



OPEN

DATA DESCRIPTOR

Integrated Network Solutions in Government Hiring Trends (INSIGHT+)

William G. Resh¹✉, Keunyoung Eli Lee²✉, Yi Ming^{1,3}, Xinyao Andy Xia⁴, Nicole Dias⁴, Kecheng Anderson Liu⁴, Darren Cao⁴ & William Huh⁵

The Integrated Network Solutions in Government Hiring Trends (INSIGHT+) database supports research on the U.S. federal civil service labor market. As of September 2024, the federal workforce included over 2.4 million civilian employees spanning more than 400 occupations, with over 85% located outside the Washington, D.C. metropolitan area. INSIGHT+ integrates micro-, meso-, and macro-level statistics from sources ranging from governmental and academic sources. Our database covers workforce dynamics from 2018 to 2023. It allows granular multivariate analyses and accommodates agency-, location-, and institution-specific tables that can be matched in the future with further detailed data on agency outputs such as discretionary grants, contracts, contingent liabilities, and various economic impacts. This article outlines the current capabilities and significance of INSIGHT+ for civil service labor market research, emphasizing ongoing enhancements to enrich analyses of civil service institutions.

Background & Summary

Introduction. The U.S. federal civilian workforce, comprising approximately 2 million employees across over 400 occupations¹, has become a focal point of intense political debate. With over 80% of civil service positions located outside the Washington, D.C. metropolitan area, the federal government plays a significant role in labor markets nationwide, accounting for 85,000 new jobs in 2023 alone². Recent events have dramatically reshaped the federal workforce landscape that highlight the importance of understanding the complex dynamics of the federal civil service labor market and the relative impacts of administrative reform.

The Trump administration has initiated a substantial reduction in the federal workforce, with agencies instructed to formulate strategies for eliminating employee roles by March 2025³. This initiative has led to a significant restructuring of the U.S. government. In response to substantial pushback from litigation and court rulings challenging their initial abrupt workforce reductions, the administration adopted alternative strategies. By implementing paid departures and temporarily sidelining employees in eliminated positions, the administration rapidly executed significant workforce reductions: roughly 124,000 federal employees opted for deferred resignation with continued compensation through September⁴, while numerous others have been placed on temporary leave as the reductions in force (RIFs) progress⁵.

These dramatic shifts underscore the importance of comprehensive datasets and analytic tools for understanding and analyzing the ongoing changes in the U.S. federal civil service labor market. Such resources can help us understand how these rapid changes impact the federal workforce, assess regional economic impacts, and inform evidence-based policy decisions.

In this article, we describe the Integrated Network Solutions in Government Hiring Trends (INSIGHT+) database as a new resource intended to support research on the U.S. federal civil service labor market. We describe its capabilities, its relevance to important questions in political science and policy research, and ongoing enhancements to ensure its enduring value for stakeholders interested in the unique features and constraints of the U. S. federal civil service labor market. Unlike the private sector, where market forces predominantly dictate labor dynamics, the federal civil service labor market is uniquely constrained by government policies, political decisions, and statutory regulations. These factors create distinct characteristics in both the supply of

¹Civic Leadership Education and Research (CLEAR) Initiative, Andrew Young School of Policy Studies, Georgia State University, Atlanta, GA, USA. ²American University, Washington, DC, USA. ³Georgia Institute of Technology, Atlanta, GA, USA. ⁴University of Southern California, Los Angeles, USA. ⁵Pomona College, Claremont, USA. ✉e-mail: wresh@gsu.edu; eli@american.edu

labor (individuals willing and able to work for the federal government) and the demand for labor (the number and types of positions the federal government needs to fill).

One of the primary factors that makes the federal civil service labor market unique is the government's monopoly over certain services and functions. The federal government holds exclusive responsibility for critical areas such as national defense, foreign diplomacy, monetary policy, and federal law enforcement⁶. This monopoly creates a unique and inelastic demand for specialized labor that cannot be supplied by the private sector. For instance, the Department of Defense requires a workforce with security clearances and specialized training not typically found in the civilian labor market⁷.

