had been periodically assessed, actions had been taken to correct weaknesses or previously identified control deficiencies, and whether specific edit and access controls were in place; (3) interviewing OPM and service center officials to determine whether OPM was performing certain internal controls processes outlined in its documentation; and (4) testing the reliability of key fields in the EHRI payroll database to identify missing data, obvious errors, and outliers. We did not assess whether an individual's hours charged to a particular time code actually reflected the hours spent on those activities. As each agency is responsible for data

---

[6] Internal controls help agency program managers achieve desired results and provide reasonable assurance that program objectives are being achieved through, among other things, effective and efficient use of agency resources. GAO, *Standards for Internal Control in the Federal Government*, GAO-14-704G (Washington, D.C.: September 2014); effective in fiscal year 2016.

[7] To be effective, all five components of internal control systems must be effectively designed, implemented, and operating in an integrated manner. Certain operating principles within each component help ensure controls are effectively designed, implemented, and operating.

[8] GAO, *Federal Information System Controls Audit Manual* (FISCAM), GAO-09-232G (Washington, D.C.: February 2009).

quality, this type of validity is managed by agencies' time and attendance policies and review processes.

We conducted this performance audit from July 2014 to October 2016 in accordance with generally accepted government auditing standards. Those standards require that we plan and perform the audit to obtain sufficient, appropriate evidence to provide a reasonable basis for our findings and conclusions based on the audit objectives. We believe that the evidence obtained provides a reasonable basis for our findings and conclusions based on our audit objectives.

Background

The Office of Personnel Management (OPM) is tasked with providing human resources, leadership, and support to federal agencies to manage their human capital functions. For OPM to effectively perform this role, executive branch agencies are required to report information on their civilian employees to OPM and ensure that workforce data meet certain standards developed by OPM.[9] OPM has developed these data to carry out its strategic goal of serving as a thought leader in data-driven human resource management and policy decision-making. The Enterprise Human Resources Integration (EHRI) system is OPM's primary repository for human capital data to support these efforts.

OPM developed EHRI to (1) provide for comprehensive knowledge management and workforce analysis, forecasting, and reporting to further strategic management of human capital across the executive branch; (2) facilitate the electronic exchange of standardized human resources data within and across agencies and systems and the associated benefits and cost savings; and (3) provide unification and consistency in human capital data across the executive branch. In addition, OPM's updated system and integrated data were expected to accrue savings to the federal government, reduce redundancy among agency systems, streamline the various processes involved in tracking and managing federal

---

[9] Under 5 U.S.C. § 1104(a)(1), the President may delegate authority for personnel management functions to the Director of OPM. In 2001, by Executive Order 13,197, the President authorized the Director of OPM to require all Executive agencies to report information related to civilian employees. The Director was to develop standards for workforce information submissions and agencies were to ensure submissions met those standards. See 5 C.F.R. § 9.2. OPM published The Guide to Data Standards Part B for Payroll submissions (last updated March 1, 2012).

employment, and facilitate human capital management activities by providing storage, access, and exchange of standard electronic information through a data repository of standardized core human capital data for most executive branch employees.

While the personnel database predated the EHRI Data Warehouse, the payroll database was newly developed for OPM's e-payroll initiative to consolidate agency payroll processes.[10] The payroll database contains individual payroll records for approximately 2.0 million federal employees and is the primary governmentwide source for payroll information on federal employees. The records consist of data elements such as an EHRI ID for linking files, agency time charge categories, and pay rates.

The consolidation of agency payroll processes—known as the e-payroll project—provided the opportunity for OPM to begin collecting standardized governmentwide payroll data.[11] As part of the e-payroll initiative, OPM consolidated the operations of 22 federal payroll system providers for the 116 executive branch agencies into four primary providers—General Services Administration's (GSA) National Payroll Center (NPC), the Department of Defense's Defense Finance and Accounting Service (DFAS), Department of Interior's Interior Business Center (IBC), and Department of Agriculture's National Finance Center (NFC). Consolidation was undertaken to simplify and standardize federal payroll policies and procedures, and better integrate payroll with other human capital and finance functions.

Processing Payroll from Time and Attendance Systems

Most federal agencies rely on one of the four payroll service centers, DFAS, IBC, NPC, or NFC, to process employee pay.[12] Payroll service centers receive employees' bi-weekly time sheets which come from a variety of time and attendance (TA) systems from the agencies they service. Generally, TA systems allow employees to specify time spent on different work and leave categories, such as the number of regular or

---

[10] Prior to EHRI, the personnel database was known as the Central Personnel Data File.

[11] In 2009, OPM completed the initiative that consolidated 22 federal payroll systems into four shared-service centers and standardized payroll policies and procedures, an effort they estimated would save the federal government $1.1 billion over 10 years.

[12] The State Department has its own payroll processing center for its employees.

overtime hours worked or the number of annual leave or sick leave hours taken in a given pay period (PP). However, the level of detail regarding the exact nature of the work or leave time varies depending on agency policies and systems for recording employee work time. While the service centers have consolidated payroll reporting, there are still variations among the centers and the TA systems agencies use to submit employee time sheets. Some systems are maintained by the service center and employees from various agencies access those systems to record their hours. For example, GSA has only one TA system that agencies use to record work and leave hours. Other systems are maintained by the agency and may reflect specific TA accounting needs of the agency. For example, NFC processes time sheets from several different TA systems, some of which are agency specific, and DFAS processes payroll for the Department of Defense, Veterans Affairs, and others through systems including the Business Management Redesign (e-Biz) TA system and the Automated Time, Attendance, and Production System (ATAAPS) (see figure 1).

Figure 1: Flow of Time and Attendance Reporting Through Payroll Service Centers to OPM's Enterprise Human Resources Integration Payroll Database

Department or agency	Time and attendance systems	Payroll service center	Office of Personnel Management (OPM)
Defense Energy Health and Human Services Veterans Affairs 10 Non-Chief Financial Officer (CFO) agencies	e-Biz DCPS ATAAPS Agency unique	DOD's Defense Finance and Accounting Service (DFAS)	
Agriculture Agency for International Development Commerce Justice Labor Homeland Security Housing and Urban Development Small Business Administration Treasury 47 Non-Chief Financial Officer (CFO) agencies	STARWeb WebTA Agency unique	USDA's National Finance Center (NFC)	Enterprise Human Resources Integration (EHRI) Payroll Database
Environmental Protection Agency Education Interior National Science Foundation National Aeronautics and Space Administration Nuclear Regulatory Commission Social Security Administration Transportation 35 Non-Chief Financial Officer (CFO) agencies	Quicktime WebTA Agency unique ETAMS	Interior's Interior Business Center (IBC) GSA's National Payroll Center (NPC)	
General Services Administration Office of Personnel Management 36 Non-Chief Financial Officer (CFO) agencies	WinTA WebTATEL	State (self)	
State	Unknown	Self	
Various Non-Chief Financial Officer (CFO) agencies			

Source: GAO analysis of OPM documentation. | GAO-17-127

Notes: Business Management Redesign (e-Biz)

Defense Civilian Pay System (DCPS)

Automated Time, Attendance, and Production System (ATAAPS)

Web-based System for Time and Attendance Reporting (STARWeb)

Electronic Time and Attendance Management System (ETAMS)

Windows Time and Attendance (WinTA)

Web-based Time and Attendance Telecommunications Line (WebTATEL)

In light of the significant variation in TA systems and service centers involved in processing TA information into payroll records, OPM established core standards for consistency in reporting and recording certain types of work and leave hours. These standards apply at the agency level as well as the service center level. Some of these standards are based on official leave authorized in statute. For example, federal employees are authorized to be absent from duty without a loss in pay or charge to leave for legal holidays and for activities such as jury duty, attendance at a military funeral, bone-marrow or organ donation, and certain union activities.[13] Other standards are based on OPM guidance for excused absences due to inclement weather or blood donation, among others, charged as administrative leave. Agencies follow common recording practices for annual leave and sick leave. These core standards at the agency level are an important part of the process for reporting to OPM because they allow the service centers to collapse certain fields in a consistent way. While agencies may have specific time codes and time keeping practices to meet their needs, core standards for service center reporting dictate how these codes should be collapsed for reporting to OPM. For example, agencies may have detailed categories for various types of administrative leave, but segments of those charge codes apply generally to the category of administrative leave, enabling service centers to aggregate these data from TA systems.

Internal Controls and Data Reliability

OPM relies on agencies and service centers to ensure that the data they submit are timely, accurate, complete, and compiled in accordance with OPM standards. However, federal internal control standards specify that even when external parties, such as service centers in this case, perform operational processes for an agency, management—in this case OPM—retains responsibility for the performance of responsibilities assigned to those organizations. Consequently, OPM management is responsible for

---

[13] See 5 U.S.C. §§ 6103-6104 (legal holiday); 6322 (court leave); 6326 (military funeral); 6327 (bone or organ donation); and 7131 (union activities).

understanding the controls each service center has designed, implemented, and operates for payroll processing and how the service centers' internal control systems impact OPM's internal controls for the payroll data. Underlying requirements for data standards and data quality efforts are the standards for internal control which apply to all executive branch government functions. According to Standards for Internal Control in the Federal Government , effective internal control systems have certain attributes, including reliable internal and external sources that provide data that are reasonably free from error and bias and faithfully represent what they purport to represent.[14] Another attribute involves management evaluating both internal and external sources of data for reliability, and obtaining data on a timely basis so that they can be used for effective monitoring.

