ELDER RESEARCH
DATA SCIENCE · AI · MACHINE LEARNING

Evaluation of Data Governance, Management, and Validity

Final Report

Prepared for the EEOC Office of Inspector General

OIG Report No. 2024-002-EOIG

June 26, 2025

Table of Contents

1. EXECUTIVE SUMMARY

This evaluation, conducted by Elder Research on behalf of the U.S. Equal Employment Opportunity Commission (EEOC or the Agency) Office of Inspector General (OIG), assessed the effectiveness of the EEOC's data stewardship across the Agency. Effective data stewardship is critical for supporting mission-critical decisions and ensuring that decision-makers rely on accurate data for enforcement, reporting, and policy decisions that align with Agency objectives. We focused on these primary objectives:

1. Assess the EEOC's management and oversight for ensuring data validity and reliability.
2. Assess the efficacy of the EEOC's efforts to ensure the validity of information entered and maintained in the Agency Records Center (ARC) and the Enterprise Data Warehouse (EDW), focusing on the most critical data.
3. Assess how well the EEOC establishes, measures, and tests the validity and reliability of Agency mission critical data.
4. Assess the efficacy and efficiency of EEOC's data governance and management efforts.

To answer these questions, we described and assessed the overall oversight and management of data validation and reliability efforts by EEOC. We also assessed the EEOC's efforts to attain, measure, and improve data reliability and validity for ARC and EDW.

During this engagement, we relied on interviews and notes, supplemented as needed by EEOC process documentation and guidance. Our evaluation team examined how stakeholders moved and used data within the Agency. This approach helped us understand how the EEOC used data to make decisions and the corresponding effectiveness. Our evaluation team conducted eight meetings with groups of stakeholders from across the Agency and collected notes aggregated from three EEOC offices to inform its evaluation.

Overall, the EEOC has a solid foundation for data governance, management, and validity with regard to ARC and EDW. However, opportunities exist for continued improvement that will further develop data capabilities. The EEOC maintains strong data management practices, while ARC data validity has the most room for improvement, which stakeholders recognize and are working to address.

The EEOC demonstrates several data governance best practices, with room for improvement. Data governance practices are actively implemented and applied in multiple EEOC offices, top-level Key Performance Indicators (KPIs)[1] that align with the Agency's strategic goals are collected and presented to users in a dashboard, and office directors and staff are aware of the data governance vision and purpose. The EEOC lacks a centralized metadata repository[2] for either system, which hampers the Agency's ability to manage data effectively across its various systems. The Agency can improve its data governance capabilities and enhance data transparency and integrity by enhancing documentation, collecting more data governance KPIs, and further codifying the Data Governance Board.

---

[1] KPIs are quantitative and qualitative measures used to review organizational performance against its goals.

[2] A metadata repository is a centralized system that stores metadata, which is information about an organization's data, including but not limited to its name, schema, lineage, content, business rules, quality, ownership, transformation processes, and integration details.

Data management practices for ARC and EDW are well implemented. The EEOC has a centralized planning and management approach, actively identifies and mitigates risks arising from data management, and reduces manual processes by integrating automations throughout the data lifecycle, which leads to data quality outcomes that are predictable to users. There are some operational inefficiencies caused by data aggregation issues, restrictive permissions, and a lack of process transparency, which, if addressed, will improve user experiences and capabilities within ARC and EDW.

Uncertainties of sufficient scope negatively impact data validity within the EEOC. User data entry and individual workarounds created by users decrease the validity of the data within ARC and EDW. Stakeholders are aware of many of these data quality concerns and implement controls and risk mitigation measures such as training sessions, manual checks, and validation tools. Despite these controls, user data entry continues to create concern for data integrity and usability, highlighting areas for further development. These existing measures provide a framework upon which the Agency can develop data validity enhancements to address these concerns.

Based on the findings and conclusions, we recommend:

1. The Office of the Chief Information Officer (OCIO) and the Office of Enterprise Data and Analytics (OEDA) add governance information to all ARC and EDW documentation to ensure that users are well-informed and have access to crucial data governance details.
2. OCIO and OEDA collect and analyze additional data governance KPIs for data completeness and compliance, as well as metadata coverage, usage, and standardization, to improve the Agency's direct and quantitative assessment of ongoing data governance improvement efforts.
3. The Data Governance Board establish written procedures for its operations to enhance transparency, consistency, and accountability in its operations.
4. OCIO and OEDA create centralized and integrated metadata repositories for ARC and EDW to ensure consistent metadata accessibility and usage.
5. OCIO provide users with Zendesk ticket status and prioritization criteria to clarify resolution timelines and improve user satisfaction within ARC processes.
6. The Agency evaluate the cost and benefits of right-sizing its computing platforms to align with user requirements for current and predictable future business priorities to improve operational efficiency.
7. The Agency evaluate the cost and benefits of employing Artificial Intelligence[3] solutions to improve operational efficiencies in data validation, reporting, and analytics.
8. OCIO implement capabilities to analyze ARC data directly, with an analytics tool within ARC or through localized tools with direct connections to ARC, to reduce the creation and use of manual workarounds by users.
9. OCIO and OEDA evaluate the feasibility and operational benefits of implementing more fine-grained access controls within ARC and EDW.

[3] Artificial Intelligence refers to the simulation of human intelligence in machines that are programmed to think, learn, and perform tasks typically requiring human cognition, such as problem-solving and language understanding.

The appendices to this report contain an EEOC response to the draft report (Appendix A: EEOC Response to Draft Report) and OIG analysis of the EEOC's response (Appendix B: OIG Analysis of the Agency's Response).

2. INTRODUCTION

This document provides analyses into the Equal Employment Opportunity Commission's (EEOC or the Agency) data governance, management, and validity within the Agency Records Center (ARC) and the Enterprise Data Warehouse (EDW). These analyses provide greater understanding of the EEOC's data systems and support informed decision-making as new challenges and opportunities arise. This offers a reference for ongoing progress, offering insights and benchmarks to guide future development.

The content of this report is structured as follows:

1. Methodology: This section outlines the approach and techniques employed to gather and analyze information. The stakeholder offices and grading criteria are provided starting on pages five and seven, respectively.
2. Findings and Conclusions: This section shares the findings and conclusions of the evaluation with context for the evaluated data systems. Grades for each finding and focus area category are provided.
3. Recommendations: This section presents actionable recommendations based on the findings and conclusions. Specific strategies and improvements to address identified opportunities for improvement are included.

3. METHODOLOGY

3.1. Stakeholder Engagement

We gathered data by engaging with stakeholders identified by the EEOC Office of Inspector General (OIG). Some stakeholders referred us to additional individuals in their offices that we also gathered information from. Stakeholders received email communications containing a set of questions and had the option to respond to the questions in writing or participate in a virtual interview. The written responses received were sufficiently detailed, and we obtained necessary clarifications through email correspondence. Additionally, we conducted two follow-up interviews to gather further insights.

The stakeholder offices contacted during this evaluation, in alphabetical order, included:

1. Office of Enterprise Data and Analytics (OEDA)
2. Office of Federal Operations (OFO)
3. Office of Field Programs (OFP)
4. Office of General Counsel (OGC)
5. Office of the Chair

6. Office of the Chief Information Officer (OCIO)

We determined that OFO was not a major stakeholder for this evaluation, as they did not use ARC or EDW for their data processes. Individuals in all other offices provided us with information for this evaluation.

3.2. Data Collection

We gathered information for this assessment from two main sources:

1. Stakeholder Interviews: Stakeholders provided rich qualitative data through their interview responses on data governance, data management, and data validity. We consolidated these insights into comprehensive notes and categorized them thematically according to core focus areas.
2. Document Review: We collected data, policy, and process documents from stakeholders, including Standard Operating Procedures (SOPs), User Guides, Release Notes, Accessibility Guides, Historical Notes, Meeting Minutes, and other materials. These documents offered additional context and substantiated findings from our stakeholder interviews. We instructed stakeholders that they should not share any documents, data, or information that posed security or privacy concerns.

3.3. Data Categorization and Analysis

We organized the collected interview responses and documents into three categories to align with the evaluation's focus areas:

1. Data Governance: The processes and mechanisms that enact and enforce data governance policies, including methods for data collection, storage, access, and quality assurance.
2. Data Management: Operational aspects such as data processing speed and accuracy, data access protocol efficiency, and data quality assurance measure effectiveness. This includes the tools and technologies for data and metadata management, such as automated systems for data validation and error detection.
3. Data Validity: Methods, processes, and external tools for quality assurance and findings documentation.

After categorizing the information, we identified patterns and insights within each category.

The Atlan Data Governance Maturity Model provides the foundation for our data governance findings and grading criteria. This model stems from data governance best practices and draws on industry-wide standards from the Data Management Book of Knowledge (DMBOK) and the Control Objectives for Information and Related Technologies (COBIT). The Data Management Association International publishes the DMBOK as a comprehensive guide outlining best practices, standards, and methodologies for managing data effectively. It provides a framework for the principles, processes, and techniques organizations need to manage data as a strategic asset. It is used by the U.S. Department of Transportation, and its concepts are included in the U.S. General Services Administration's Data Governance guidelines. The COBIT is a globally recognized

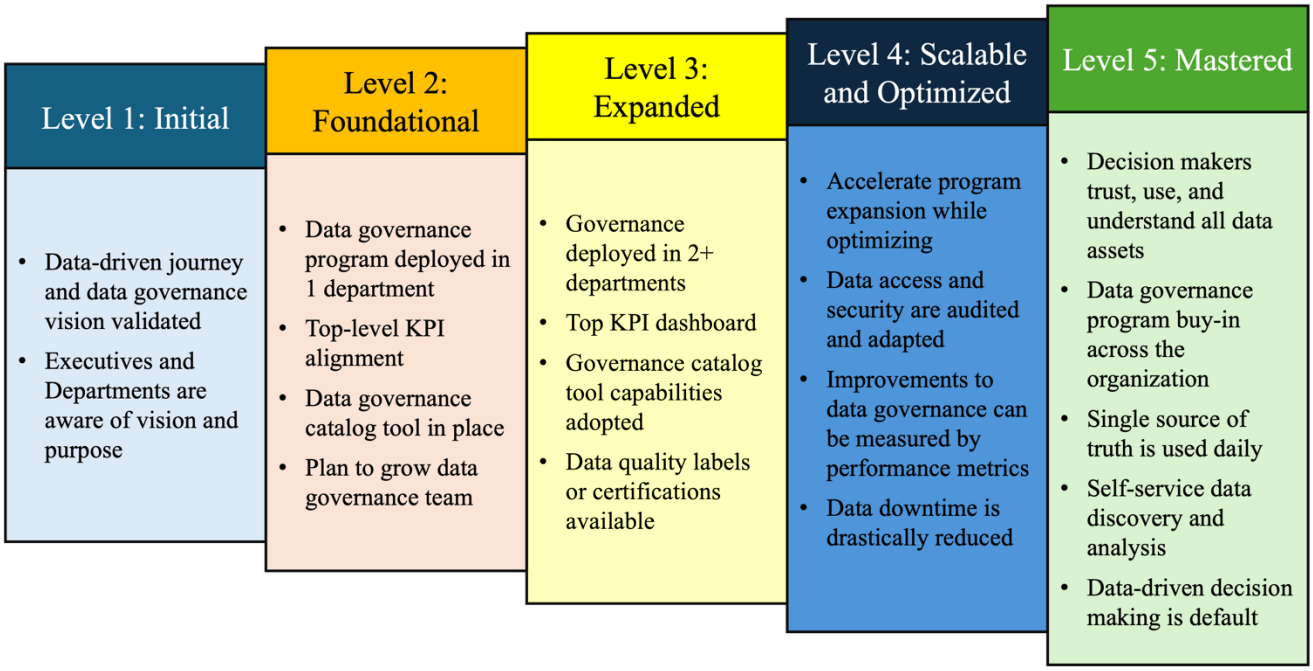
framework for Information Technology (IT) governance and management. It provides best practices, principles, and tools for organizations to effectively manage and govern their IT systems, including data and information resources.