Moreover, the demand for federal labor is politically determined. The federal budget, which sets funding levels for agencies and programs, is established through a political process involving both the Executive Branch and Congress. Political priorities and negotiations influence appropriations bills, directly affecting staffing levels across federal agencies⁸. Legislative mandates can create new agencies or expand programs, increasing labor demand, while efforts to reduce government size or eliminate programs can decrease it⁹.

Fiscal policies and economic conditions also play a significant role. The federal government may increase hiring during economic downturns as part of fiscal stimulus measures, thereby acting counter-cyclically compared to the private sector¹⁰. Conversely, concerns over national debt and budget deficits can lead to hiring freezes or workforce reductions, affecting labor demand independent of market conditions¹¹.

On the supply side, the federal civil service offers attributes that both attract and limit the pool of potential employees. Job security and comprehensive benefits, including robust retirement plans like the Federal Employees Retirement System (FERS), make federal employment attractive to individuals seeking stability¹². Those same incentives can be disrupted by political upheaval and instability. At the same time, the federal hiring process presents barriers to entry that can limit the supply of eligible candidates. The merit-based system requires applicants to meet stringent qualifications, undergo extensive background checks, and often obtain security clearances¹³. These requirements can deter potential applicants who may find more accessible opportunities in the private sector.

Wage structures in the federal civil service further impact labor supply. The General Schedule (GS) pay scale establishes fixed salaries that may not keep pace with private sector wages, especially in high-demand fields such as technology or healthcare¹⁴. The inability to negotiate salaries beyond established pay grades can dissuade highly skilled workers who might command higher compensation elsewhere.

Importantly, geographic distribution also affects labor supply. A significant proportion of federal positions are concentrated in specific regions, such as the Washington, D.C., area or locations with military installations. These locations can also constrain supply by the concentrated demand for talent required by specialist agencies with operations in particular (and often remote) locations chosen both in terms of political and operational pressures or exigencies. This concentration can limit the local labor pool and necessitate relocation, which not all potential applicants are willing or able to undertake¹⁵.

The political economy within which the federal civil service operates profoundly influences both supply and demand. Political ideologies regarding the size and role of government directly impact labor demand. Administrations favoring an expanded governmental role may increase federal employment, while those advocating for limited government may reduce it¹⁶. Policy priorities shift labor demand by focusing on areas such as cybersecurity, healthcare, or environmental protection (as well as most recently artificial intelligence¹⁷), creating surges in demand for specialized labor^{18,19}.

Legislative and regulatory constraints further shape the labor market. Political impasses can result in hiring freezes or budget sequestration, abruptly altering labor demand regardless of agency needs²⁰. Civil service protections make it challenging to adjust the workforce quickly in response to changing needs, leading to a less flexible labor market compared to the private sector²¹.

Unionization in the federal civil service differs markedly from the private sector. Federal employee unions have limited collective bargaining rights, particularly concerning wages and benefits, due to legal restrictions²². The prohibition of strikes removes a key bargaining tool, potentially affecting the attractiveness of federal employment for workers seeking strong labor representation²³.

Public perception and trust in government institutions also influence labor supply. Negative perceptions or political rhetoric criticizing federal employees may reduce the pool of applicants²⁴. Conversely, a strong motivation for public service can attract individuals to federal employment despite potential drawbacks, affecting the labor supply uniquely compared to the private sector²⁵.

The federal civil service labor market also exhibits wage rigidity relative to general occupational labor markets due to standardized compensation structures, leading to labor shortages in high-demand occupations where private sector wages are significantly higher¹⁴. The government's slower pace in adjusting compensation or creating new positions hinders its ability to respond to rapidly changing labor market conditions, such as technological advancements requiring new skill sets.

An oversupply of applicants may occur in positions where federal compensation is competitive or superior to the private sector, leading to increased competition for those roles. Policies such as veteran preference in hiring can influence the composition of the applicant pool, potentially limiting opportunities for non-veterans and affecting overall supply dynamics²⁶.