OPM Has Not Made EHRI Payroll Data Widely Available Even Though They Have Potential to Support OPM's Strategic and Open Data Goals

Use of the EHRI payroll database has been limited because OPM has not made it widely available. This is in contrast to other, related OPM datasets, such as the EHRI personnel database, which OPM has prepared for use and made publicly available through multiple mechanisms including FedScope, an online tool for data analytics. Because the EHRI payroll database has potential to be used for accountability, research, and data-driven human resource management and policy decision making, making it available would support OPM's strategic and open data goals.

Unlike Other, Related Databases, OPM Has Not Made the EHRI Payroll Data Widely Available

The EHRI payroll data have rarely been used since the database became operational in 2009. We identified four instances where the data have been used, primarily by OPM or by GAO to respond to Congressional requests for information. Specifically, (1) OPM used EHRI payroll data to calculate rough estimates of official time—paid time that employees spend on union-related activities—for its 2009 to 2012 reports to Congress; (2) we made similar use of EHRI payroll data to estimate use of official time in selected federal agencies in a 2014 report, which

---

[14] GAO, *Standards for Internal Control in the Federal Government*, GAO-14-704G (Washington, D.C.: September 2014).

revealed limitations in OPM's method of estimating the governmentwide costs of official time; (3) we used EHRI payroll data in a 2014 review that found inconsistencies in how agencies recorded and reported the use of administrative leave; and (4) we used EHRI payroll data in 2016 to report on the use of administrative leave at the Department of Homeland Security (DHS). In this final case, our use of EHRI data enabled a more detailed examination of DHS's use of administrative leave and helped verify the reliability of the information obtained from DHS.

Aside from these four instances, the data remain largely unused because OPM does not make the data available to the larger research community or to federal agencies. This is in contrast to other OPM data, such as the EHRI personnel data, which has been widely used since OPM made it available. OPM has taken specific steps to make these other human resources-related data available, but has not taken any of these steps for the EHRI payroll database. For example, OPM has made the EHRI personnel database available by (1) cleaning it and preparing it for statistical analysis; (2) integrating it with other data in a repository known as the EHRI Statistical Data Mart; (3) making deidentified data—that is, data without personally identifiable information—accessible through FedScope, an online data analytics tool that draws on data in the Statistical Data Mart;[15] (4) listing data that are available (either by download or by request) on the OPM website and on Data.gov; and (5) sharing requested data with other parties, such as think tanks and academic researchers.

Data in the EHRI Statistical Data Mart are also processed and repackaged to make them more available and usable. Specifically, data received by OPM from agencies and stored in the EHRI Data Warehouse are further processed and cleaned and placed into a format better suited for analysis. This process involves additional corrections and generates additional data elements likely to be useful for statistical analysis. These processed and prepared data are then submitted to the EHRI Statistical Data Mart, which forms the basis for FedScope. Both the FedScope analytics tool and the downloadable datasets are accompanied by

---

[15] Deidentification involves removing personally identifiable information from the data (such as SSN, name, address, and email).

documentation that clarifies the meanings of the data elements and limitations associated with the data.

OPM also makes other EHRI data available through its website and Data.gov. Established in 2009, Data.gov is administered by the General Services Administration (GSA) as a public portal for government data in accordance with the government-wide open data policy. It includes information about and links to datasets from executive branch agencies. As of September 2016, users are able to find references to the EHRI personnel and retirement data on Data.gov. From Data.gov, users can follow links to the personnel data in FedScope and request access to the retirement data.[16] OPM also offers a suite of analytic tools for agencies to perform workforce analyses and forecasting on the data in the EHRI Data Warehouse.

Unlike the EHRI personnel and retirement data, the EHRI payroll data have not been made available in any of these ways. The documentation that accompanies the EHRI Statistical Data Mart specifically notes the absence of payroll data as a limitation, warning users that the data elements related to pay reflect only annualized rates of pay, and that employees' actual pay may be lower or higher due to such factors as overtime or leave without pay, which would be addressed if the payroll data were integrated into the Statistical Data Mart. The four databases within the EHRI Data Warehouse were designed with linking identifiers to enable such integration. Even though this capability exists, for the past seven years, the payroll database has not been linked with other EHRI databases, is not incorporated into the Statistical Data Mart, and has been left largely unchecked and unused. Until the payroll data are made available, such as by incorporating them into the Statistical Data Mart and linking them to the other EHRI databases as designed, OPM will not be able to crosscheck data across the databases for accuracy, and the data will not benefit from the processing and preparation for statistical analysis that is performed for data in the Statistical Data Mart.

Because reliability issues are often identified during use of the data, greater use of the EHRI payroll data by other parties would also have the

---

[16] According to metadata updated in September of 2015, the retirement data are classified on Data.gov as "restricted public" and are available under certain conditions or to certain audiences (such as researchers who sign a waiver).

benefit of helping to improve and establish the data's reliability. As noted in GAO's *Assessing the Reliability of Computer-Processed Data*, past users can be valuable sources of information about the completeness, accuracy, limitations, and usability of a dataset. For example, our prior reports that utilized the EHRI payroll data uncovered reliability issues, and OPM itself discovered a reliability issue when it attempted to use the data to analyze the use of sick leave (an issue we describe later in this report).[17]

EHRI Payroll Data Have Potential to Support Research and Analysis in Areas Related to OPM's Strategic and Open Data Goals

The EHRI payroll data have potential to be used for research and analysis on topics related to OPM's strategic and open data goals. In particular, EHRI payroll data include detailed information on pay, incentives, leave, work activities, telework, and other aspects of the federal workforce that could support OPM's strategic and open data goals for data-driven research in areas such as audits and human resource analytics and decision making, according to our review of literature and interviews with OPM officials. OPM's strategic goals call for the agency to develop and provide access to data systems that support human resources–related research and analytics both within and outside of OPM. As part of its strategic goal to serve as the thought leader in research and data-driven human resource management and policy decision making, OPM's strategies include (1) developing data systems to support such analysis and (2) fostering partnerships with work groups, agencies, universities, and industry to access and analyze data. As part of its strategic goal to manage information technology systems efficiently and effectively, OPM's strategies include providing greater access to human resources data and enabling data analytics to inform policy and decisions. In addition, OPM has open data goals that involve making data available and usable, in part to help ensure governmentwide accountability. In particular, OPM's flagship enterprise information management initiative, a part of its most recent Open Government Plan, includes (1) ensuring that data are easily retrievable and highly usable for analytics and decision making, (2) promoting a culture of collaboration and partnerships with external stakeholders, and (3) releasing data to foster a broader conversation with the public by allowing third parties to

---

[17] GAO-15-9. GAO, *Federal Paid Administrative Leave: Additional Guidance Needed to Improve OPM Data,* GAO-15-79 (Washington, D.C.: Oct 2014).

conduct their own analyses and even create their own applications using OPM data.

EHRI Payroll Data Have the Potential to Support Governmentwide Accountability

The utility of the EHRI payroll data for governmentwide accountability is demonstrated in audits that have been conducted using agency-specific data. For example, the Department of Defense Inspector General (DOD IG) used agency payroll data to conduct an audit of the Defense Finance and Accounting Service (DFAS), DOD's payroll provider. Specifically, by using agency payroll data, the DOD IG was able to identify improper payments to federal civilian employees. Improper payments occur when funds go to the wrong recipient, the recipient receives an incorrect amount of funds, or the recipient uses the funds in an improper manner. DFAS determined that payments were being made to accounts with invalid social security numbers, to employees under the legal employment age, and to multiple employees into the same bank account.[18] These improper payments amounted to more than $15 million over a six-year period. We identified similar audits for improper payments conducted by the Small Business Administration (SBA) and the Social Security Administration Office of Inspector General (SSA IG).

Agency-specific payroll data have also supported audits of the use of retention incentives. For example, in 2011, the Department of Veterans Affairs (VA) Office of Inspector General examined retention incentives paid to VA employees in fiscal year 2010. Using agency payroll data, it found that officials responsible for reviewing and approving retention incentives did not adequately justify and document awards in accordance with VA policy. Also, VA officials did not always terminate retention incentives at the end of set payment periods. As a result of their review, the VA Inspector General questioned the appropriateness of nearly 80 percent of incentives it reviewed. These incentives totaled about $1.06 million in FY 2010.[19]

---

[18] Department of Defense, Inspector General, *Validity of DOD Civilian Employee Accounts*, D-2009-092 (Arlington, VA: July 15, 2009).