We base our data management findings and grading criteria on the Data Management Maturity (DMM) Model. The DMM is a framework that helps organizations assess the maturity of their data management practices and draws best practices from the DMBOK and COBIT, with a focus on government-specific requirements. The model evaluates how well an organization manages, governs, and utilizes its data to meet business objectives and comply with relevant regulations.

Industry-standard data quality frameworks found in the DMBOK and the International Organization for Standardization (ISO) 8000 form the basis of our data validity grading criteria. The National Institute of Standards and Technology issues ISO 8000, widely recognized as the global standard for data quality. These models emphasize the impact of uncertainty, validity, accuracy, completeness, and data sufficiency—key aspects of any data quality assessment.

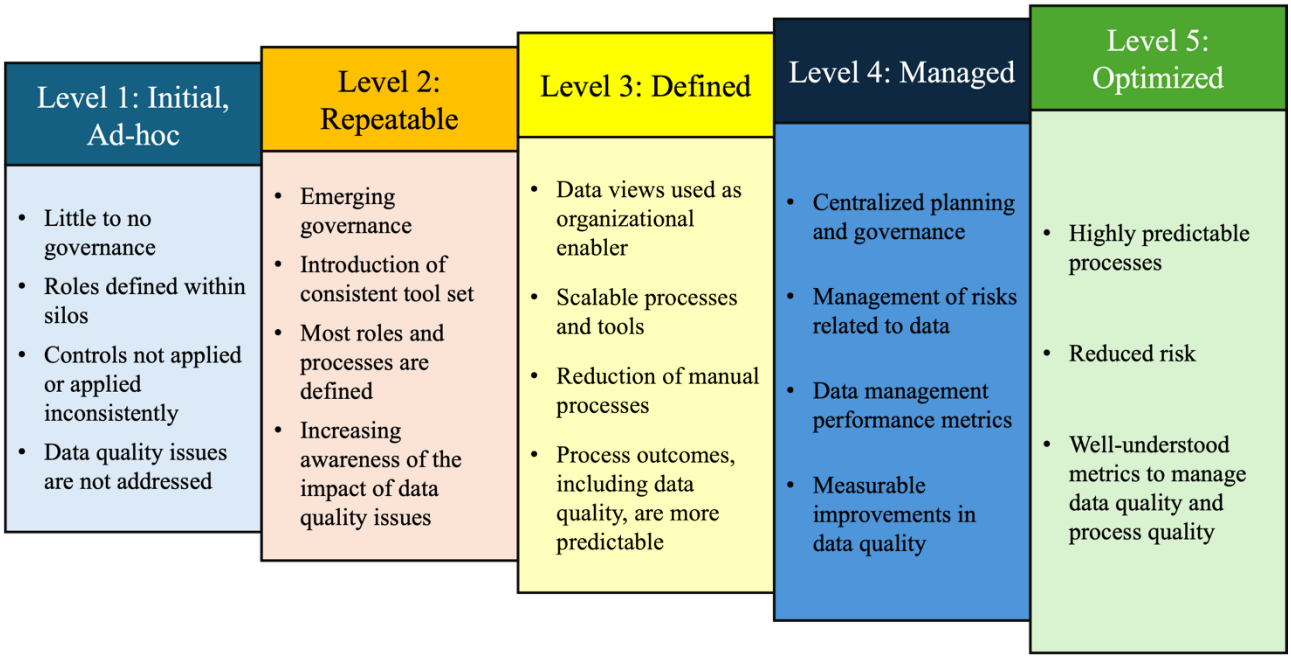
Each category uses the following grading levels and corresponding criteria:

Figure 1: Data Governance Maturity Model



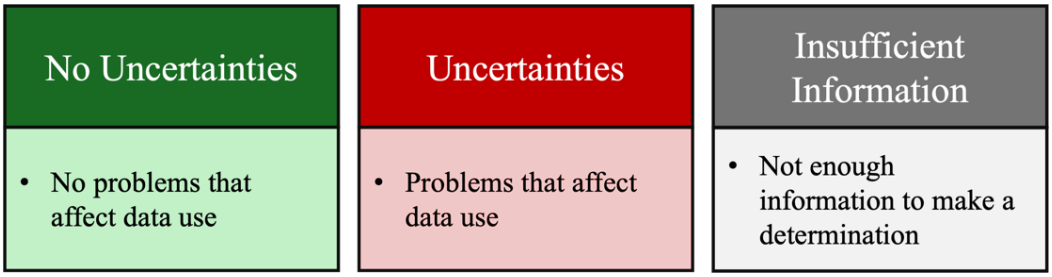
Source: Atlan Data Governance Maturity Model

Figure 2: Data Management Maturity Model



Source: DMBOK

Figure 3: Data Validity Outcomes



Source: Elder Research-created model based on DMBOK and ISO 8000

3.4. Grading

Grading is based on a comprehensive analysis of findings and conclusions from each focus area category (data governance, data management, and data validity). These grades indicate the EEOC's performance but are not prescriptive. Rather, they are intended to provide a more complete understanding of the EEOC's efforts and effectiveness in each focus area category.

For data governance and data management, we provide grades for individual findings and a total grade. For data validity, the grading outcome is more concise, wherein we provide key findings and a direct statement that summarizes the findings.

Data governance and data management findings generally reflect characteristics of grading criteria, and each of these characteristics has a corresponding level between one through five, as shown in Figure 1: Data Governance Maturity Model and Figure 2: Data Management Maturity Model. If a finding does not correspond directly to a characteristic outlined in the grading criteria, it is because we deemed the finding to be of sufficient importance to warrant a recommendation beyond the initial criteria. For these findings we applied grades we determined were aligned with the grading criteria based on our subject matter expertise.

To calculate the final grade for each category, we averaged the finding grades within each focus area category and rounded down (e.g., 3.5 would be rounded down to 3). Rounding down offers clarity on the overall state, reduces ambiguity, and encourages a culture of continuous improvement in the context of this data stewardship evaluation.

For data validity, we did not apply the five-level grading scale used for data governance and data management, as shown in Figure 3: Data Validity Outcomes. Instead, we identified significant key findings from the information gathered and provided a direct statement on data validity. This statement summarizes the findings and offers a clearer understanding of the current state of data validity within the EEOC.

3.5. Background

The EDW serves as a central repository for the EEOC to store and analyze charge data and EEO employer and union workforce reports data. The EDW stores various data including EEO-1 Component 1[4] (2019-2023), EEO-1 Component 2 (2017-2018), EEO-3 (Local Union Report, 1998 to 2022), EEO-4 (State and Local Government Report, 1973 to 2021), EEO-5 (Elementary-Secondary Staff Information Report, 1992 to 2016), and Private Discrimination Charge Data, which comprises all existing Integrated Mission System (IMS) data. The IMS is the EEOC's legacy database system that tracks and manages employment discrimination charge information. The EDW hosts dashboards that visualize EEO-1 Component 1 data and discrimination charge data.

ARC is the central records management system for the EEOC's enforcement work. It serves as a charge and case management system that supports users with private sector data, Fair Employment Practices Agency (FEPA)[5] Services data, and litigation data. The EEOC deployed ARC as a replacement for IMS in 2022. Data migrations from IMS to ARC are ongoing, including federal hearings and appeals data.

The EDW contains EEO data that the Agency collects and processes through various data validation and cleaning procedures. After ingestion into the EDW, the data is further processed for analytics & reporting needs, and to create Public Use Files (PUFs). ARC consists of EEOC inquiry and charge data, charge data from FEPA charges that are cross filed with the EEOC, and litigation

[4] Throughout this report, "EEO-1 Component 1" is used to refer to the workforce demographic data, and "EEO-1" refers to the EEO-1 data collection process at large.
[5] FEPA refers to state and local agencies that enforce laws prohibiting employment discrimination within their jurisdictions.

data. OFP staff in field offices manage inquiry and charge data in ARC. OGC manages litigation data in ARC. ARC also ingests data from the legacy system, IMS, through the Intermediate Reporting Database (IRD). OFP, OGC, OEDA, and OCIO validate the data that is input into ARC. Subsets of ARC data are then processed into the EDW and published in dashboards.

The EEOC uses the data in ARC and EDW for various reporting, analysis, and decision-making activities by numerous offices and stakeholders. Reports and publications that depend on these critical data and dashboards include, but are not limited to, the Annual Financial Report, Strategic Planning Process, Staffing Projections, private sector charge reports, Outreach and Hearings reports, and Webstats Enforcement and Litigation Statistics.

We created Table 1: Office Relationships with ARC/EDW below to contextualize office interactions with ARC and EDW. It provides an overview of the offices' roles and relationships with each data system to support a better understanding of ARC and EDW structures.

Table 1: Office Relationships with ARC/EDW

	ARC	EDW
Chair	Used for data needed to ensure operations align with mission, obtain performance insights, and support public policy.	Used for centralized data access to comprehensive data, to obtain performance insights, and to support public policy.
OEDA	Ensures policy compliance, provides data-driven insights and reports using ARC data, and transfers some ARC data into EDW. Has primary responsibility for data governance.	Used for centralized data for reporting, provides oversight and updates for the system, and provides training and support. Has primary responsibility for data governance.
OFP	Used for charge data and systemic case management, annual reviews for standards compliance, and data retention.	Used for public facing datasets and systemic investigations of employer discrimination.
OGC	Used for litigation docket, activity logs, and external portal integration.	Provides charge and litigation data.
OCIO	Ensures system reliability, provides data analytics capabilities, develops features, and contributes to data governance.	Provides technical management and data integration.