Technological change and automation impact the federal labor market differently than in the private sector. The federal government's procurement and adoption processes for new technologies can be slower due to regulatory requirements, affecting demand for certain labor skills²⁷. Additionally, budgetary and political constraints may limit investment in workforce reskilling, impacting the supply of qualified internal candidates for evolving job requirements¹⁸.

For all of the above reasons, we argue that civil service labor markets are uniquely constrained relative to the general occupational labor markets with which they overlap. As well, the allocation of expertise across agencies

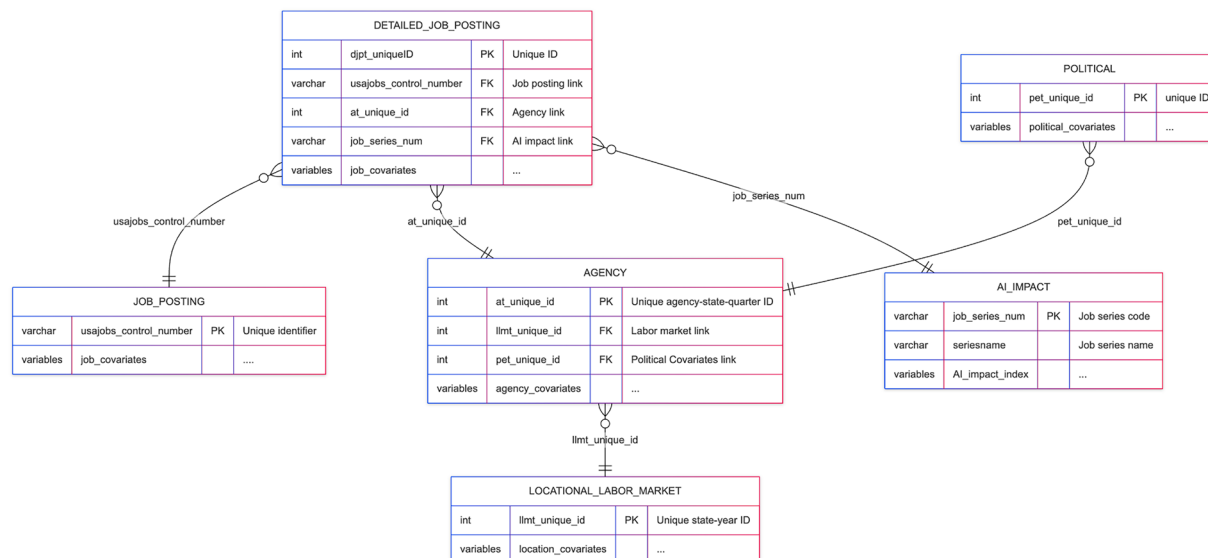


Fig. 1 Relational Diagram: how data files (tables) are connected with each other.

and by location is political. Just as other government allocations are negotiated through distributive politics, so too can agency employment be, to some extent, beyond the operational needs of government. Regional agency headquarters, VA hospitals, military bases, and the like can have substantial economic impacts on a community in which they are located.

In the following, we provide a data infrastructure that can be replicated across large, general-purpose governments. As of April 2024, government payrolls accounted for 14.7% of the total payrolls in the economy, with local governments employing the majority (64%), followed by state governments (23.4%) and the federal government (12.9%)²⁸. Across the board, governments are facing an unprecedented hiring crisis in competition for new talent²⁹. Despite the clear demand, there exists a significant shortfall in the availability of comprehensive data needed to effectively understand and manage the public sector labor market's supply and demand dynamics. Thus, this project is designed to integrate, develop, and create variables from the scattered data sources across both professional practice in the public sector and the academic fields of public administration, labor economics, and political science that can create utility to the multiple stakeholders who study or operate within these market dynamics.

Our project is designed to match relevant data across multiple levels of analysis, including job-specific, bureau, geographic, and market dimensions. Furthermore, it works *retrospectively* in terms of tracking and collecting observations over time and is designed for *prospective* analytics by incorporating a proprietary data engine that forecasts workforce development using generative artificial intelligence (AI), enabling proactive responses to evolving labor market trends.