[19] Department of Veterans Affairs, Office of Inspector General, *Department of Veterans Affairs: Audit of Retention Incentives for Veterans Health Administration and VA Central Office Employees*, 10-02887-30 (Washington, D.C.: November 14, 2011).

Agency-specific payroll data have also been used to audit agencies' reports on the amounts they have withheld or deducted from employees' pay for retirement, health benefits, and life insurance. For example, the DOD IG, in collaboration with OPM, has checked payroll data against Official Personnel Files to assess whether withholdings from pay appear reasonable for employees in multiple departments whose pay is processed through DFAS. The Department of Agriculture IG has done a similar audit for employees in multiple departments whose pay is processed through the National Finance Center (NFC). Identifying payroll fraud, monitoring retention incentives, and assuring accuracy of withholdings are issues that can affect all agencies. The EHRI payroll database contains data elements necessary for conducting such audits, with the advantage that it includes data for all executive branch agencies, thus enabling governmentwide reviews.

EHRI Payroll Data Have the Potential to Support Human Resource Analytics and Decision Making

Our review of the literature and interviews with OPM officials suggest that key elements in the EHRI payroll data have the potential to be used to understand a variety of human capital outcomes in the federal government. We identified studies that could have benefitted from the availability of the EHRI payroll data to examine (1) the relationship between demographic characteristics and pay disparities and costs of compensation; (2) the use of flexibilities, such as telework, on employee retention and motivation; and (3) the use of various types of leave. Because the EHRI payroll data were unavailable, the studies we identified tended to make use of data that were less precise, less directly relevant, or less comprehensive than the EHRI payroll data. These studies illustrate some of types of research that could be done more precisely or more comprehensively if EHRI payroll data were made available.

According to our review of the literature, EHRI payroll data could also inform studies of disparities in pay among demographic groups in the federal workforce and enable more precise analysis of the costs of federal compensation. For example, a 2015 article in the Journal of Public Administration Research and Theory compared long-term professional mobility between federal employees who received veterans' preferences and those who did not. To measure mobility, the study relied on employees' General Schedule (GS) grade levels from OPM's Personnel

data system.[20] A 2012 study in the Internal Review on Public and Non-Profit Marketing examined the relationship between performance-based pay initiatives and discrimination complaints in selected federal agencies using agency data related to equal employment opportunity records.[21] A 2009 study in the American Review of Public Administration examined potential drivers of the narrowing pay gap between men and women in the federal government, including changes in seniority, differences in fields of study, and women's migration into traditionally male fields. To measure the pay gap, these researchers relied on data on employees' self-reported annual salary.[22]

Although these studies of disparities provide insight into the impact of human capital decisions concerning hiring and pay, they had to rely upon GS grade levels, administrative data, or average annual salary levels to assess outcomes. Each of these measures has limitations that could have been addressed if EHRI payroll data had been available. This is because the EHRI payroll data were designed to provide standardized, governmentwide information by pay period regarding actual pay, incentive pay, telework and leave hours, and numerous other data elements related to federal work activities and compensation. The data used in the three studies, however, did not provide precise measures of compensation because pay ranges overlap grade levels in the GS system and individuals with different grades can receive similar pay rates. In addition, approximately 30 percent of federal employees are not covered by the GS system and would be excluded from such assessments by default. Available measures of annualized salary used in these studies are also imprecise. An employee's actual earnings may include other forms of pay (for example, overtime or shift differentials) not included in adjusted basic pay, or may be less than the annualized rate because of the employee's work schedule (for example, less than full time non-seasonal) or

---

[20] Tim Johnson, "Service after Serving: Does Veterans' Preference Diminish the Quality of the US Federal Service?," *Journal of Public Administration Research and Theory*, vol. 25, no. 3 (2015).

[21] Andrew I. E. Ewoh and Stephen Sonnenfeldt-Goddard, "Pay-for-Performance Reform and Organizational Discrimination: An Exploratory Analysis of the United States Federal Agencies," *International Review on Public and Non-Profit Marketing*, vol. 9, no. 1 (2012).

[22] Gregory B. Lewis and Seong Soo Oh. "A Major Difference? Fields of Study and Male-Female Pay Differences in Federal Employment," *American Review of Public Administration,* vol. 39, no. 2 (2009).

individual circumstances (for example, leave without pay). Also, incorporating administrative records of pay into analyses can be challenging because methods of collecting and reporting otherwise similar payroll information varies significantly across federal agencies.[23] In contrast, if EHRI payroll data had been available for these studies, they could have addressed some of these limitations because the data are designed to reflect actual pay, including any special pay, overtime pay, or other incentives and awards an individual might receive each pay period. They are also centralized and designed to be consistent across agencies.

Studies of the impact of workforce flexibilities, such as telework, on employee retention and motivation also demonstrate potential uses of the EHRI payroll database. A 2010 article in Public Manager described how agencies, such as the U.S. Nuclear Regulatory Commission, have used telework to improve retention of employees with critical skills.[24] Similarly, a 2013 study published in the American Review of Public Administration found that federal employees who engaged in frequent or infrequent telework were no more likely than their counterparts who do not telework to express an intention to leave, while a 2012 study in the same journal concluded that employees at the Department of Health and Human Services who telework were not significantly more motivated than those who choose not to telework.[25] Instead of using time spent teleworking drawn from time and attendance records, both of these studies relied on data from the Federal Employee Viewpoint Survey, in which federal employees self-report whether they engage in telework "frequently" or "infrequently." These studies would have benefitted from EHRI payroll data, which include telework fields that, if reliable, could yield more precise findings by allowing researchers to use the actual number of hours or days employees teleworked per pay period, rather than employees' generalized descriptions of their telework frequency.

---

[23] Ewoh and Sonnenfeldt-Goddard, "Pay-for-Performance Reform and Organizational Discrimination," 75.

[24] Pat Galagan, "Bridging the Skills Gap: Part II," *Public Manager,* vol. 29, no. 2 (2010).

[25] James Caillier, "Are Teleworkers Less Likely to Report Leave Intentions in the United States Federal Government than Non-Teleworkers Are?" *American Review of Public Administration,* vol. 43, no.1 (2013). James Caillier, "The Impact of Teleworking on Work Motivation in a U.S. Federal Government Agency," *American Review of Public Administration* vol. 42, no. 4 (2012).

The telework fields in the EHRI payroll data could also help OPM to meet statutory requirements to monitor and report on governmentwide use of telework, and OPM has recently issued a memo to agencies indicating that it will start using the payroll data to do so.[26] In our 2012 report on OPM's ability to meet this requirement, we found that estimates of telework among federal employees were limited to data calls to agencies because some agencies did not track telework in their time and attendance systems. In that report, we concluded that the accuracy of telework participation and frequency data for some agencies was questionable. The EHRI payroll reporting requirements now include data elements for continuous and episodic telework. The availability and use of these data elements for analysis would allow for more accurate and efficient assessments to meet statutory reporting requirements, and would be of use to policymakers.

Our review of literature also indicates that the EHRI payroll data are potentially useful for analyzing the use of leave. OPM officials told us that they would use the data to analyze use of sick leave and annual leave across the federal government if they had sufficient resources. In the past, officials conducted such analyses by requesting data on an ad hoc basis from agencies, but, according to OPM officials, that process was too resource-intensive to continue. Our review of recent studies suggests that researchers share OPM's interest in these topics. For example, a 2015 study in the Journal of Occupational and Environmental Medicine used time and attendance records from an unidentified federal agency to examine sick leave use among different demographic groups.[27] Using survey data, another study examined the impact of different office designs on sick leave use among Swedish workers, finding that open office plans were associated with significantly higher reported rates of sick leave

---

[26] The Telework Enhancement Act of 2010 requires the Office of Personnel Management (OPM) to submit an annual report to Congress addressing the telework program of each executive agency. Pub. L. No. 111-292, 124 Stat. 3165 (Dec. 9, 2010).

[27] Kim Gajewski, Dara Burris, D. Ross Spears, Kevin Sullivan, Oluremi Oyinloye, Carrie McNeil, Paul Meechan, Eli Warnock, Joanthan Trapp, K.C. Decker, and Sandy Chapman, "Demographic Trends of Sick Leave Absenteeism among Civil Service Employees at a Federal Agency from 2004 to 2012," *Journal of Occupational and Environmental Medicine,* vol. 57, no. 3 (2015). Among other things, the authors found that employees on the Civil Service Retirement System (CSRS) and the Federal Employees Retirement System (FERS) used sick leave at comparable rates, leading the authors to conclude that the latter system does not incentivize employees to save sick leave.

use.[28] However, research on trends in leave use and impacts of certain policies on leave use across the federal government has been limited in the past by a lack of comprehensive and standardized leave use data, which could be addressed if EHRI payroll data were made available to agencies and researchers.