Source: Elder Research-created table, as reported to Elder by EEOC

4. FINDINGS AND CONCLUSIONS

4.1. Data Governance

4.1.1. Summary

The EEOC's data governance practices for ARC and EDW demonstrate a solid foundation for continued development. The following four findings and their conclusions highlight both strengths and areas for improvement:

1. Data governance practices are actively implemented and applied in multiple EEOC offices. OCIO provides a wide array of data governance documentation to stakeholders across various offices, and OEDA offers both documentation and training sessions (Level 3: Expanded). This facilitates cross-team collaboration with greater operational efficiency. However, data governance documents do not consistently include important governance information.
2. The EEOC aligns collected and measured top-level KPIs with the Agency's overall strategic goals and provides users with a dashboard reflecting these KPIs (Level 3: Expanded). This enables the Agency to self-evaluate operational and data governance efficacy. The Agency should collect additional data governance KPIs for data completeness and compliance, as well as KPIs for metadata coverage, usage, and standardization.
3. Office directors and staff are aware of the data governance vision and purpose (Level 1: Initial), as evidenced by their membership on the Data Governance Board, which defines and communicates these principles. This understanding supports data governance prioritization, resource allocation, and organizational alignment. However, the Agency has not comprehensively codified the Data Governance Board.
4. The EEOC lacks an easily accessible centralized metadata repository for all ARC and EDW users (Level 1: Initial). This creates risks of inconsistent data understanding and duplication of efforts.

The first three findings are foundational practices that position the EEOC for ongoing growth and refinement. By establishing and actively referencing a metadata repository for which metadata is available to all users, even those without access to data, the Agency will provide users with an essential resource to improve data understanding and mitigate potential errors. The EEOC's ARC and EDW data governance practices earn a total grade of 2 out of 5.

4.1.2. Finding 1: Implementation Across Offices

Data governance practices are actively implemented and applied in multiple EEOC offices. OCIO provides a wide array of data governance documentation to stakeholders across various offices, and OEDA offers both documentation and training sessions. However, data governance documents do not consistently include important governance information.

The EEOC benefits from a wide array of documentation for both ARC and the EDW that are consistently shared with users and developers across multiple offices. Available documentation to

users includes user guides, release notes, templates, and Standard Operating Procedures. OCIO and OEDA are responsible for developing and disseminating such resources, and they do so regularly.

Our team received and reviewed all data and dashboard user guides provided by OEDA and OCIO. ARC user guides provided by OCIO contain valuable information such as document update dates and 508 compliant features including alternative text for images. However, these guides inconsistently provide key information regarding the responsible party for 508 compliance and how to obtain support beyond submitting a helpdesk ticket. The user guides do not provide contact information or details for addressing further questions or concerns.

All data and dashboard user guides, as well as all seven PUF user guides provided by OEDA, lack the following information: specific contact information for questions, date of document's last update, who is responsible for enforcing expectations outlined in the guide, historical issues or changes, and roles and responsibilities. Some user guides direct readers to the OEDA helpdesk via email for questions but lack information regarding responsible parties for the documentation and systems, including titles or groups and how they may be contacted for matters beyond helpdesk questions. The EEOC can improve its data governance capabilities by ensuring that all ARC and EDW documentation contain important governance information.

OEDA conducts periodic training for field staff and data entry personnel across various offices to ensure accurate data collection that follows updated standards. OEDA also reviews updates to data lifecycle processes to ensure that they fall within data governance standards and regulations. OEDA ensures policies align with federal guidelines, such as the Federal Data Strategy.

Recommendation 1, Data Governance Documentation, offers suggestions for improvement related to this finding.

4.1.3. Finding 2: Key Performance Indicators

The EEOC aligns collected and measured top-level KPIs with the Agency's overall strategic goals and provides users with a dashboard reflecting these KPIs. The Agency should collect additional data governance KPIs for data completeness and compliance, as well as KPIs for metadata coverage, usage, and standardization.

The EEOC measures several KPIs to evaluate the effectiveness of its data governance efforts. These KPIs support the Agency by facilitating several objectives of data governance: 1) to inform decision-making, 2) to quantitatively assess data governance alignment with business objectives and regulatory requirements, 3) to drive continuous improvements of data governance practices through more readily identifiable opportunities, and 4) to gather employee and stakeholder feedback and engagement for process improvements. The Agency also maintains dashboards that provide users with live access to the top-level KPIs. This allows stakeholders to easily and quickly assess the status of key processes without requesting and reviewing various reports.

Examples of data governance KPIs that are currently being captured, as reported by OEDA and OCIO, include:

1. Number of New User Boardings: Tracks how many new users are onboarded to the EDW platform each month.
2. Time to Make Data Available: Measures the average time from when a dataset is received until it is made available to EDW users.
3. Number of Active Users: Tracks the number of users who have actively accessed EDW within a specific time window (e.g., weekly, monthly).
4. Number of User Logins: Measures total logins to EDW, indicating engagement and usage patterns.
5. EEO Data Collection Metrics: Tracks response rates, number of filers, roster or frame size, volume, type, and helpdesk ticket ages for each EEO data collection.
6. Data Dissemination KPIs: The number of datasets published annually, data compliance levels, percentage of data requests fulfilled at different levels of detail, and the accuracy of published data.
7. Extract, Transform, Load (ETL)[6] and Data Replication[7] Metrics: Tracks the performance and efficiency of ETL and data replication processes for data accuracy and accessibility.

The current inventory of collected KPIs is well-established and allows the Agency to monitor its data governance efforts. There are additional KPIs that the EEOC could measure and analyze, particularly in areas such as data quality, compliance, and metadata management, which would augment the Agency's data governance awareness.

Recommendation 2, Data Governance KPIs, offers suggestions for improvement related to this finding.

4.1.4. Finding 3: Office Awareness

Office directors and staff are aware of the data governance vision and purpose, as evidenced by their membership on the Data Governance Board, which defines and communicates these principles. This understanding supports data governance prioritization, resource allocation, and organizational alignment. However, the Agency has not comprehensively codified the Data Governance Board

The EEOC has established a Data Governance Board (DGB) whose mission is to "(provide) executive leadership and oversight for the development and implementation of the policies and processes which govern the collection or creation, management, use, and disclosure of EEOC data."[8] The DGB meets once a quarter to decide on the Agency's data strategy and review key data governance issues and updates. This is necessary for data governance prioritization, resource allocation, and organizational alignment. The DGB has been a key part of the Agency's efforts to enhance data governance, transparency, and decision-making processes, and has been awarded for

[6] ETL is a process that pulls data from different sources, cleans and organizes it, and then stores it in a central location where it can be easily accessed and analyzed.
[7] Data replication is the process of copying data from one location to another, ensuring that the same data is available in multiple places. This helps improve access, reliability, and backup of the data across different systems.
[8] EEOC Data Governance Board Charter (April 19, 2018)

data transparency. This recognition highlights the DGB's commitment to ensuring consistency and coordination.

The Chief Data Officer leads the DGB, with governance from the Office of the Chair. DGB members are representatives from offices including the Offices of the Chair, the Chief Financial Officer, Communications and Legislative Affairs, Field Programs, General Counsel, the Chief Information Officer, Inspector General, Legal Counsel, Enterprise Data and Analytics, and Federal Operations.

While the DGB maintains a Charter and consistent meeting minutes, it lacks other codification documents, and any such codification documents are inaccessible to EEOC members who are not in the DGB. There is no formal documentation regarding the DGB's processes for managing the agenda or implementation of changes. The Charter states that the Chair is responsible for disseminating material but does not specify what material. Existing documents do not clarify meeting frequency, whether additional artifacts exist, or how stakeholders can access such artifacts. The DGB has no procedure for tracking requests and their resolution, despite frequently receiving requests. The DGB is structured such that the Chair decides priorities, but no governing documents detail this process, which would enable predictable prioritization.

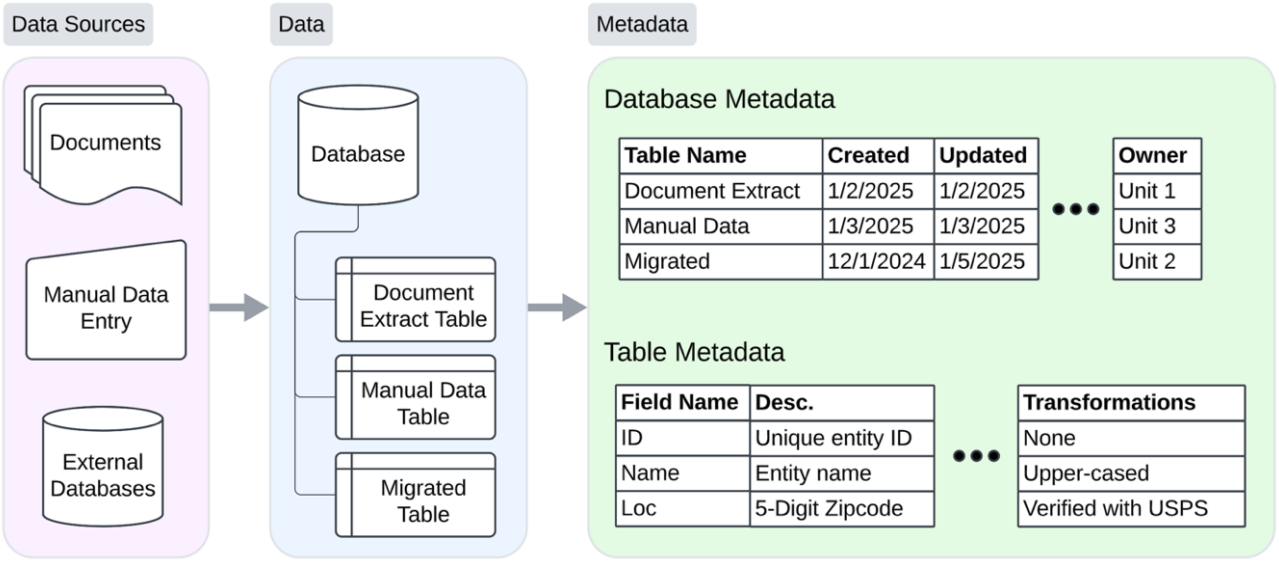
Recommendation 3, Data Governance Board Documentation, offers suggestions for improvement related to this finding.

4.1.5. Finding 4: Centralized Metadata Repository

The EEOC lacks an easily accessible centralized metadata repository for all ARC and EDW users.

Metadata is data that provides context and information about other data, including source, format, relationships, and usage. A metadata repository is a centralized system that stores metadata, including its structure, content, business rules, transformation processes, and integration details. When integrated directly with the system it supports, this repository becomes easily accessible to the users when needed. To illustrate the types of information commonly captured as metadata, we have included Figure 4: Metadata Diagram:

Figure 4: Metadata Diagram



Source: Elder Research

After reviewing all provided reference documents from OEDA and OCIO, we determined that the EEOC lacks a dedicated centralized metadata repository for ARC or EDW. While some documents such as the IMS Private Business Rules Manual and the EDW Source Data Inventory Report include metadata information such as definitions and table structures, no centralized resource exists for users to reference for comprehensive metadata information pertaining to ARC or the EDW. OEDA representatives report that they generate and share metadata information with external agencies that request it on a case-by-case basis.