Additionally, the project provides a data infrastructure that enables various stakeholders—including policy-makers, hiring authorities, educational institutions, and social science researchers—to access, analyze, and visualize this labor market data. This structured and collaborative environment fosters innovative research, supports informed strategic decisions in public-sector employment, and improves the alignment of educational programs with real-world labor market demands. By simplifying data collection and integration, the tools allow both domain experts and non-specialists to identify patterns and derive actionable insights, ultimately enhancing public administration effectiveness and efficiency.

Method — Database Development

The INSIGHT + database is designed to serve multiple stakeholders' needs. It utilizes an extensive “Extract, Transform, and Load (ETL)” pipeline to aggregate and refine disparate data streams. Sources such as public sector hiring data from diverse platforms (e.g., Monster, USAJobs.gov, USA Staffing), the U.S. Office of Personnel Management (US OPM) FedScope Cube, Administrative Conference of the United States (ACUS)'s Federal Administrative Procedure Sourcebook, and scholarly research on agency ideology, reputation, independence, and partisan disagreement measures^{30–34} are methodically extracted, transformed to align with uniform metrics, and loaded into the comprehensive INSIGHT + database. It is also complemented by macroeconomic statistics from the private sector labor market, such as occupational employment and wage statistics from the U.S. Bureau of Labor Statistics (U.S. Bureau of Labor Statistics OEWs).

Our database organizes different dimensions of data into separate tables with predefined relationships³⁵ (Fig. 1). It comprises six interconnected files (data tables): Job Posting, Detailed Job Posting, Agency, Locational Labor Market, and AI Occupational Impact tables. Each data file (table) has rows (records) and columns (fields) identified by a unique ID column, also known as primary keys in relational database management system (RDBMS) terminology, and data files are linked together through foreign keys, which reference primary keys in other data files to maintain relationships and ensure data integrity. The Job Posting Table uses the USAJobs Control Number