Collectively, these studies demonstrate the value of fields within the EHRI payroll database—such as pay, telework, leave, and other compensation data—in assessing human capital outcomes. However, in all of the studies we identified, researchers relied on proxy, annualized, or self-reported measures, as opposed to actual measures of pay, telework, leave, and other key variables. Further, researchers typically relied upon data covering a limited number of federal agencies or employees. Compared to the EHRI payroll data, these sources do not provide governmentwide data or the same level of precision or detail for assessing policy outcomes among federal employees, which can affect the results of analysis. Studies relying on annualized salary may over or understate actual compensation given the timing of personnel actions, such as hires, separations, promotions, and leave, which can affect the actual amount of pay employees receive in a year. Studies relying on information about a small subset of the federal workforce may not provide reliable insights about overall federal human capital trends or policy effects. We and others have noted the importance of appropriate methods and data in comparing benefits and wages among federal employees and their private-sector counterparts.[29]

Other data sources that have been used instead of the EHRI payroll data—such as OPM's EHRI personnel database and the Federal Employee Viewpoint Survey (FEVS)—also have limitations that could be addressed if the payroll data were made available. For example, the EHRI personnel database does not contain information on the amounts of time spent on sick leave, annual leave, administrative leave, official time activities, and telework, among other variables relevant to compensation

---

[28] Christina Bodin Danielsson, Hlendro Singh Chungham, Cornelia Wulff, and Gugo Westerlund, "Office Design's Impact on Sick Leave Rates," *Ergonomics*, vol. 57, no. 2 (2014).

[29] GAO, *Federal Workers: Results of Studies on Federal Pay Varied Due to Differing Methodologies*, GAO-12-564 (Washington, D.C.: June 22, 2012). Kelley, Collen. 2012. A Balanced Pay System Serves Our Nation. *Public Administration Review* 72 (6): 782.

and time use studies. OPM's FEVS—a governmentwide database of federal employee perceptions on their agency's policies and practices—contains data elements related to pay, but limitations on the reliability of these data elements have been identified.[30]

In addition to the specific studies we reviewed in detail, we identified hundreds of articles on topics that correspond to EHRI payroll data fields.[31] For example, we identified 276 articles with the phrases "administrative leave," "annual leave," "court leave," "family leave," "medical leave," "military leave," or "unpaid leave" in their titles that have been published in peer-reviewed journals since 2009, when OPM launched EHRI. In addition, we identified 37 peer-reviewed studies with the term "telework" in their titles and six with the phrase "performance-based pay" in their titles. Although an in-depth assessment would be necessary to determine the reliability of individual fields for any of the specific purposes noted above, our basic reliability testing suggests that several key fields in the EHRI payroll data are reasonably complete and contain data within expected ranges—and therefore would have potential to support research on these topics if the EHRI payroll data were made available. (See appendix I for more detailed results of our electronic tests of EHRI payroll data reliability.) As long as the EHRI payroll data remain unavailable, federal pay and work-related research will be limited and OPM will continue to miss opportunities to support its strategic and open data goals.

---

[30] Sergio Fernandez, William G. Resh, Tima Moldogaziev, and Zachary W. Oberfield, "Assessing the Past and Promise of the Federal Employee Viewpoint Survey for Public Management Research: A Research Synthesis," *Public Administration Review*, vol. 75, no. 3 (2015).

[31] Keywords in publication titles are considered a sign of the significance of terms to a work. *See* John Whittaker, Jean-Pierre Courtial, and John Law, "Creativity and Conformity in Science: Titles, Keywords, and Co-Word Analysis," *Social Studies of Science* vol.19, no. 3 (1989).

Weaknesses in Internal Controls Complicate OPM's Ability to Leverage EHRI Payroll Data in Support of Its Mission

OPM has designed and implemented some control activities to ensure the reliability of EHRI payroll data, but weaknesses in these controls limit OPM's ability to fully leverage these data in support of its mission. As described earlier, we assessed OPM's internal controls on the payroll data against two of the five elements of the Standards for Internal Control in the Federal Government: control activities and monitoring. Control activities are the actions management establishes through policies and procedures to achieve its objectives, including appropriate documentation of internal controls. Control activities help agencies ensure the reliability of data within information systems, such as the EHRI payroll system. Monitoring is necessary to promptly resolve the findings of audits and other reviews so that corrective actions necessary to achieve objectives are taken in a timely manner. A deficiency exists when the design, implementation, or operation of a control does not allow management or personnel to achieve control objectives or address related risks.

Internal Control Weaknesses Increase the Risk of EHRI Payroll Data Reliability Issues

While OPM internal controls provide some assurance of the reliability of EHRI payroll data, weaknesses in the design or implementation of certain control activities and monitoring controls for the EHRI payroll database increase the risk of reliability issues that may limit OPM's ability to fully leverage the data in support of its mission. Specifically, (1) weaknesses in control activities have resulted in limited quality checks and acceptance of unreliable data into the EHRI payroll database; and (2) weaknesses in monitoring activities have resulted in failure to address these reliability issues and increased risk that these issues may compound over time. Table 1 lists these control components and related activities, along with an assessment of whether they provide reasonable assurance of OPM's ability to achieve its objectives in these areas.[32]

---

[32] To assess OPM's control activities, including documentation and procedures for assuring the reliability of EHRI payroll data, we relied on specific "look for" indicators of information system controls outlined in the *Federal Information System Controls Audit Manual* (FISCAM). The FISCAM contains guidance for reviewing information system controls that affect the confidentiality, reliability, and availability of computerized information, which are aligned with the broader controls outlined in *Standards for Internal Control in the Federal Government.*

Table 1: Assessment of Selected Internal Controls for Enterprise Human Resources Integration Payroll Data

Internal control component	Internal control attributes	Indicator of control activity	Design	Implementation
Control activities	Management designs control activities over the information technology infrastructure to support the completeness, accuracy, and validity of information processing by information technology.	Parameters and tolerances are configured and error conditions and messages are defined, and data input errors are identified in error reports.	○	○
	Management designs control activities to limit user access to information technology through authorization control activities.	Roles and responsibilities of users are clearly documented, including written procedures for access and use of the database.	●	○
	Management designs appropriate types of control activities, such as for appropriate documentation of transactions and internal controls.	All documentation and records should be properly managed, updated, and maintained.	○	○
Monitoring	Management performs ongoing monitoring of the design and operating effectiveness of the internal control system as part of the normal course of operations.	System generated reports are reviewed and resolved or resubmitted in a timely manner to reasonably assure the integrity of production data and transaction processing.	●	○
	Management evaluates and documents the results of ongoing monitoring and separate evaluations to identify internal control issues.	Periodic reviews should be made to help reduce the risk of errors, fraud, misuse, or unauthorized alteration.	○	○

Source: GAO Analysis of OPM information | GAO-17-127

●= Reasonable assurance control objective will be achieved

○= Insufficient assurance control objective will be achieved

Control Activities: Controls for Completeness, Accuracy, and Validity of Information Processing

OPM guidance includes requirements for automated controls in support of data quality, such as defining data parameters and tolerances, identifying data errors, checking for completeness, and taking corrective actions when necessary. According to EHRI documentation and OPM officials, automated edit checks are performed by the data system software to check the validity of individual data elements, the proper relationship of values among associated data elements, and data format specifications. The rules check the value and format of fields, including record identifying fields, such as birthdate and agency, as well as non-record identifying fields, such as hours of leave. Specifically, they check to make sure that fields are formatted as numbers, dates, or text, depending on the designed content of the field, and that all values in a field fall within a defined range of possible values. Further, OPM applies three relational edits to ensure (1) that actions taken to add an employee to the system are not associated with an employee already in the system, (2) that

actions taken to correct a record are associated with an existing record, and (3) that actions taken to delete a record are associated with an existing record. See table 2 for a description of the fields that are checked, the rule that is applied, and the action taken if the rule identifies an error.

Table 2: Payroll Data Edit Rules: Excerpts from the Office of Personnel Management Guide to Human Resources Reporting

Field name	Edit rule	Action taken if edit rule fails
Social security number	None	N/A
Birth date	Check if valid date	Reject record
Agency Code	Check if code value is valid	Reject record
NUMBER(n)	Check if value is numeric	Report error; If value is null or not a number, store zero
DATE	Check if valid date	Report error; If date is null, store "No Data Reported"; If date is invalid, store "Invalid Data"

Source: OPM's Guide to Human Resources Reporting | GAO-17-127

Data that fail OPM's automated checks are considered errors, and the edit rules for the EHRI payroll system specify that data with error rates greater than 3 percent will not be accepted. However, according to OPM officials, the payroll data enters the EHRI Data Warehouse with very few other edits. OPM officials noted that they intend to define additional edits that could be applied to the payroll data during the data loading process. Federal standards for internal control state that management should design control activities to achieve objectives and respond to risks. However, due to the limited nature of these edits EHRI payroll data have a higher risk of data reliability issues that may limit OPM's ability to fully leverage the data in support of its mission. For example, the results of our electronic testing of data from 2010-2015 found fields with missing data, logical errors, and out-of-range values. (For selected results of electronic testing of the data, see appendix I.)