The EEOC faces significant challenges due to the absence of a centralized metadata repository. Data ownership spans throughout the Agency, but users cannot consistently access information about ownership and responsibility for specific data systems. This complicates efforts to analyze the Agency's operations, monitor data lifecycle processes, and identify areas for improvement.

A representative in the Office of the Chair noted that it was common to find data points or fields that were not labeled accurately or consistently, potentially causing users to enter or use data differently. According to OEDA, the EEOC is working on implementing metadata management improvements, such as ensuring that North American Industry Classification System (NAICS) or industry codes are captured consistently during the intake of discrimination charges. The lack of a centralized metadata repository hampers the Agency's ability to manage data effectively across its various systems.

OCIO representatives acknowledged the potential benefit of a centralized metadata repository. They stated that the tool would significantly enhance their ability to standardize customer addresses, deduplicate records, and ensure the overall quality of data integrated in ARC and across all Portals. By providing accurate mapping, NAICS codes, and employee numbers, a metadata repository would provide valuable business intelligence.

Recommendation 4, Centralized Metadata Repositories, offers suggestions for improvement related to this finding.

4.1.6. Conclusions

The EEOC has a strong foundation in its data governance practices, but several key areas require improvement, resulting in an overall data governance maturity grade of 2 out of 5.

The Agency's extensive trainings and documentation, produced and disseminated to users, contribute to comprehensive, accessible, and transparent data governance practices. This facilitates cross-team collaboration with greater operational efficiency and reflects a data governance "Level 3: Expanded" grading criterion. However, documentation resources should be improved by incorporating additional governance details such as contact information for support, dates of latest document updates, historical issues or changes, roles and responsibilities, and 508 compliance considerations.

The EEOC measures various KPIs that align with strategic goals and provides users with a dashboard reflecting these KPIs. The Agency is able to self-evaluate operational and data governance efficacy using KPIs, reflecting a data governance grading criterion under "Level 3: Expanded." Though the current inventory is well-established, there are additional KPIs that the EEOC could measure and analyze, particularly in areas such as data quality, compliance, and metadata management.

Leadership across various offices are well-informed of the Agency's data governance vision. There is an established channel for feedback and communication through the DGB. While the DGB is crucial in supporting the Agency's data governance framework across offices, the DGB's processes are not comprehensively codified. The lack of documentation makes past decision and ongoing request inquiries challenging and may result in inconsistencies in data governance over time. The overall informed status of data governance purpose, communication channels for feedback, and limited process documentation and transparency reflect the data governance grading criterion in "Level 1: Initial."

The lack of a centralized, easily accessible metadata repository for all ARC and EDW users hinders effective data management and consistency across the Agency. Without a centralized metadata repository, the EEOC faces issues such as replications of effort, delays, reduced data integrity, and unrealized opportunities for operational efficiencies. We evaluate this finding to be at a "Level 1: Initial," as a grading criterion within "Level 2: Foundational" is for there to be a data governance catalog tool in place. The creation of a centralized metadata repository for each data system would provide a single, unified source of truth for managing and accessing metadata, ensuring consistency, improving data governance, and optimizing processes across the EEOC.

Overall, the EEOC has established foundational governance practices in documentation, KPI tracking, and executive oversight. Enhancing governance documentation, expanding measured KPIs, further codifying the DGB, and implementing a centralized metadata repository are critical steps toward further improving the Agency's data governance capabilities and will lead to greater data transparency and integrity.

4.2. Data Management

4.2.1. Summary

The EEOC's data management practices for ARC and EDW are well-defined and administered, resulting in more predictable, consistent, and reliable data quality outcomes. While there are opportunities for improvement, including implementing more fine-grained access controls within ARC and EDW, increasing Zendesk process transparency, and addressing data aggregation issues, the current practices maintain robust data accessibility and scalability. The five data management findings and their conclusions are as follows:

1. The EEOC has a centralized planning and governance approach that is executed by the Data Governance Board and leaders in OEDA, OCIO, and OFP (Level 4: Managed). This promotes consistent decision-making and streamlines operations by improving scalability and flexibility.
2. The EEOC actively identifies and mitigates risks arising from data management (Level 4: Managed). This improves long-term sustainability and scalability.
3. The EEOC reduces manual processes by integrating automations throughout the data lifecycle (Level 3: Defined). This increases operational efficiency and reduces the risk of user error.
4. The EEOC's well-defined and properly managed processes lead to data quality outcomes that are more predictable, consistent, and reliable (Level 3: Defined). This leads to improved decision-making and greater confidence in the Agency's data-driven insights.
5. Data is vulnerable to inconsistency, inaccuracy, and inefficiencies due to data aggregation issues (Level 1: Initial, Ad Hoc). Addressing these aggregation issues would improve data quality and avoid continued inconsistencies, inaccuracies, and inefficiencies.

The first four findings show that the EEOC has established strong data management and has taken steps to streamline operations, make process outcomes more reliable, and reduce the risk of user error. The last finding is a point of improvement that, if appropriately addressed, will increase the consistency, accuracy, and efficiency of the Agency's data. The total grade for the EEOC's ARC and EDW data management practices is 3 out of 5.

4.2.2. Finding 1: Centralized Planning and Governance

The EEOC has implemented a centralized planning and governance approach executed by the Data Governance Board and leaders in OEDA, OCIO, and OFP.

The EEOC maintains a central governing authority that oversees data initiatives across all offices. As mentioned previously, the EEOC's DGB provides leadership and oversight for the handling of EEOC data. Representatives in OEDA, OFP, and OCIO act as the governing bodies of ARC and EDW. These representatives collaborate to achieve several objectives: 1) to make strategic decisions regarding data management, 2) to set standardized data processes and policies, 3) to take ownership and accountability for data management tasks, 4) to ensure that data governance frameworks address risks and privacy concerns, and 5) to promote cross-departmental data integration.

Some ARC users outside of OEDA, OFP, and OCIO serve as high-level representatives and assist with prioritizing and directing development updates in ARC. When these representatives identify opportunities for improvements and uncover issues with ARC, whether with the application or data structure, they collaborate with ARC leadership to identify root causes, ideate on solutions, and define enhancement requirements.

Zendesk, the ticketing and request system for ARC, does not have transparent processes. While the process of submitting a ticket is straightforward, the post-submission process for tickets is unclear to users. Users are not provided with information regarding how decisions are reached, the prioritization of tickets, or the status of submitted tickets. Being transparent with users regarding these processes would enhance confidence in the system reliability.

Recommendation 5, Zendesk Transparency, offers suggestions for improvement related to this finding.

4.2.3. Finding 2: Risk Identification and Mitigation

The EEOC actively identifies and mitigates risks arising from data management.

ARC and EDW access is controlled, based on defined permissions and roles. Data users and data entry personnel do not have unfettered access to all data in either system. It is important to distinguish between data entry personnel (i.e., field staff) who input data into ARC, and data users (i.e., HQ staff including OCH and OEDA) who primarily view or analyze the data but do not contribute to its entry. Data entry personnel carry a higher risk to data integrity, as inaccuracies introduced at this stage can compromise the validity of the data. However, data users also impact data integrity through interpretation, analysis, and reporting. For instance, when users conduct ARC or EDW data analyses outside of the respective systems, they risk generating inaccurate or inconsistent outputs. Thus, both data entry personnel and data users can influence data quality.

Within each EEOC office, there is a designated official who manages role-based ARC access permissions for users within their office. In addition to EEOC staff users, ARC database administrators, system administrators, developers, and help desk support personnel have access to ARC systems, as needed. Each user is granted roles or permission within ARC that controls access to data at the application and form level. Field employee access levels are determined by the EEOC office director. For headquarter employees, access level is determined by their office director in conjunction with the specific application's system owner. The access level is based on the user's role and the extent to which need-to-know indicates the necessity of data access to perform their job duties.

Restricting data access permissions is an effective way to reduce data management risks. However, stakeholders in the Office of the Chair, OFP, and OEDA have encountered operational inefficiencies due to limited access to live data and customized reporting tools. Staff in the Office of the Chair stated that being given easy access to relevant EDW and ARC dashboards would be beneficial, rather than needing to wait for quarterly reports or make ad-hoc requests. Interviewed stakeholders submit requests as needed to OEDA or OCIO for more data permissions within the EDW or ARC, respectively. Many users rely on time-intensive support to request specific data points or report enhancements and have stated that delivery timelines can take weeks. There is also

growing reliance on external tools such as Excel or R workarounds, particularly when ARC's reporting functionalities do not meet user needs. Implementing more fine-grained access controls within ARC and EDW could improve operational efficiency and data integrity within the Agency.

While access is restricted to relevant users, a critical point of vulnerability for errors is user entry of data. Data entry errors are a common issue arising from manual processes and can result in poor data integrity when left unmitigated. As such, OFP, OEDA, and OCIO actively work to minimize such errors and enhance data integrity within ARC by jointly issuing quarterly guidance to field staff providing data validation instructions and conducting regular data validations. In addition, OFP takes measures to ensure data is correctly entered into ARC by providing training, refreshers, and updates to users. Data validation measures, such as type-to-search lookup functions[9], have also been deployed at the user entry level.

Recommendation 8, ARC Analytics Tools, and Recommendation 9, Data Permissions, offer suggestions for improvement related to this finding.

4.2.4. Finding 3: Reduction of Manual Processes

The EEOC reduces manual processes by integrating automations throughout the data lifecycle, resulting in more consistent process outcomes and improved operational efficiency.

Several manual processes for ARC and EDW have been automated. OCIO has developed automated scripts and tools which are used to refresh data from IMS to ARC incrementally. This process uses content-aware data migrations to limit time and resource utilization when refreshing ARC data from IRD. OCIO's automated incremental data loading minimizes processing time and ensures data consistency.

OEDA employs SAS and R automation to conduct validation of multiple records simultaneously using rigorous data validation protocols within the EDW. These automation scripts are designed to identify common issues such as missing data, duplicate records, and format inconsistencies. Running these scripts concurrently on different segments of data reduces overall validation time for the EDW.

OEDA has also developed a system to stream data from ARC into the EDW, automatically integrating, validating, and making available for reporting new updates from employers and charge data collections.

To accommodate changes to data that necessitate updates and refreshes in the EDW, OEDA has developed automated pipelines to execute the complete process of loading, validating, and publishing data to dashboards within minutes. This customizable automation allows OEDA to execute limited portions of the system when only specific capabilities are needed.

Recommendation 7, Artificial Intelligence, offers suggestions for improvement related to this finding.

[9] Search features that show accepted data input results while typing a query.