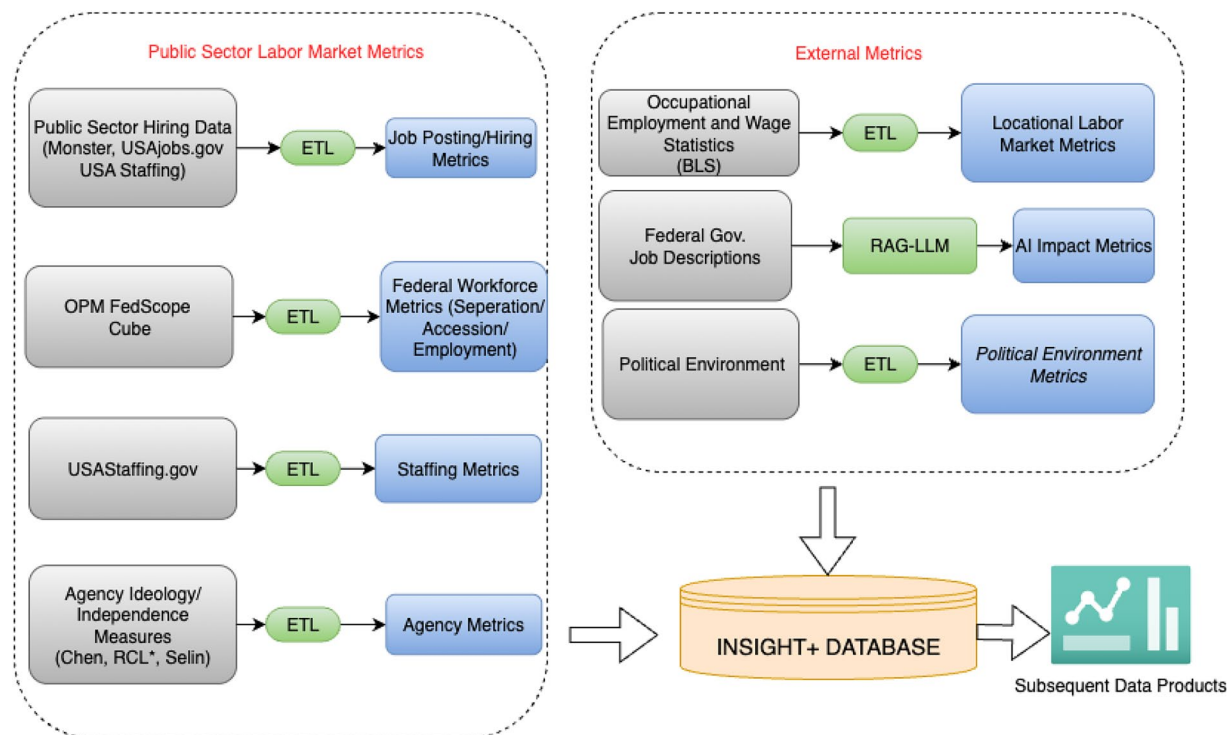


Fig. 2 The Major Components of INSIGHT+.

as the primary key (PK). The Detailed Job Posting Table expands postings by specific locations and job series, linking to Job Posting (USAJobs Control Number as foreign key (FK)), Agency (“AT_Unique_ID” as FK), and AI Impact (Job Series Number as FK). The Agency Table aggregates agency-level data by agency/state/quarter, linked to the Locational Labor Market Table (“LLMT_Unique_ID” as FK) and to the Political Environment Table (Election Cycle (Year)–State as FK). The Locational Labor Market Table contains socioeconomic employment data (“LLMT_Unique_ID” as PK).

The primary data file in our database is “Job Posting Table”. It represents micro-level data, with attributes such as pay grade and occupation (more details will be explained in the “Data Records” section). Moreover, job postings and agencies are located within states, where occupational salary levels and the total number of employees shape hiring dynamics. The structure of our database integrates these interdependent data levels to support the analysis of public sector employment patterns.

The INSIGHT database consists of six primary data files (tables) that represent different dimensions of public sector labor markets. Each table serves a unique purpose, contributing to the integration and analysis of data. The Job Posting Table serves as the foundational table. It contains job postings, where each row corresponds to a unique job posting. The unique key ID, “USAJobs Control Number”, links each entry to its corresponding job posting on the USAJobs website. This table provides the essential structure for connecting additional data dimensions, such as job characteristics and agency information, while maintaining the integrity of the source data. Furthermore, when merging across tables, users should rely on the unique IDs across data files (tables) to identify the corresponding records from other tables. Each data file has its own unique ID column, also known as, primary key, which uniquely identifies rows in that file, and foreign keys that reference primary keys in related data files. For example, to merge the Job Posting Table with the Detailed Job Posting Table, researchers should use the column USAJOBS Control Number, which serves as the primary key in the Job Posting Table and as a foreign key in the Detailed Job Posting Table. Similarly, to merge the Agency Table with the Detailed Job Posting Table, the column AT Unique ID functions as the primary key in the Agency Table and as a foreign key in the Detailed Job Posting Table.

The Detailed Job Postings Table focuses on job information categorized by location and job series. This table originates from the Job Posting Table but has been expanded to address the complexity of job postings on the web, where multiple job series and locations are often bundled in a single entry. To enable detailed analysis, the dataset has been separated by job series and location. This approach allows for a more granular understanding of which jobs are available in specific locations and occupational categories.

The Agency Table contains detailed information about federal agencies. Here, federal agencies include sub-agency levels within cabinet departments and large independent agencies. Key variables include ideological orientation^{30,32}, external reputation³³, structural and political independence³¹, partisan alignment³⁴, and other structural characteristics³⁶. As well, the Agency Table includes temporal workforce dynamics such as separations, accessions, and staffing characteristics that are pulled from the United States Office of Personnel Management’s FedScope. The Locational Labor Market Table draws from the U.S. Occupational Employment and Wage Statistics (OEWS) program³⁷. This table includes data on average annual/hourly wages, employment

counts, and occupational distributions by state, enabling insights into the locational dynamics that shape the public sector labor market. The Political Environment Table captures partisan and representation dynamics at both a state and national level over time. The AI Impact Table categorizes generative AI impacts (complementarity, augmentation, substitutivity) across job knowledge, skills, and abilities, linked by Job Series Number to the Detailed Job Posting Table. This relational structure works well for connecting multi-level data and supports seamless queries and analysis across interconnected dimensions³⁸.