OPM's automated rules also require the system to check the number of records for each agency every pay period, and are designed to reject submissions and generate an automated report for the service centers when the number of records has changed by more than 5 percent. While

the reports are generated, OPM officials told us that resource constraints preclude them from having the same level of controls in place for the payroll data as they do for other EHRI data, including lack of capacity to follow up on missing payroll submissions. Federal standards for internal control state that management should evaluate information processing objectives to meet the defined information requirements, such as for completeness and accuracy. However, the EHRI payroll system accepted multiple submissions of data that should have been rejected by this rule. Specifically, our testing of the data found that, for nine separate pay periods in fiscal year 2014, payroll data records for agencies contained less than 1 percent of the affected agency's total civilian workforce. In all, 17 of the 24 CFO Act agencies were affected by this problem at least once in fiscal year 2014 (see table 3). Without these data, government-wide analytics that cover any of these impacted dates will be similarly limited and incomplete.

Table 3: Number of Pay Periods with Missing Data for the 24 CFO Act Agencies, Fiscal Year 2014

Agency	Number of pay periods affected
Agriculture	2
U.S. Agency for International Development	2
Commerce	2
Defense	0
Education	1
Energy	1
Environmental Protection Agency	0
General Services Administration	0
Health and Human Services	0
Homeland Security	2
Housing and Urban Development	2
Interior	1
Justice	2
Labor	2
National Aeronautics and Space Administration	1
Nuclear Regulatory Commission	1
National Science Foundation	1
Office of Personnel Management	0
Small Business Administration	2
Social Security Administration	1

Agency	Number of pay periods affected
State	0
Transportation	1
Treasury	2
Veterans Affairs	0

Source: GAO analysis of OPM data | GAO-17-127

Although these submissions should have been flagged and rejected by OPM's edit check for having a greater than 5 percent change in the number of records from one pay period to the next, OPM was unaware of the missing data until we identified the problem. This was due, in part, to inadequate monitoring controls, which are described in more detail below. In addition, while OPM designed these control activities to meet requirements for completeness and accuracy, the controls have not always met their objective and therefore do not provide sufficient assurance that completeness requirements will be achieved. Failing to evaluate information processing objectives to see if they meet the defined information requirements for completeness increases the risk that some EHRI payroll data will be unreliable.

Control Activities: Controls on User Access

OPM has established roles and responsibilities for users and other controls safeguarding accountability for data quality and security, and maintains information on data access and use activity. As designed, these user controls are intended to provide reasonable assurance that control objectives will be achieved if OPM monitors them. According to OPM officials, EHRI payroll database users must complete an application to gain access and service provider points of contact are given access credentials once access forms are submitted and approved. Applications of users requiring administrative privileges on information system accounts receive additional scrutiny. In addition, OPM documentation establishes processes for updating the list of authorized users when accounts are created, deactivated, or deleted—for example, specifying that account passwords are to expire after 60 days and that accounts that are inactive for 60 days are to be deactivated.

OPM has also designed controls to capture and save some metadata tied to data loading and data provider submissions and user access. Automated processes capture metadata on user access and those logs are stored in system audit tables which are archived on a monthly basis and retained indefinitely, according to OPM officials. Logs detailing access for the three most recent months are available online. Within the

data warehouse program where payroll data reside, applications have auditing functionality for user activity which captures what the user did as well as what was accessed. Web application activity is also tracked, and logs are retained indefinitely. Reports issued to providers can be reconstituted in real time and this information can be used for investigation. However, OPM officials told us they have not used this information for such investigations or reviews. As a result of this incomplete implementation of access controls, OPM does not know whether these controls are working appropriately.

Control Activities: Appropriate Documentation of Transactions and Internal Controls

The two primary sources of documentation that guide submissions of EHRI payroll data into the system—the Guide to Human Resources Reporting (GHRR), and the Guide to Data Standards Part B (GDS)—are not up-to-date and do not provide sufficient assurance that control objectives will be achieved. OPM relies on the service centers and agencies to assure the accuracy of payroll data submissions, as outlined in these documents. For example, the GHRR outlines each data element, its required format, and whether it must be included in the database. The GDS outlines the required format for submissions to EHRI, including the file content, notification of transmission to OPM, file naming conventions, and transmission frequency. The GDS also acknowledges that the edits outlined for service centers and agencies in that document constitute the minimum required level of quality control and encourages agencies to supplement them based on the specifics of their internal programs and operations. The GHRR was last updated July 2013 to reflect the inclusion of telework variables, but the GDS has not been updated since March 2012. Given the changes to the system and the control weaknesses noted above, OPM officials noted that all payroll data standards will have to be reworked to ensure they are robust for data collection and programming by the payroll providers. OPM officials did note their intention to update these guides to align with system, regulatory, and other changes, but did not have a detailed plan or timeframe for doing so. Federal standards for internal control state that management should document internal controls to ensure that all transactions, documentation, and records are properly managed and maintained. OPM cannot ensure that the data quality control changes made to the system are fully implemented without updating its guidance documents. Out-of-date documentation does not provide sufficient assurance that control objectives will be achieved. Until OPM updates this documentation, it faces increased risk that data submissions will not be consistent with current requirements and recent changes to the system, which could affect the reliability of data submissions.

Monitoring: Ongoing Monitoring During Normal Operations

As described above, the EHRI payroll system is designed to reject data and produce data quality reports when data error rates exceed 3 percent or when the number of records at an agency changes by more than 5 percent from one pay period to the next. According to OPM officials, biweekly EHRI data quality control reports and error files are made available to payroll providers on the EHRI portal. This quality control reporting is kept in the EHRI Data Warehouse indefinitely and the quality control reports issued to providers can be reconstructed. The GHRR directs payroll providers to monitor these reports for deviations from previous norms and analyze them to identify potential issues in systems that gather and send EHRI data from the agency to OPM. OPM officials told us that they do not monitor these reports to identify and resolve problems, and resource constraints prevent the agency from following up with the payroll providers. While inconsistent implementation of control activities allowed incomplete data to be accepted by the EHRI system, this limitation in monitoring controls led these incomplete submissions to remain undetected and unaddressed by OPM. In addition, without timely review and correction of problems identified in these reports, OPM risks errors compounding with each biweekly data submission, as the error tolerance checks involve comparison of each new submission to the most recent submission, which itself may have been incomplete. Further, without timely identification and correction of such problems, missing data may not be recoverable. For example, in response to the missing data issue noted above, OPM contacted the relevant service centers to locate the missing files. However, service centers only retain data for 18 months from the original date of submission. If the controls were working as designed, the service center would have been required to provide a corrected submission before the end of this retention period, and OPM would have reasonable assurance that the data for these pay periods were complete. Because of the delay in identifying this error, when OPM finally did request the data from the service centers, corrected data submissions were expected to require a significant amount of work because the retention period had passed. OPM was unable to provide the data within the time frames of this engagement and it is unclear whether OPM will be able to retrieve the missing data from the relevant service centers. Federal standards for internal control state that management should establish monitoring activities for the internal control system and evaluate the results, and should remediate identified internal control deficiencies on a timely basis. Without appropriate efforts to review and respond to system generated reports, OPM does not have sufficient assurance that the control objective will be achieved and the risk of submissions of inaccurate or incomplete data is increased.

Data Reliability Issues Limit OPM's Ability to Fully Leverage EHRI Payroll Data in Support of Its Mission

While OPM's internal controls provide some assurance of the reliability of some of the EHRI payroll data, the weaknesses in control activities (controls for completeness, accuracy, and validity of information processing and appropriate documentation) and monitoring controls (ongoing during normal operations) may increase the risk for data reliability issues to arise and persist in the EHRI payroll data. We have also identified several data reliability issues through electronic testing of EHRI payroll data, in past GAO work, and through interviews with OPM officials. Collectively, these issues present challenges for fully leveraging the EHRI data, and may limit OPM's ability to utilize the data for some analyses in support of its mission and strategic goals.

We found a variety of potential data reliability issues from our electronic testing of EHRI payroll data, as illustrated in appendix I. In some cases, these issues indicate the potential for reliability issues that may limit OPM's ability to fully leverage the data in support of its mission. For example, we found that the EHRI payroll data includes records for six entities that should not be in the system due to exemptions from OPM reporting requirements, as shown in table 4. When using EHRI payroll data, the unintentional inclusion of these entities could impact some analyses and limit OPM's ability to draw valid conclusions from the data.

Table 4: Entities Excepted from Enterprise Human Resources Integration (EHRI) Reporting that are Included in EHRI Payroll Data, by Fiscal Year

	2011	2012	2013	2014	2015[a]
Legislative Branch	✓	✓	✓	✓	✓
Executive Office of the President			✓		
Commission on Security and Cooperation in Europe	✓	✓	✓	✓	✓
Commission on the People's Republic of China	✓	✓	✓	✓	✓
John C. Stennis Center for Public Service Training and Development	✓	✓	✓	✓	✓
U.S. Court of Appeals for Veterans Claims	✓	✓	✓	✓	✓

Source: GAO analysis of OPM data | GAO-17-127

[a]Partial year data for FY 2015

We also found a small number of instances of social security numbers being assigned to multiple EHRI records, as shown in table 5 below.