4.2.5. Finding 4: More Predictable Data Quality Outcomes

The EEOC's well-defined and properly managed processes lead to data quality outcomes that are more predictable, consistent, and reliable.

Both ARC and EDW users have access to trainings to stay abreast of changes to data entry, data collection, new tools and technologies, and agency standards.

External data collection for ARC follows standardized procedures with clear entry points through systems such as the Public Portal, E-File for Attorneys, and the Respondent Portal. In several interviews, Agency stakeholders were able to accurately relay how they and external users entered data into ARC. Interviewed stakeholders had a strong understanding of the existing data validation process as it pertained to their work. ARC leaders were aware of the occurrence of data entry errors by users and were pursuing multiple approaches to reduce and efficiently address the errors.

For both ARC and EDW, stakeholders could identify relevant offices or individuals in offices they contact regarding questions or concerns they had, including the observation of data accuracy issues. Regarding the EDW, stakeholders did not express concerns about data quality that would fall under the EEOC's purview. For example, a leader from the Office of the Chair acknowledged that the EEO data the Agency collected was often incomplete because individuals sometimes failed to file reports despite the mandatory filing requirements.

4.2.6. Finding 5: Aggregation Issues

Data is susceptible to inconsistency, inaccuracy, and inefficiencies due to data aggregation issues.

Migration from the legacy system, IMS, to ARC has created challenges with aggregated data[10] fields; it is common to find that data points or data fields are not accurately or consistently labeled. This results in varied interpretations of data fields, leading to data entry inaccuracies.

Regarding EDW data, the aggregate count data of the EEO-1 Component 1 employer reports that OEDA receives directly from the Commission force the data into broad categories and unnecessarily creates zero-count cells, as reported by OEDA. These issues greatly increase the size and processing time of the data relative to the expected time were each record correctly limited to a single individual employed by the employer. The Agency faces high runtimes of data due to the aggregate count structure of EEO-1 Component 1 data and the current computing resources the Agency uses. The current compute capacity of the SAS server and local laptops available to OEDA are frequently insufficient; runtimes of the EEO-1 Component 1 data can range between five to ten minutes up to 12 hours. Such long run-times significantly hinder efficiency, increase resource consumption, and cause delays in decision-making.

[10] Data aggregation is the process of collecting and combining data from various sources, then summarizing or transforming it into a format that is more useful for analysis, reporting, or decision-making.

Recommendation 6, Computing Resources, offers suggestions for improvement related to this finding.

4.2.7. Conclusions

The EEOC's data management processes are well-defined and administered, resulting in more predictable, consistent, and reliable data quality outcomes. The EEOC's overall data management maturity level for ARC and EDW is 3 out of 5.

The Agency's centralized planning and governance approach ensures consistent and standardized data management practices across the Agency, promoting efficiency, compliance, and alignment with business objectives. It supports consistent decision-making and streamlines operations by improving scalability and flexibility. Centralized planning and governance reflect a data management grading criterion in "Level 4: Managed." One way the EEOC can improve user confidence in the reliability and predictability of their data processes, which is a grading criterion in "Level 5: Optimized," is by offering more transparency and clarity in the Zendesk ticketing system.

The EEOC actively identifies and mitigates risks arising from data management. Access permissions for both ARC and EDW are role-based, reducing the risk of unauthorized data modifications. Both data entry personnel and data users influence data integrity. Data entry personnel potentially introduce inaccuracies during input, and data users can affect data integrity through varying interpretations and external analyses, which can lead to conflicting information across the organization. These permissions ensure data security while aligning with user roles and responsibilities. However, operational efficiency is hindered when user permissions are too restrictive, such as when individuals require task-specific heightened access to data. This indicates that the available user roles and access control customization are not yet sufficiently granular to adequately represent all real use cases.

While access is appropriately restrictive for defined roles, user entry of data remains vulnerable to errors. The Agency provides trainings, refreshers, and updates to users and has implemented validation measures to ensure data is entered correctly into ARC. These practices improve long-term sustainability and scalability of the EEOC's data systems and reflect a data management grading criterion in "Level 4: Managed."

The EEOC automates key data lifecycle functions, such as data validation and importing. The reduction of manual processes with effective automations increases operational efficiency and reduces the risk of user error for both ARC and EDW. This reflects a data management grading criterion in "Level 3: Defined." The EEOC could continue to build on this foundation by exploring additional potential Artificial Intelligence enhancements to further reduce manual processes in areas such as reporting and analytics.

Standardized processes for data entry and validation have made data quality more predictable, though user errors remain an ongoing concern, particularly for ARC. These errors are being mitigated through trainings, manual checks, and validation tools. There is potential for improvement as the errors create concern for data integrity and usability, but the current measures are a reasonable foundation for further development. Having standardized processes that make

data quality more predictable results in improved decision-making and greater confidence in the Agency's data-driven insights. This reflects a data management grading criterion in "Level 3: Defined."

Data aggregation issues leave data prone to inconsistency, inaccuracy, and inefficiencies. Aggregated data from the IMS to ARC migration process have led to data field issues and are exacerbated by source data quality and inconsistent data field labeling. The current aggregate structure of EEO-1 Component 1 data significantly increases dataset size and processing times, as reported by OEDA. We evaluate this finding to be at a "Level 1: Initial, Ad Hoc," as these data quality issues are not currently being addressed. To improve data quality and avoid continued inconsistency and inefficiencies, the EEOC should address data aggregation issues, and evaluate computing resources for improvement. Left unaddressed, the Agency will continue to face inefficiencies that lead to higher costs and reduced data reliability.

The EEOC's data management system for both ARC and EDW is strongly defined and well-managed. It demonstrates a satisfactory balance between security, scalability, and accessibility. While there are some areas for improvement, such as customizing ARC and EDW user permissions, increasing Zendesk process transparency, and addressing data aggregation issues, the current practices provide robust scalability, security, and usability.

4.3. Data Validity

4.3.1. Summary

The EEOC has demonstrated a commitment to data validity, though both ARC's and EDW's data validity could be improved. The following two findings and their relevant conclusions highlight current strengths and opportunities for improvement:

1. The EEOC is aware of data accuracy and validity challenges within ARC and employs controls to mitigate risks. The EDW system has a robust and iterative data validation process that ensures high data integrity. The EEOC is well-positioned to ensure consistent data quality across both systems, enhancing the overall reliability and integrity of Agency data in the long-term.
2. Some users create workarounds when their needs extend beyond ARC's current functionalities. User reliance on manual workarounds compromises the overall integrity of ARC data usage.

EEOC data validity is negatively impacted by user entry errors and manual workarounds. The controls for mitigating data error risks in ARC and the robust validation processes in the EDW position the EEOC well for long-term data reliability and integrity. Reducing user needs for workarounds will further enhance data consistency, accuracy, and reliability.

4.3.2. Finding 1: Data Accuracy and Validation Controls

The EEOC is aware of data accuracy and validity challenges within ARC and employs controls to mitigate risks. The EDW system has a robust and iterative data validation process that ensures high data integrity.

ARC relies on user accuracy[11] for data entry, which creates risks of incomplete, inconsistent, or erroneous data. OCIO reported that as of October 4, 2024, the ARC application held over six million records of Charging Party, Respondent, and Representative addresses, and grew at an average of 745,000 records per year. OCIO stated that duplicate records, poor data integrity, and inconsistent data formatting result in operational ineffectiveness, miscommunications, and unreliable business intelligence.

To counter these issues, the EEOC implements the following controls: 1) training programs for staff to ensure consistent data entry and compliance with agency standards, 2) annual performance reviews of data quality and compliance, and 3) use of validation tools such as SAS and automated R scripts for ongoing data integrity checks. When concerns regarding ARC data or its PowerBI reports arise, OFP conducts manual reviews of the data and collaborates with OCIO to ensure appropriate solutions are implemented. According to OEDA, OEDA also works in collaboration with OFP to perform manual reviews of the data. Addressing these issues takes considerable time and resources.

The EEOC faces an additional data validity challenge in the reliability of current EEO-1 Component 1 and ARC data linkages. OEDA noted that the EEO-1 Component 1 Online Filing System employs heavy data validation and standardization (e.g., using the Google API to validate addresses and accepting only valid NAICS codes). While data standardization at the ARC data entry stage is valuable, addressing broader data linkage challenges may benefit from a more centralized and consistent approach to standardization across the data systems. OCIO acknowledged that a possible solution would be to integrate EEO-1 data collection with centralized data standardization and enhancement services. This would significantly improve the quality of the EEO-1 Component 1 data and accurately link it to reports and respondents in ARC.

OEDA teams and contractors perform multiple rounds of EDW data checks to ensure data accuracy. After each data collection period, contractors conduct substantive checks for data issues before providing data files to the EEOC. The Employer Data Team within OEDA then reviews the files, using programmatic checks for data quality issues. The Employer Data Team then remits the file to the Data Policy and Access Team for additional review, and upon completion provide the file to OCIO for loading into EEOC systems for use.

When transferring data to the EDW, OEDA performs additional data profiling. Data is often loaded in parallel batches to allow simultaneous data profiling and validation. This involves analyzing the data for any irregular patterns, errors, or inconsistencies not detected by prior checks. Data profiling further ensures the data loaded into the EDW is accurate and suitable for use in internal analytics and reporting tools.

[11] User accuracy is the degree to which a user correctly enters data into a system.

The EEO Online Filing System collects data used for burden estimates[12], allows filers to access historical reports between collection, collects the mandatory EEO-1 Component 1, EEO-3, EEO-4, and EEO-5 reports, performs real-time validation, and produces the reports upon certification. The EEO Online Filing System must pass User Acceptance Testing before data collection begins, which ensures the system accurately captures data provided by filers. Measures that are checked during testing include data validation and built-in error messages. OEDA also ensures that the data provided is accurately captured and preserved once stored.

After data dissemination, the EEOC provides ongoing monitoring and feedback mechanisms. Collaborations, such as the Memorandum of Understanding with the Bureau of Labor and Statistics to match EEO-1 Component 1 data with the Quarterly Census of Employment and Wages, enable cross-validation of EDW data with external sources, enhancing its reliability throughout its life cycle. OEDA also evaluates each stage of the data life cycle through feedback from internal teams and external data users. This continuous feedback loop allows OEDA to identify areas where processes can be enhanced, such as the need for more streamlined data collection methods or enhanced validation checks.

Recommendation 4, Centralized Metadata Repositories, offers suggestions for improvement related to this finding.

4.3.3. Finding 2: Manual Workarounds

Some users create workarounds when their needs extend beyond ARC's current functionalities. User reliance on manual workarounds compromises the overall integrity of ARC data usage.

Stakeholders reported that ARC's functions did not meet all their needs, leading them to develop workarounds to carry out job functions. Relying on manual workarounds for data analytics as compared to relying on ARC as a single source of truth undermines the consistency of the information. When users perform analyses or generate reports using different methodologies, this can result in conflicting outputs for the same data point(s), which reflects a vulnerability in data integrity. Validity of knowledge extracted from ARC can be ensured by pulling additional analyses and steps into ARC.