When using EHRI Payroll data, this could indicate that some individuals appear in the data more than once, potentially impacting some analyses and limiting OPM's ability to draw valid conclusions from the data.

Table 5: Social Security Numbers Associated with More than One Enterprise Human Resources Integration Records, by Fiscal Year

	2011	2012	2013	2014	2015[a]
Percent	.01	.01	.01	.01	.01
Number	356	280	236	231	126

Source: GAO analysis of OPM data | GAO-17-127

[a]Partial year data for FY 2015

In a 2014 report, we found that weakness in OPM's documentation for transactions and internal controls led to inconsistent reporting of administrative leave data and inclusion of some excepted agencies' data in payroll feeds. Specifically, in our report of agencies' use of administrative leave we found differences between agencies' leave recording practices and what OPM officials consider paid administrative leave.[33] In response, OPM issued guidance to agencies to review how they record administrative leave and clarify that administrative leave should not be routinely used for an extended time. This guidance can help agencies and payroll providers to provide more consistent data on administrative leave, and improve the usefulness of EHRI payroll data for related analyses.

OPM officials also told us about data reliability issues beyond those identified in this review. For example, OPM officials told us that, in 2015, they discovered a problem related to data on sick leave. Specifically, due to a programming error, the data received from payroll providers that sum the number of sick leave hours an employee used in a year was populating an unrelated field in the EHRI payroll database. As a result, according to OPM officials, the amount of sick leave an employee used in any given year was not accurate. One of these officials told us that this problem may also apply to other variables. This suggests that OPM's edit checks, which were designed to maintain a minimum level of quality control, may not sufficiently reduce the risk of these types of errors.

---

[33] GAO-15-79

OPM officials told us they had plans to update EHRI security protocols, payroll documentation, testing for reliability issues, and data standards, but OPM has not documented these plans or created a schedule to implement them. For example, OPM officials told us that they planned to update their documentation beginning in FY 2016, working collaboratively with OPM's program policy office, federal agencies, and shared service centers, and other stakeholders. Although the EHRI payroll data was not a part of the 2015 OPM data breach, agency officials told us that the agency is evaluating its current security posture and making necessary changes to protect the privacy and integrity of all the data they manage. According to OPM officials, these plans include preparing to deploy a new secure portal to applications and tools; improving use of encryption; masking and redaction when appropriate and prudent; consolidating data from multiple data sets into more secure databases; utilizing better and more secure user management tools and audit trail logging; and providing new forms of user authentication, among other potential security and access measures.

OPM officials also said they are planning to correct the issue they had identified with the sick leave variable and that they were in the process of testing other variables to see if they had the same problem. However, OPM officials told us that these actions would require resources and reprioritization of the existing workload, and that a project plan and timeframes had not yet been developed. Further, OPM officials noted that the agency has a critical leadership role in addressing the complete data life cycle, and that agencies and service centers also play a critical role in assuring data quality. Accordingly, OPM officials said they were seeking a comprehensive solution that includes agency and service center actions to ensure accurate data are submitted to EHRI.

As yet, OPM has not linked EHRI payroll correction activities back to specific agency objectives or created a schedule for implementing these changes. GAO's schedule assessment guide notes that a well-planned schedule is a fundamental management tool that can help government agencies gauge progress, identify and resolve potential problems, and determine the amount and timing of resource needs.[34] Without a well-planned schedule—developed in consideration of how it will contribute to

---

[34] GAO-16-89G

OPM's objectives and risks—OPM may not be able to appropriately prioritize and execute the necessary changes.

Conclusions

Although use of EHRI payroll data has been limited to date, it carries significant potential to support governmentwide accountability and human resource analytics and decision making. In our review of peer-reviewed journals we identified hundreds of articles on topics that correspond to EHRI payroll data fields. Unfortunately, however, EHRI payroll data will continue to be underutilized until—consistent with its own strategic and open data goals—OPM makes the data available to potential users, as it does other databases within the EHRI system. While data collection and storage is not without cost, EHRI's centralized, standardized, and comprehensive features offer the promise of efficient, cost effective, and more precise analytics. In preparing the data to make them available, OPM will need to take steps to process and clean them as it does for the EHRI personnel data. This is the first step toward improved reliability. Our basic reliability testing suggests that several key fields in the EHRI payroll data are reasonably complete and contain data within expected ranges—and therefore could have potential to support research on these topics.

However, while some fields in the current EHRI payroll data may be sufficiently reliable for certain types of audits and workforce analytics, other fields suffer from reliability issues that limit the range of purposes for which the data can be used. This is because OPM has not designed sufficient control activities to assure data quality, has not evaluated or consistently implemented the control activities it has designed, and has not updated key documentation to support quality submissions of data. Compounding these problems is OPM's failure to monitor ongoing operations, for example, by reviewing system generated reports. Without timely identification and correction, data quality problems will continue undetected and remain uncorrected. While OPM officials noted their intention to address these shortcomings, they do not have plans with specific actions and time frames for doing so. Without a schedule specifying when these planned changes will be made, OPM officials will be unable to gauge progress, identify and resolve potential problems, or determine the amount and timing of resource needs related to the desired changes. As a result, OPM faces an increased risk of implementing ineffective or contradictory changes, and of facing delays in completing these activities. Until relevant changes are made, existing problems can continue to compound as data for 2 million federal civilian employees are received biweekly. Without available and reliable payroll data, OPM and others must continue to rely on data that are more costly, imprecise, or

limited in scope—missing opportunities to leverage centralized, standardized data that is essential for accountability and well-informed management and policy decisions.

Recommendations for Executive Action

GAO is making five recommendations to the Director of OPM.

GAO recommends that the Director of OPM take the following action to support its strategic and open data goals:

- Improve the availability of the EHRI payroll data—for example, by preparing the data for analytics, making them available through online tools such as FedScope, and including them among the EHRI data sources on the OPM website and Data.gov.

GAO recommends that the Director of OPM take the following two actions to improve internal controls for data quality:

- Update EHRI payroll database documentation to be consistent with current field definitions and requirements, including the Guide to Human Resources Reporting and the Guide to Data Standards, Part B; and

- Consistently monitor system-generated error and edit check reports and ensure that timely action is taken to address identified issues.

GAO recommends that the Director of OPM take the following two actions to integrate the payroll data into the larger suite of EHRI databases:

- Develop a schedule for executing these plans; and

- Evaluate existing internal control activities and develop new control activities for EHRI payroll data, such as implementing transactional edit checks that leverage the information in the other EHRI datasets.

Agency Comments and Our Evaluation

We provided a draft of this report for review and comment to the Director of OPM. OPM agreed with our recommendations. In its comments (reproduced in appendix III), OPM noted that a lasting and effective solution for enhancing the quality of payroll data requires consistent data quality not just in the "last mile" after delivery to the EHRI system, but also at the origination of the data. OPM also noted that implementation of these recommendations will require collaboration between various stakeholders and appropriate resources. We agree. As we note in the report, while payroll processing is more consolidated than in the past, agencies still use a variety of time and attendance (TA) systems, which

can vary in the level of detail with which work or leave time is recorded depending on agency policies and systems. In addition, there are variations in the systems and processes of the payroll providers. These variations across agencies and across payroll providers underscore the importance of updated documentation for reporting and consistent monitoring of error reports. In addition, through its leadership role in the OPM-managed Human Resources Line of Business, OPM can consider action for ensuring data quality—for example, by including data quality indicators among its performance measures for the payroll providers. OPM also provided technical comments which we incorporated, as appropriate.

We will send copies to the appropriate congressional addressees and the Director of the U.S. Office of Personnel Management, as well as to other interested parties. In addition, the report is available at no charge on the GAO website at http://www.gao.gov. If you or your staff have any questions about this report, please contact me at (202) 512-2700 or KingsburyN@gao.gov. Contact points for our Offices of Congressional Relations and Public Affairs may be found on the last page of this statement. The names of GAO staff who made key contributions to this report are listed in appendix IV.

Nancy Kingsbury
Managing Director, Applied Research and Methods

Appendix I: Selected Data Reliability Testing Results

This Appendix presents selected results from our electronic testing of the EHRI payroll data. The results are grouped into three categories: (1) tests for missing data, (2) tests for logical errors, and (3) tests for potentially invalid values.

1. Missing Data Tests

Incomplete data can limit both the ability to conduct desired analyses and the usefulness of any analysis conducted. For example, if a large amount of data is missing, as was the case with data missing for entire agencies for some pay periods, it may not be possible to complete analysis of that agency for the missing periods. If a smaller proportion of data is missing, analysis may still be possible. However, any analysis completed using these data will be limited in its accuracy and validity, which may increase the risk of drawing inappropriate or invalid conclusions.

Because the EHRI payroll data have been identified as potentially useful for government-wide studies of telework behavior, we tested for missing data in telework-related fields. As shown in table 6, we found data in these fields to be missing entirely in 2011 and largely incomplete in 2012, years for which reporting on this variable was not required by OPM. The percentage missing is based on the number of records without values out of all records within a fiscal year. As a result, estimates of governmentwide telework participation, as shown in table 7, are likely to be inaccurate for these years.

Table 6: Percent and Number of Missing Records for Telework Related Fields in the EHRI Payroll Data

	2011	2012	2013	2014	2015[a]
Routine telework hours in pay period	100 (55,704,002)	85.56 (45,430,244)	0	0	0
Routine telework instances in pay period	100 (55,704,002)	85.56 (45,430,244)	0	0	0
Situational telework hours in pay period	100 (55,704,002)	85.56 (45,430,244)	0	0	0
Situational telework instances in pay period	100 (55,704,002)	85.56 (45,430,244)	0	0	0

Source: GAO analysis of OPM data | GAO-17-127

[a]Partial year data for FY 2015

Table 7: Percent and Number of Records Indicating Use of Telework in the EHRI Payroll Data

	2011	2012	2013	2014	2015[a]
Routine telework hours in pay period	0	0.04 (21,246)	0.71 (391,177)	1.23 (641,117)	1.58 (274,036)
Routine telework instances in pay period	0	0.04 (21,246)	0.71 (391,801)	1.23 (641,134)	1.58 (274,040)
Situational telework hours in pay period	0	0.02 (9,918)	0.33 (180,721)	0.48 (252,342)	0.61 (105,975)
Situational telework instances in pay period	0	0.02 (9,918)	0.33 (181,523)	0.48 (252,365)	0.61 (105,982)

Source: GAO analysis of OPM data | GAO-17-127

[a]Partial year data for FY 2015

2. Logical Error Tests

Logical testing can reveal data reliability issues among and within individual records. For example, logical testing can assess whether there are duplicates among records in the data system. Our electronic testing assessed whether there were duplicate records in the EHRI payroll data. We also looked for records with multiple payments, either from the same agency or from different agencies, in a single pay period—a form of duplication that could also indicate problems with data reliability. As shown in tables 8, 9, 10, and 11, we found no instances of a complete duplicate records, few instances of EHRI IDs associated with more than one social security number, and that generally less than 1 percent of records were associated with multiple payments in a single pay period.

Table 8: Complete Duplicate Records (All Fields the Same) in the EHRI Payroll Data

	2011	2012	2013	2014	2015[a]
Percent	0	0	0	0	0
Number	0	0	0	0	0

Source: GAO analysis of OPM data | GAO-17-127

[a]Partial year data for FY 2015

Table 9: EHRI IDs Associated With More Than One Social Security Number in the EHRI Payroll Data

	2011	2012	2013	2014	2015[a]
Percent	0.01	0.01	0.01	0.01	0.01
Number	356	280	236	231	126

Source: GAO analysis of OPM data | GAO-17-127

[a]Partial year data for FY 2015

Table 10: Records with Multiple Payments Per Pay Period (Any Agency) in the EHRI Payroll Data

	2011	2012	2013	2014	2015[a]
Percent	0.51	0.48	0.48	1.19	0.51
Number	281,343	256,045	264,266	623,855	89,020

Source: GAO analysis of OPM data | GAO-17-127

[a]Partial year data for 2015

Table 11: Records with Multiple Payments Per Pay Period Within the Same Agency in the EHRI Payroll Data

	2011	2012	2013	2014	2015[a]
Percent	0.13	0.14	0.14	0.84	0.10
Number	73,223	71,847	76,534	441,178	17,312

Source: GAO analysis of OPM data | GAO-17-127

[a]Partial year data for FY 2015

Logical testing can also uncover data reliability issues within individual records. As shown in table 12, we found a number of variables with questionable values, given the values of other variables for the same record. For example, we found instances where the amount of annual leave used was greater than the amount available, and other instances where the data indicate an agency contribution to Federal Employee Group Life Insurance (FEGLI) for an employee who has not made a contribution. In both cases, this should not be possible under typical circumstances, and may indicate a data reliability issue.

Table 12: Percent and Number of Selected Variables with Questionable Values Given Acceptable Ranges and Tolerances in the EHRI Payroll Data

	2011	2012	2013	2014	2015[a]
Annual leave hours used > Annual leave balance hours	3.52 (1,960,306)	3.51 (1,862,361)	3.4 (1,863,464)	3.47 (1,814,034)	3.49 (604,351)
Sick leave hours used > Sick leave balance hours	4.88 (2,720,821)	5.05 (2,682,368)	5.08 (2,787,612)	5.02 (2,629,143)	5.02 (869,716)
No data reported for Federal Tax Filing Status	11.6 (6,478,042)	11.55 (6,134,757)	11.44 (6,274,615)	11.57 (6,054,692)	12.98 (2,250,379)
Retention Incentive amount >= Adjusted Basic Pay amount	3.85 (2,144,171)	3.73 (1,979,512)	3.91 (2,142,584)	4.22 (2,209,366)	3.62 (627,317)
Relocation Incentive amount >= Adjusted Basic Pay amount	3.86 (2,148,971)	3.73 (1,982,202)	3.91 (2,145,520)	4.23 (2,212,497)	3.63 (628,497)
Recruitment Incentive amount >= Adjusted Basic Pay amount	3.86 (2,151,457)	3.73 (1,982,990)	3.91 (2,146,046)	4.23 (2,213,236)	3.63 (629,077)
Employee FEGLI contribution = 0 and Agency FEGLI contribution is not =0	0.25 (137,943)	0.24 (125,352)	0.22 (121,853)	0.22 (112,939)	0.17 (29,443)
Total Earnings is < Salary Total within a pay period	0.43 (239,145)	0.43 (228,188)	0.43 (236,955)	0.45 (234,545)	0.42 (72,988)

Source: GAO analysis of OPM data | GAO-17-127

[a]Partial year data for FY 2015

3. Outlier and Out of Range Tests

Tests for invalid formats and values can reveal obvious errors about data. For example, as shown in table 13, we tested the format of the social security numbers (SSN) in EHRI, which should all be nine digit numbers, and found some cases where these numbers were not properly formatted, indicating a potential data reliability issue that could prevent analysis of individuals when attempting to match SSNs.

Table 13: Social Security Numbers with Invalid Formats in the EHRI Payroll Data

	2011	2012	2013	2014	2015[a]
Number	21,738	15,770	3,358	0	0
Percent	0.04	0.03	0.01	0	0

Source: GAO analysis of OPM data | GAO-17-127

[a]Partial year data for FY 2015

We also tested the values of fields in the EHRI payroll data to determine whether any records were outside of the expected ranges. As shown in table 14, we found several variables where the minimum value was below the expected possible floor or the maximum value was above the expected ceiling. For example we found a maximum value for Student Loan Repayments of $100,000, which is well above the expected ceiling for such payments. We also found that the lowest minimum value for total salary in a pay period was negative $99,140, while all salary values should generally be positive for normal records (non "correction" records).

Table 14: Minimum and Maximum Values Outside of Expected Ranges in the EHRI Payroll Data

	2011		2012		2013		2014	
	Minimum	Maximum	Minimum	Maximum	Minimum	Maximum	Minimum	Maximum
Salary total amount (amt) pay period (pp)	(99,140)	340,236	(50,000)	382,582	(64,149)	424,423	(48,986)	359,764
Salary total year-to-date (ytd) amt	(76,482)	791,328	(16,556)	827,212	(84,325)	620,541	(22,578)	503,006
Locality pay amt pp	(88,716)	124,898	(34,663)	146,507	(51,220)	111,249	(148,761)	115,282
Locality pay amt ytd	(88,716)	485,759	(27,794)	473,814	(27,794)	449,448	(28,358)	448,490
Special rate sppl amt pp	(3,223)	15,674	(373)	9,616	(1,348)	11,159	(2,263)	11,498
Special rate sppl amt ytd	(2,885)	42,986	(2,885)	44,703	-	44,346	-	44,640
Total earnings pp	**(213,550)**	**458,106**	**(337,592)**	**927,085**	**(243,471)**	**424,423**	**(319,969)**	**610,737**
Total earnings ytd amt	**(126,416)**	**1,021,795**	**(128,615)**	**990,491**	**(180,046)**	**1,014,664**	**(375,585)**	**615,779**
Award amt	(19,829)	100,000	(10,333)	119,925	(6,075)	155,646	(7,609)	220,953
Award hours earned	(1)	744	-	840	-	960	-	1,000
Award hours used	(3)	516	(27)	528	(27)	514	(1)	514
Award hours ytd bal	(96)	6,064	(92)	6,088	(96)	6,088	(90)	6,088
Recruitment incentive amt	(136,771)	136,771	(30,000)	108,289	(15,964)	80,000	(11,518)	135,235
Relocation incentive amt	(80,000)	175,596	(31,507)	97,212	(101,288)	101,288	(38,160)	130,155
Retention incentive amt	(65,484)	74,567	(128,484)	71,304	(87,257)	65,459	(74,302)	33,878
Student loan repay pp amt	(6,335)	100,000	(10,000)	10,000	(10,000)	20,000	(10,000)	16,886