A stakeholder in the Office of General Counsel noted that ARC was designed as a workflow and data capturing tool rather than a data analytics or reporting tool. To account for their office's analytical needs, they utilized R-Shiny to develop analytics tools.

Because ARC itself has very rudimentary reports, OCIO staff have built PowerBI dashboards with ARC data. In the Office of the Chair, one professional uses these PowerBI reports, but due to limitations with the reporting system, extracts and exports the data using Excel to facilitate further analyses.

[12] Burden estimates are calculations that encompass the total cost of employing someone, including not only their salary but also indirect costs like payroll taxes, benefits, and other expenses.

Staffing data is managed using an Excel spreadsheet rather than ARC. This significantly limits performance analysis, as direct comparisons cannot be made between this data and data housed in ARC. Staff in the Office of the Chair noted that they would benefit from seeing additional trend-analysis capabilities in ARC dashboards beyond those currently available.

The highly hierarchical approach to ARC data usage is a key limitation for users. Users have limited access and control for report generation, preventing many from effectively engaging with the data. Representatives in the Office of the Chair have noted that OCIO will design and develop data structures that are not necessarily aligned with user business needs.

Recommendation 8, ARC Analytics Tools, and Recommendation 9, Data Permissions, offer suggestions for improvement related to this finding.

4.3.4. Conclusions

EEOC data validity is negatively impacted by user entry errors and manual workarounds. However, the EEOC is proactively taking steps to identify and address issues. While substantial progress is being made to improve data validity, challenges remain, and there is considerable opportunity for further enhancement, particularly for ARC.

The EDW's data lifecycle process is robust, with iterative data validation processes ensuring data integrity, while the data lifecycle processes of ARC raise data validity concerns. Recurring issues with ARC data consistency and labeling underscore the need for a comprehensive data dictionary and enhanced metadata governance. User data entry within ARC creates risks of incomplete, inconsistent, and erroneous data. These concerns are being addressed by several EEOC offices through training, validation tools, and multiple rounds of checks throughout the data lifecycle. The availability of comprehensive documentation that aligns with training would reinforce standards and enhances data-entry quality. By actively identifying data validity challenges and applying measures to address them, the EEOC is well-positioned to ensure consistent data quality across both systems, supporting the overall reliability and integrity of the data in the long-term.

User reliance on manual workarounds reduces ARC's validity as a single source of truth. These workarounds introduce inconsistencies and inefficiencies that compromise the overall integrity of ARC data usage. When users perform analyses or generate reports using different methodologies, this can result in conflicting outputs for the same data point(s), which reflects a vulnerability in data integrity. Reducing reliance on manual workarounds by building out ARC functionality that addresses user needs would improve data reliability and usability, ultimately fostering more efficient and consistent data management practices.

The EEOC faces data integrity issues stemming from inconsistent data entry and system limitations but has identified these challenges and implemented several measures to address them. This demonstrates that while there are uncertainties of sufficient scope that negatively impact data validity within the EEOC, the EEOC is taking active steps to improve. Furthermore, improving system capabilities, standardizing data understanding, and reducing the need for manual workarounds will greatly enhance data validity within the EEOC, ensuring more efficient and effective decision-making and reporting.

5. RECOMMENDATIONS

The EEOC has established a strong foundation for data management, governance, and validity regarding to ARC and EDW. The following recommendations address key opportunities for improvement based on data evaluation findings.

1. Data Governance Documentation

This recommendation relates directly to Data Governance Finding 1: Implementation Across Offices.

We recommend OCIO and OEDA add governance information to all ARC and EDW documentation to ensure that users are well-informed and have access to crucial data governance details. ARC and EDW documentation that should have governance information added to them include documents that users reference to understand the systems and documents derived from the systems' data products. Governance information that should be added includes, but is not limited to, contact information for support, dashboard roles and responsibilities, data owners, and 508 compliance information.

This change has two components. First, gathering the necessary information for each data or analytics object during the creation of centralized governance information. Second, scheduling maintenance of this data governance information with other routine data management tasks to ensure the information collected remains current. Both steps can be performed using either a phased approach that systematically updates documentation during creation and maintenance incrementally, or through a more concerted initiative for either or both components to expedite completion.

2. Data Governance KPIs

This recommendation relates directly to Data Governance Finding 2: Key Performance Indicators.

To improve the Agency's direct and quantitative assessment of ongoing data governance improvement efforts, we recommend that OCIO and OEDA collect and analyze additional data governance KPIs for data completeness and compliance, as well as KPIs for metadata coverage, usage, and standardization. The Agency is currently collecting various KPIs to measure its effectiveness in achieving strategic goals and objectives and to support strategic decision-making. We advise collecting and monitoring these additional KPIs to augment data governance awareness:

1. Data Completeness: Assesses whether all required data elements are present and available, ensuring no missing or incomplete information
2. Compliance with Regulatory Standards: Measures the extent to which data and processes adhere to relevant laws, regulations, and industry standards

3. Metadata Coverage: Assesses the rate of data assets that are fully documented with complete metadata (e.g., "90% of data assets are fully documented with complete metadata")
    a. Metadata coverage typically includes, but is not limited to, information such as data definitions, ownership, purpose, applicable restrictions, validation rules, lineage information, and access requirements.
4. Metadata Access Frequency: Tracks how often metadata is accessed or used
    a. Metadata access frequency is the direct assessment of usage of the information described in metadata coverage, such as a query count for a metadata table.
5. Metadata Standardization Rate[13]: Measures the percentage of metadata that adheres to defined organizational metadata standards (e.g., "90% of all metadata follows the defined organizational metadata standards")

KPI prioritization requires EEOC stakeholders decide which data governance KPIs best align with Agency needs. Implementation strategies may vary by KPI. For example, metadata access frequency and data completeness may be directly measured, while the application of standards and assessment of metadata coverage require more evaluation.

3. Data Governance Board Documentation

This recommendation relates directly to Data Governance Finding 3: Office Awareness.

We recommend that the Data Governance Board establish written procedures for its operations to enhance transparency, consistency, and accountability in its operations. Implementing a systematic approach to tracking and addressing requests will allow the Agency to effectively log, categorize, and monitor all past and current requests, and facilitate examination of past requests and DGB actions by staff outside of the DGB. Recognizing that not all requests or actions should be public, the DGB may choose to specifically limit publication of submitted requests or DGB actions as necessary.

The creation of formalized processes, including a code of conduct and a clear statement of core values, provides clear guidelines for decision-making and ensures the DGB's standards are upheld. A clear framework for succession planning ensures leadership transitions are well executed. Designing and documenting a procedure for tracking and addressing requests will enable the Agency to effectively log, categorize, and monitor all past and current requests.

4. Centralized Metadata Repositories

This recommendation relates directly to Data Governance Finding 4: Centralized Metadata Repository and Data Validity Finding 1: Data Accuracy and Validation Controls.

[13] Metadata standardization rate refers to the overall metadata coverage rate across all applicable data.

We recommend that OCIO and OEDA create centralized and integrated metadata repositories for ARC and EDW to ensure consistent metadata accessibility and usage. Some existing EEOC documentation include metadata definitions and descriptions and are distributed upon individual requests. However, we recommend the establishment of a comprehensive and integrated metadata repository that is accessible to all users and would serve as a single, authoritative resource for understanding and managing all metadata for each system. This would provide consistency and accessibility for all users, as metadata would be available to all users, including those without direct access to specific data, facilitating data discovery. By establishing centralized metadata repositories that are easily accessible to all users when they are interfacing with ARC or EDW data, the EEOC can improve operational efficiency, data integrity, and data management. Below are key actions OCIO and OEDA should undertake in the development and management of these repositories:

1. Develop the metadata repositories, either in a software tool or as documents or tables. Best practice is to integrate metadata directly into the data warehouse itself in tabular form. This makes it accessible to users within the scope of their typical system interactions. When metadata is integrated directly into the data warehouse itself, queries may report fields directly by short names, long names, or descriptions without referencing a separate document. Queries can also report missing data directly and routinely.
2. Include both static and dynamic information in the repositories.
    a. Static metadata includes the full data dictionary, data lineage, accountable parties, business rules, and data models. This static metadata can allow users to understand how data within ARC and EDW are structured.
    b. Dynamic metadata includes required coverage, such as the evolving set of entities the data is intended to cover, and at what frequency. Dynamic metadata results are derivable by query after each load cycle and include both coverage and quality metrics; it enables governance of data collection by comparing expected coverage with actual coverage over time.
3. Include the validation methods employed to verify data integrity in the repositories.
    a. This should include the status of the validations (e.g., results from the latest data assessment) and any known issues or limitations. This will enable greater transparency to users, such that they are better able to understand the quality and reliability of the data, ultimately supporting informed decision-making and more effective data management practices.
4. Ensure the repositories are accessible to all users.
    a. Reference and integrate the metadata repositories with other relevant governance and management documentation. The metadata repositories should be made into key reference points in EEOC data documentation and systems. This is critical for utilizing metadata repositories for data discovery: a user should be able to identify data resources even if they do not have access to the data itself.
5. Note: Creating a centralized metadata repository can be performed (1) iteratively in a phased approach, (2) in a more concerted effort if greater resources are allocated, or (3) as an adjustment to existing data review and documentation processes.
    a. These approaches offer flexibility between resource allocation and overall time to completion.
    b. Similar flexibility exists around data selection. Prioritizing data which currently generates the most inquiries, has the most knowledge-based obstacles to utilization, or has some other indication of demand for metadata, will maximize the utility

gained from this recommendation. Alternatively, focusing on data by general usage statistics, beginning with the most widely used data, would allow for maximized coverage with the least effort.

5. Zendesk Transparency

This recommendation relates directly to Data Management Finding 1: Centralized Planning and Governance.

We recommend that OCIO provide users with Zendesk ticket status and prioritization criteria to clarify resolution timelines and improve user satisfaction within ARC processes. Offering real-time status updates on tickets, such as when they are being reviewed, prioritized, or developed, keeps users informed. Clearly communicating the criteria and decision-making process behind ticket prioritization sets appropriate expectations. This information can be included in existing or new governance documentation for ARC. Providing users with more clarity on these aspects would foster a sense of transparency and engagement, ultimately improving user trust and satisfaction.

To implement more transparent processes for Zendesk tickets, OCIO should review the Zendesk configurations for opportunities to provide users with updates during the existing ticket resolution process. A common practice is to pre-populate forms with information on how inputs will be utilized and how they may affect prioritization. This practice may also be paired with a public-facing dashboard showing analytics and aggregate statistics for different inputs.

6. Computing Resources

This recommendation relates directly to Data Management Finding 5: Aggregation Issues.