	2011		2012		2013		2014	
	Minimum	**Maximum**	**Minimum**	**Maximum**	**Minimum**	**Maximum**	**Minimum**	**Maximum**
Overtime hours worked	(404)	11,931	(248)	18,825	(157)	13,046	(491)	8,912
Overtime pay amt	(22,927)	89,590	(23,924)	91,050	(37,312)	104,142	(50,328)	134,879
Annual leave (AL) balance hours	(602)	11,601	(1,344)	9,357	(1,344)	2,814	(1,992)	3,416
AL carryover hours	(271)	6,900	(531)	3,059	(531)	2,587	(297)	2,587
AL hours earned	(386)	158	(350)	903	(96)	126	(50)	332
AL hours used	(320)	472	(872)	1,344	(1,090)	872	(1,360)	1,099
AL hours used ytd	(320)	1,881	(872)	1,997	(416)	1,911	(720)	2,039
Sick Leave (SL) balance hours	(3,792)	14,353	(3,185)	14,307	(2,722)	14,587	(2,648)	97,349
SL carryover hours	(795)	5,949	(1,248)	7,775	(2,507)	6,029	(1,214)	9,062
SL hours earned	(176)	720	(196)	720	(921)	720	(20)	720
SL hours used	(693)	875	(728)	1,131	(564)	1,051	(920)	352
SL hours used ytd	(32)	1,584	(32)	488	(8)	688	(24)	496
Comp time accrued ytd	(421)	2,097	(421)	2,097	(184)	1,467	(184)	1,371
Comp time earned	(421)	540	(177)	218	(228)	192	(183)	192
Comp time used	(124)	152	(99)	222	(128)	8,377	(94)	277
Comp time used ytd	(100)	1,258	(100)	1,059	(128)	8,379	(120)	8,407
Comp time travel accrued ytd	(109)	1,428	(80)	1,920	(105)	1,704	(105)	1,190
Comp time travel earned	(339)	144	(134)	162	(179)	191	(109)	207
Comp time travel used	(92)	104	(111)	112	(138)	87	(96)	120
Comp time travel used ytd	(62)	1,824	(111)	1,824	(90)	1,764	(92)	1,600

Source: GAO analysis of OPM data | GAO-17-127

Appendix II: Standards and Methods for Data Reliability Assessment

Data reliability assessments—a process consistent with internal control standards—gather and evaluate the information needed to determine whether data can be used to answer specific research questions. In this context, reliability means that data are reasonably complete and accurate to answer the intended questions that OPM, agencies, policy organizations, and academics might have about the federal workforce. Reliability assessments are specific to the context of the particular characteristics of the research project and the risk associated with the possibility of using insufficiently reliable data. Errors are considered acceptable if the associated risk has been assessed and a conclusion reached that the errors are not substantial enough to cause a reasonable person, aware of the errors, to doubt a finding, conclusion, or recommendation based on the data. To determine whether data are sufficiently reliable for a specific research purpose, one must consider the expected importance of the data in the final report; corroborating evidence; level of risk of using the data; and the results of assessment work conducted to date.

Completeness, accuracy, and validity are all components of reliability.

- Completeness refers to the extent to which relevant records are present and the fields in each record are populated appropriately. For example, are the payroll records for all on-board employees at an agency recorded for every pay period within the calendar year?

- Accuracy refers to the extent to which recorded data reflect the actual underlying information.[1] For example, do the recorded hours of annual leave in an employee's payroll record accurately reflect the number of annual leave hours they reported in their time and attendance form?

- Validity, for the purposes of this report, refers to whether the data actually represent what is being measured. For example, if we are measuring the extent of overtime in the federal government and we use a field that records a certain type of administratively

---

[1] Consistency, a subcategory of accuracy, refers to the extent to which the data entry requirements are clear and well defined to yield similar results, even if there are multiple entry points. If data are entered at multiple sites, inconsistent application of data entry rules can lead to data that, taken as a whole, are unreliable. For example, agencies may vary on what activities are considered as administrative leave, versus some other type of leave.

uncontrollable overtime, does that represent the extent of overtime use or might there be other ways overtime is recorded?

Data reliability assessments as a process include (1) reviewing existing information about the data and conducting interviews with officials from the entity or entities that collect the data, (2) reviewing selected system controls, and (3) performing tests on the data, such as advanced electronic analysis and tracing to and from source documents.

Appendix III: Comments from the Office of Personnel Management

Sidney Schwartz, Director
Center for Design Methods and Analysis,
U.S. Government Accountability Office
441 G Street, NW
Washington, DC 20548

Mr. Schwartz:

Thank you for providing us the opportunity to respond to the Government Accountability Office (GAO) draft report, **"Federal Human Resources Data: OPM Should Improve the Availability and Reliability of Payroll Data to Support Accountability and Workforce Analytics,"** (GAO-16-725).

GAO and OPM are in agreement regarding the need to enhance the quality of government payroll data. However, while we wholeheartedly concur with your recommendations, we also note that a lasting and effective solution for enhancing the quality of payroll data must also require consistent data quality at the origination, in addition to the "last mile" after delivery to the OPM Enterprise Human Resources Integration (EHRI) system.

We recognize that even the most well run programs benefit from external evaluations and we appreciate your input as we continue to enhance our programs. Responses to your recommendations are provided below. In addition to our responses, we note implementation of these recommendations will require collaboration between various stakeholders and appropriate resources.

Recommendation 1: GAO recommends that the Director of OPM take the following action to support its strategic and open data goals: Improve the availability of the EHRI payroll data—for example, by preparing the data for analytics, making them available through online tools such as FedScope, and including them among the EHRI data sources on the OPM website and Data.gov.

Management Response:

We concur. OPM will develop a comprehensive strategy and schedule to improve the availability of EHRI payroll data for analytics, as prescribed under Open Data. The strategy will include steps to:

- Inventory, review, and update payroll data standards following OPM's data governance model;
- Evaluate existing and develop different internal controls for data quality;
- Update payroll database documentation to be consistent with current field definitions and

requirements, including the Guide to Human Resources Reporting and the Guide to Data
Standards, Part B;
- Increase the availability of appropriate data to the public for research through enhanced
automation with real-time updates and online search capabilities.

Releasing payroll data to a publicly accessible database will require that all entities involved in
the chain of custody of payroll data (agencies, shared service centers, and OPM) comply with
agreed standards for timeliness and data quality. As described above, OPM will undertake what
is within its control; however, it cannot release data publicly that does not meet the standards.

Recommendation 2: GAO recommends that the Director of OPM take the following [action] to
improve internal controls for data quality: Update EHRI payroll database documentation to be
consistent with current field definitions and requirements, including the Guide to Human
Resources Reporting and the Guide to Data Standards, Part B.

Management Response:

We concur. We will address this recommendation as we implement the first
recommendation. Before updating EHRI payroll database documentation, OPM will validate
information and use resources and tools to correct information.

Recommendation 3: GAO recommends that the Director of OPM take the following [action] to
improve internal controls for data quality: Consistently monitor system-generated error and edit
check reports and ensure that timely action is taken to address identified issues.

Management Response:

We concur. OPM will develop a plan to monitor system-generated errors and edit check reports
and follow up with shared service centers and agencies concerning their reports. OPM will
identify and inform the agencies of edit checks on their data submissions, which will require
agencies to modify their systems and provide corrected data to OPM in a consistent and accurate
fashion.

Recommendation 4: GAO recommends that the Director of OPM take the following [action] to
integrate the payroll data into the larger suite of EHRI databases: Develop a schedule for
executing these plans.

Management Response:

We concur. We will address this recommendation as we implement the first recommendation.
This implementation plan will be in the OPM multiyear data strategy plan.

Recommendation 5: GAO recommends that the Director of OPM take the following [action] to
integrate the payroll data into the larger suite of EHRI databases: Evaluate existing internal
control activities and develop new control activities for EHRI payroll data, such as implementing
transactional edit checks that leverage the information in the other EHRI datasets

Mr. Sidney Schwartz

We concur. This recommendation is captured in 1, 3 and 4 above and answers apply to this
recommendation also.

I appreciate the opportunity to respond to this draft report. If you have any questions regarding
our response, please contact Ashu Goel, (202) 418-4326, Ashu.Goel@opm.gov.

Sincerely,

David DeVries
Chief Information Officer

Attachment:
    (1) Technical Comments on GAO Report (GAO-16-725)

Appendix IV: GAO Contact and Staff Acknowledgments

GAO Contact

Nancy Kingsbury, (202) 512-2700 or KingsburyN@gao.gov

Staff Acknowledgements

In addition to the contact named above, the following individuals made important contributions to this report: Sidney Schwartz, Director; Rebecca Shea, Assistant Director; Russ Burnett; Steven Putansu; David Blanding, Hiwotte Amare; Joanna Berry; Amy Bowser; Tim Carr; Melinda Cordero; Sara Daleski; Lorraine Ettaro; Dani Greene; Donna Miller; Laura Pacheco; and Jeffrey Schmerling.