We recommend the Agency evaluate the costs and benefits of right-sizing its computing platforms to align with user requirements for current and predictable future business priorities to improve operational efficiency. A comparative cost analysis is a useful evaluation tool as this represents a potentially significant financial investment. Current computing power limitations are substantially slowing projects that utilize data. The costs of platform enhancements may be significantly or entirely offset by efficiency improvements of the workforce and improvements to organizational effectiveness. This evaluation should also consider scalability options and limitations, future requirements, and the Agency's broader technology infrastructure to ensure that the chosen platform supports evolving data needs.

This evaluation entails three elements. First, EEOC would determine from ARC and EDW users the scope and extent of impacts from computing resource limitations on their efficiency and effectiveness at executing tasks. This could also include OEDA staff, who use downstream ARC and EDW datasets. Second, EEOC would evaluate the resources necessary to mitigate these impacts at present and with predictable future demand. Resource forecasting should include projections of current trends in resource utilization combined with workload characteristics.

Resource forecasting may be enhanced where feasibly determinable with future application requirements and induced demand effects. Finally, the EEOC would estimate benefits from enhanced user efficiency and effectiveness and the associated computational resource costs.

7. Artificial Intelligence

This recommendation relates directly to Data Management Finding 3: Reduction of Manual Processes.

We recommend the Agency evaluate the costs and benefits of employing Artificial Intelligence (AI) solutions to improve operational efficiencies in data validation, reporting, and analytics. AI system deployment methods require significant model governance, technical capabilities, and resources. With EEOC data sensitivity, using externally hosted AI models is inadvisable; therefore, this recommendation must consider the prior recommendation regarding computational resources. Additionally, the evolving regulatory environment necessitates ongoing monitoring of official government resources, communications, and regulations.

Operational efficiency improvements to consider include knowledge retrieval and data validation automations. Knowledge retrieval can be enhanced using a large language model (LLM)[14] with Retrieval Augmented Generation (RAG)[15] for textual data summarization, accelerating access to key insights with references to the exact source of the information. These tools can assist with reporting and analytics by quickly analyzing data and generating clear, concise summaries, insights, or answers to specific questions. Programmatically applying an LLM to human-entered data can make data validations more impactful. For example, the LLM can identify a specific subset of records that should be manually validated, which can be particularly helpful for validating descriptive plain-text data. Monitoring tools can be applied to detect significant changes to data completeness or other quality measures and reported automatically to accountable parties. The operational benefits need not be limited to these examples; an evaluation of AI solutions could consider identifying additional manual processes throughout the data lifecycle that might be automated to achieve more consistent outcomes and improved operational efficiency.

This evaluation comprises several distinct elements. First, the EEOC would identify opportunities for AI solutions to be implemented that improve efficiency and effectiveness. An optimal approach to gathering this information begins with identifying applicable technologies and use cases. Then, the EEOC would discuss this information with stakeholders and users to identify additional use cases and gather information to evaluate impact on data governance and usage. The Agency should then analyze use cases to understand the impact and minimize the systems necessary to provide coverage. These steps may be performed iteratively; the EEOC can seek user and stakeholder input as additional use cases and systems are identified. Implementation methods may include internal systems, external vendors, and combinations as appropriate.

The regulatory environment around AI is evolving and must also be considered. The most recently issued Executive Order, “Removing Barriers to American Leadership in Artificial Intelligence”

[14] An LLM is trained on vast amounts of text data to understand and generate human-like language.
[15] A RAG system enhances an LLM by incorporating internal information without requiring fine-tuning.

(14179), requires that obstacles to promoting human flourishing, economic competitiveness, and national security be removed. This order was implemented in response to the revoked Executive Order "Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence" (14110). Prior orders, regulations, rulings, and resolutions regarding AI infrastructure, cybersecurity, trustworthiness, and research and development facilitation must be included in this analysis. Future orders, regulations, rulings, and resolutions, including those which are broadly applicable and not limited to AI, must also be considered.

With this information, the EEOC may estimate the necessary resources to plan, develop, launch, and maintain AI solutions to facilitate comparison with operational improvements. Upon synthesizing this information, the EEOC should evaluate the costs and benefits of AI solutions with attention on the regulatory landscape to reach conclusions.

8. ARC Analytics Tools

This recommendation relates directly to Data Management Finding 2: Risk Identification and Mitigation and Data Validity Finding 2: Manual Workarounds.

We recommend that OCIO implement capabilities to analyze ARC data directly, with an analytics tool within ARC or through localized tools with direct connections to ARC, to reduce the creation and use of manual workarounds by users. This would allow users to perform statistical analyses directly within ARC or utilize local computing resources with active connections to ARC data rather than building external tools. Utilizing data directly, as opposed to exporting it, allows maintenance and updates to data without deprecating prior efforts, while centralizing analytics improves efficiency by reducing reproduction of efforts and prevents competing sources of information from arising by maintaining a single source of truth. Centralized analytics tools often include governance capabilities, such as certifying dashboards and data, protecting approved content from further changes, and deploying enterprise analytics apps, all of which further improve data validity and governance.

To best meet user needs, existing stakeholders should be consulted on what additional analytics functions would most benefit their work. This can include surveying stakeholders and users to understand the primary ways in which they rely on external tools for data analytics to best understand how to satisfy their needs within ARC. For analytics capabilities already licensed or available at EEOC, the implementation costs would be limited to labor. If new systems and software need to be licensed, additional costs may need consideration.

9. Data Permissions

This recommendation relates directly to Data Management Finding 2: Risk Identification and Mitigation and Data Validity Finding 2: Manual Workarounds.

We recommend that OCIO and OEDA evaluate the feasibility and operational benefits of implementing more fine-grained access controls within ARC and EDW. Fine-grained access

control is a method of controlling access to resources based on multiple factors such as user role, data type, and specific actions. Role-based access controls, wherein resource access is based solely on roles, are a widely recognized best-practice, but individuals with the same general role may need specialized privileges for different specific tasks. Increasing permission granularity allows individuals to inherit access from a primary role, and gain additional access as needed for secondary roles, in line with the least-privilege principle. The least-privilege principle is an information security concept which maintains that users, applications, and systems should only have access to the minimum resources and permissions necessary to perform their tasks.

6. Appendix A: EEOC Response to Draft Report

U.S. EQUAL EMPLOYMENT OPPORTUNITY COMMISSION
Washington, D.C. 20507

Office of the Chair

May 23, 2025

TO: Joyce Willoughby
Inspector General
Office of Inspector General

FROM: Brett Brenner
Deputy Chief Operating Officer
Office of the Chair

BRETT BRENNER Digitally signed by BRETT BRENNER Date: 2025.05.23 11:36:03 -04'00'

CC: Andrea Lucas
Acting Chair
Office of the Chair

Jiashen You, PhD
Chief Data Officer
Office of Enterprise Data and Analytics

Pierrette McIntire
Chief Information Officer
Office of the Chief Information Officer

Thomas Colclough
National Director
Office of Field Programs

Andrew Rogers
Acting General Counsel
Office of the General Counsel

Subject: **Response to Draft Report for the Evaluation of Data Governance, Management, and Validity (OIG Report No. 2024-002-EOIG)**

Thank you for the opportunity to review the Office of Inspector General's (OIG) draft report titled "Evaluation of Data Governance, Management, and Validity" (OIG Report No. 2024-002EOIG). We appreciate the OIG's efforts in identifying areas for improvement and have carefully considered each recommendation.

Enclosed below is our management response, which outlines our concurrence or nonconcurrence with each recommendation and planned corrective actions where applicable.

Recommendation 1: Data Governance Documentation

Management Response: Currently, OCIO and OEDA provide a wide array of data governance documentation and training sessions to stakeholders across various offices.

Concurrence: Concur.

Corrective Action(s): OCIO and OEDA will collaborate on adding governance information to all ARC and EDW documentation.

Responsible Office(s): OCIO, OEDA

Recommendation 2: Data Governance KPIs

Management Response: OEDA and OCIO are aware of existing data quality and completeness issues, and some tracking is already in place. In response to the recommendation around establishing data governance KPIs, including completeness, compliance, and metadata usage, we can work together to formalize reporting and develop dashboards that track trends over time. These could include metrics like changes in completeness rates for key variables, how often analytic reports are accessed or updated, and usage patterns across tools such as Tableau and R Shiny. This would help establish a more structured and measurable approach to data governance improvement.

Concurrence: Concur.

Corrective Action(s): OEDA and OCIO will collaborate to formalize data governance KPIs and develop dashboards that track data completeness, compliance, and metadata usage trends over time.

Responsible Office(s): OCIO, OEDA

Recommendation 3: Data Governance Board Documentation

Management Response: The EEOC Data Governance Board has been sharing access to its operational documents for transparency, consistency, and accountability with its members in a private group setting since its inception. The industry model chosen by the evaluation team lacks

emphasis on regulatory compliance for an enforcement agency and insufficiently recognizing established government data frameworks such as the Federal Data Strategy. However, DGB Charter could use an update, given the applicable new laws, regulations, and Executive Orders, as well as best practices from the Federal CDO Community. OEDA will lead this update in consultation with OCH.

Concurrence: Non-concur.

Recommendation 4: Centralized Metadata Repository

Management Response: A centralized metadata repository is already in place for our current data infrastructure and remains under active development. This effort was conducted extensively as part of the IMS project and continues to evolve across both the IRD and EDW environments. In our view, the EDW already functions as a centralized metadata repository. Integration between EDW and ARC is not straightforward; however, it may be possible if both EDW and IRD are fully managed in-house. This is something that can be explored further and will continue to be developed and enhanced over time.

Concurrence: Non-concur.

Recommendation 5: Zendesk Transparency

Management Response: The development priorities for the ARC are led by the product owners from the respective Program Offices. Product owners hold the authority for making decisions related to the prioritization of features, delivery timeline, and acceptance criteria. They play a crucial role in ensuring that the development efforts align with the strategic goals and needs of the Agency.

Concurrence: Concur.

Corrective Action(s): OCIO will work with Program Offices on the prioritization criteria and responses.

Responsible Office(s): OCIO

Recommendation 6: Computing Resources

Management Response: ARC applications leverage a modern Kubernetes platform on Azure Cloud services along with PowerBI for data analytics. These services provide a scalable, flexible, and modern enterprise platform. This setup ensures that the applications can efficiently handle varying workloads, support rapid development and deployment, and leverage a wide range of cloud services for enhanced functionality and performance.

From a data analytics perspective, OEDA recognizes OIG's assessment regarding the need for increased compute capacity. For example, the current SAS setup allocates only 4 cores and 4 GB of RAM per session, which is often insufficient for performing complex or time-sensitive analysis. Also, some staff use R and Python locally, but moving these tools to a centralized server environment would significantly improve performance and reduce strain on individual workstations. In many cases, we shift work to the EDW, which has more capacity but was originally designed to serve up to 300 external researchers, and not for EEOC analysts responding to investigative needs or leadership requests on tight timelines. It is possible the current SAS environment was scaled too broadly, resulting in relatively lower per-user resource. Overall, there remains a strong need for a more robust and scalable analytics infrastructure.

Concurrence: Non-concur.

Recommendation 7: Artificial Intelligence

Management Response: OCIO recognizes the potential for AI to improve operational efficiencies. However, implementing AI solutions requires an initial investment in technology and infrastructure, as well as ongoing maintenance and updates. There may also be a need for specialized training for staff.

Concurrence: Concur.

Corrective Action(s): Exploring, pending available resources.

Responsible Office(s): OCIO, (OEDA)

Recommendation 8: ARC Analytics Tools

Management Response: Currently, ARC applications reporting and analytics integrate with multiple central and localized analytical tools, including PowerBI, SAS, and Tableau. These tools are used on both ARC and IRD data, allowing for comprehensive data analysis and visualization. Azure SQL Database allows for the storage and analysis of large volumes of data, while PowerBI offers interactive visualizations and business intelligence capabilities. This combination ensures that the ARC applications have powerful, scalable, and user-friendly reporting and analytics capabilities.

Additionally, it is important to recognize that users will have specific needs to export data and work locally. By providing the capability to export data, users can perform more detailed and customized analyses using their preferred tools and environments. This flexibility will further enhance user satisfaction and support a wide range of analytical needs.

Concurrence: Non-concur.

Recommendation 9: Data Permissions

Management Response: This evaluation is ongoing. Fine-grained access controls already exist within IRD concerning data. EDW already tracks all user activities, such as login and read access to data, while import and export requests are tightly controlled and monitored.

Concurrence: concur.

Corrective Action(s): OCIO and OEDA will continue to evaluate the feasibility and operational benefits of fine-grained access controls.

Responsible Office(s): OCIO, OEDA

7. Appendix B: OIG Analysis of the Agency's Response

The Agency provided the OIG its formal response to the recommendations on May 23, 2025. The Agency's formal response is included in Appendix A to this report. The OIG's analysis of the Agency's response, in consultation with Elder Research, is as follows:

Recommendation 1: Data Governance Documentation

We recommend OCIO and OEDA add governance information to all ARC and EDW documentation to ensure that users are well-informed and have access to crucial data governance details.

Agency Response: EEOC Management concurred with this recommendation. The Agency stated that OCIO and OEDA provide a wide array of data governance documentation and training sessions to stakeholders across various offices. The Agency also stated that OCIO and OEDA will collaborate on adding governance information to all ARC and EDW documentation.

OIG Analysis: The Agency's proposed corrective actions are responsive to this recommendation.

Recommendation 2: Data Governance KPIs

To improve the Agency's direct and quantitative assessment of ongoing data governance improvement efforts, we recommend that OCIO and OEDA collect and analyze additional data governance KPIs for data completeness and compliance, as well as KPIs for metadata coverage, usage, and standardization.

Agency Response: EEOC Management concurred with this recommendation. The Agency stated that OEDA and OCIO are aware of existing data quality and completeness issues, and some tracking is already in place. In response to the recommendation around establishing data governance KPIs, including completeness, compliance, and metadata usage, the Agency can work together to formalize reporting and develop dashboards that track trends over time. These could include metrics like changes in completeness rates for key variables, how often analytic reports are accessed or updated, and usage patterns across tools such as Tableau and R Shiny. This would help establish a more structured and measurable approach to data governance improvement. To address the recommendation, the Agency stated that OEDA and OCIO will collaborate to formalize data governance KPIs and develop dashboards that track data completeness, compliance, and metadata usage trends over time.

OIG Analysis: The Agency's proposed corrective actions are responsive to this recommendation.

Recommendation 3: Data Governance Board Documentation

We recommend that the Data Governance Board (DGB) establish written procedures for its operations to enhance transparency, consistency, and accountability in its operations.

Agency Response: EEOC Management did **not** concur with this recommendation. The Agency stated that the EEOC DGB has been sharing access to its operational documents for transparency, consistency, and accountability with its members in a private group setting since its inception. The industry model chosen by the evaluation team lacks emphasis on regulatory compliance for an enforcement agency and insufficiently recognizing established government data frameworks such as the Federal Data Strategy. However, the Agency stated that the DGB Charter could use an update, given the applicable new laws, regulations, and Executive Orders, as well as best practices from the Federal CDO Community. OEDA will lead this update in consultation with OCH.

OIG Analysis: An update to the DGB, as described in the EEOC management response, will be responsive to much of the recommendation. In addition, the recommendation would be implemented in full if the DGB update includes appropriate transparency policy so that Agency staff outside the DGB could identify the source of, and reason for, new data governance policy.

Recommendation 4: Centralized Metadata Repository

We recommend that OCIO and OEDA create centralized and integrated metadata repositories for the ARC and EDW to ensure consistent metadata accessibility and usage.

Agency Response: EEOC Management did **not** concur with this recommendation. The Agency stated that a centralized metadata repository is already in place for its current data infrastructure and remains under active development. This effort was conducted extensively as part of the IMS project and continues to evolve across both the IRD and EDW environments. The Agency believes that the EDW already functions as a centralized metadata repository. Integration between EDW and ARC is not straightforward; however, it may be possible if both EDW and IRD are fully managed in-house. The Agency response stated that this can be explored further and will continue to be developed and enhanced over time.

OIG Analysis: While metadata can certainly be located in a warehouse(s), broad user accessibility to the metadata is the key aspect that makes the warehouse valuable and is the focus of the recommendation. As stated in the final report, "A metadata repository is a centralized system that stores metadata, including its structure, content, business rules, transformation processes, and integration details. When integrated directly with the system it supports, this repository becomes easily accessible to the users when needed." The repositories should be accessible by everyone in the agency, **regardless of whether they have access to the data described by the metadata**. Providing users metadata access enables them to identify needed data and prevent attempts to

reproduce, recalculate, infer, or otherwise obtain this information. OIG will work with the Agency to ensure the intent of the recommendation is implemented.

Recommendation 5: Zendesk Transparency

We recommend that OCIO provide users with Zendesk ticket status and prioritization criteria to clarify resolution timelines and improve user satisfaction within ARC processes.

Agency Response: EEOC Management concurred with this recommendation. The Agency stated that the development priorities for the ARC are led by the product owners from the respective program offices. Product owners hold the authority for making decisions related to the prioritization of features, delivery timeline, and acceptance criteria. They play a crucial role in ensuring that the development efforts align with the strategic goals and needs of the Agency. To address this recommendation, OCIO will work with program offices on the prioritization criteria and responses.

OIG Analysis: The Agency's proposed corrective actions are responsive to this recommendation.

Recommendation 6: Computing Resources

We recommend the Agency evaluate the costs and benefits of right-sizing its computing platforms to align with user requirements for current and predictable future business priorities to improve operational efficiency.

Agency Response: EEOC Management did **not** concur with this recommendation. The Agency stated that ARC applications leverage a modern Kubernetes platform on Azure Cloud services along with PowerBI for data analytics. These services provide a scalable, flexible, and modern enterprise platform. This setup ensures that the applications can efficiently handle varying workloads, support rapid development and deployment, and leverage a wide range of cloud services for enhanced functionality and performance.

EEOC Management also stated that from a data analytics perspective, OEDA recognizes OIG's assessment regarding the need for increased compute capacity. For example, the current SAS setup allocates only 4 cores and 4 GB of RAM per session, which is often insufficient for performing complex or time-sensitive analysis. The Agency also recognized that, while some staff use R and Python locally, moving these tools to a centralized server environment would significantly improve performance and reduce strain on individual workstations. In many cases, the Agency shifts work to the EDW, which has more capacity but was originally designed to serve up to 300 external researchers, and not for EEOC analysts responding to investigative needs or leadership requests on tight timelines. It is possible the current SAS environment was scaled too broadly, resulting in relatively lower per-user resource. Overall, there remains a strong need for a more robust and scalable analytics infrastructure.

OIG Analysis: The OIG edited the recommendation for clarification. The Agency's analysis of the issues (i.e., the need for increased computing capacity) and proposed corrective action (i.e., moving tools to a centralized server) align with OIG's recommendation. If implemented, right-sizing the computing platforms to user requirements will improve operational efficiencies.

Recommendation 7: Artificial Intelligence

We recommend the Agency evaluate the costs and benefits of employing Artificial Intelligence (AI) solutions to improve operational efficiencies in data validation, reporting, and analytics.

Agency Response: EEOC Management concurred with this recommendation. The Agency stated that OCIO recognizes the potential for AI to improve operational efficiencies. However, implementing AI solutions requires an initial investment in technology and infrastructure, as well as ongoing maintenance and updates. There may also be a need for specialized training for staff.

OIG Analysis: The Agency's proposed corrective actions are responsive to this recommendation.

Recommendation 8: ARC Analytics Tools

We recommend that OCIO implement capabilities to analyze ARC data directly, with an analytics tool within the ARC or through localized tools with direct connections to the ARC, to reduce the creation and use of manual workarounds by users.

Agency Response: The EEOC Management did not concur with this recommendation. The Agency stated that ARC applications reporting and analytics integrate with multiple central and localized analytical tools, including PowerBI, SAS, and Tableau. These tools are used on both ARC and IRD data, allowing for comprehensive data analysis and visualization. Azure SQL. Database allows for the storage and analysis of large volumes of data, while PowerBI offers interactive visualizations and business intelligence capabilities. This combination ensures that the ARC applications have powerful, scalable, and user-friendly reporting and analytics capabilities.

Additionally, the Agency stated it is important to recognize that users will have specific needs to export data and work locally. By providing the capability to export data, users can perform more detailed and customized analyses using their preferred tools and environments. This flexibility will further enhance user satisfaction and support a wide range of analytical needs.

OIG Analysis: While it can be productive to perform additional data analysis using downstream tools, manual workarounds pose risks and may lead to replication of efforts and multiple sources of the same information. Sharing applications across the Agency will prevent additional effort being used to recreate the same result by a different area of the organization. To reduce these risks, where feasible, the Agency should move upstream in EEOC systems calculated and/or derived metrics for direct data files used to conduct agency business. We believe this will ensure consistent application of the method in a materialized view, enabling the Agency to use one definition for

each metric, preserving the integrity and validity of data. OIG will work with the Agency to ensure the intent of the recommendation is implemented.

Recommendation 9: Data Permissions

We recommend that OCIO and OEDA evaluate the feasibility and operational benefits of implementing more fine-grained access controls within the ARC and EDW.

Agency Response: EEOC Management concurred with this recommendation. The Agency stated fine-grained access controls already exist within IRD concerning data and that EDW already tracks all user activities, such as login and read access to data, while import and export requests are tightly controlled and monitored. To address the recommendation, the Agency stated that OCIO and OEDA will continue to evaluate the feasibility and operational benefits of fine-grained access controls.

OIG Analysis: The Agency's proposed corrective actions are responsive to this recommendation.