All data files then undergo a rigorous quality check (see technical validation part) before being stored in the database. This ensures not only data accuracy but also the flexibility required to adapt as the scope of the research evolves. The figure below shows the ecosystem of our data infrastructure and shows how our facilities serve our stakeholders.

Figure 2 illustrates the overall architecture of the INSIGHT + database system. The framework is divided into two major domains: *Public Sector Labor Market Metrics* and *External Metrics*. Within the public sector domain, data from platforms such as USAJobs.gov, USA Staffing, Richardson, Clinton, and Lewis (RCL)'s agency ideology score, and US OPM FedScope are processed through ETL (Extract, Transform, Load) pipelines to produce structured metrics on job postings, workforce dynamics, staffing, and agency characteristics. The external domain incorporates additional labor market indicators from sources like the Bureau of Labor Statistics and political datasets, as well as semantic insights derived from federal job descriptions using an RAG-LLM (Retrieval-Augmented Generation with Large Language Models) approach^{39,40}.

Development of tables across multiple dimensions. In developing our database, we identified the similarities and differences between various data sources, diagrammed their comparability, and unified and standardized their formats. This process is known as Extract, Transform, and Load (ETL), which involves extracting data from diverse sources, transforming it into a usable and trusted format, and then loading it into systems that end-users can access for further analysis and decision-making⁴¹.

As data volumes, types, and sources continue to grow, the ETL process becomes important to ensure that data is not only accurate and up-to-date but also reliable and ready for analytical and machine-learning initiatives. Furthermore, ETL is essential in sustaining long-term viability as ETL ensures the continual availability of current and accurate information and tackles traditional data silos effectively⁴². This is relevant in our context, where ETL facilitates connections across various data sources, from the micro-level of individual agencies to the meso-level of interagency interactions that encompass political environments, up to the macro-level of intersectoral analysis that reveals the competitiveness of public sector jobs. In this section, we provide a detailed description of each table and its constituent variables.

Job posting table. The first table draws from the Federal Hiring Assessments and Selection Outcome Dataset, developed by the General Services Administration⁴³. This dataset integrates data from multiple sources, including USAJOBS.gov (managed by the U.S. Office of Personnel Management), Monster, and the Federal Aviation Administration (FAA) Hiring Assessment and Selection Outcome Data. It encompasses job posting records for the 24 CFO Act agencies and spans the years 2018 to 2023. This six-year dataset provides a detailed snapshot of public sector recruitment patterns across distinct periods.

The dataset records information related to public sector job postings. At the core of this table, the variable *usajobs control number* acts as the primary key. This ensures a unique identifier for each job posting. The dataset organizes variables into seven categories to structure insights into different aspects of the recruitment process. Below is a summary of the variables in each category.

The first category, Announcement Information, includes variables such as *date open*, which records the date the job announcement was opened, and *announcement type*, which defines the type or category of the announcement. It also contains *announcement open date* and *announcement close date*, which indicate the exact opening and closing dates, and *announcement open date fy* and *announcement close date fy*, which correspond to the fiscal years of these dates.

The second category, Position Details, captures data related to the specific job position. Variables in this category include *pay_plan*, which indicates the pay structure, *job series title*, which denotes the title linked to the *job series number*, and *grade*, which defines the grade level of the position. Additionally, *vacancy job title* specifies the official title of the vacant position, *vacancy announcement types* describes the types of vacancy announcements, and *service type* identifies the type of service, such as competitive or excepted.

The third category, Application Process, focuses on details regarding applicant eligibility and application limits. Variables such as *application limit dummy* and *application limit num* specify if there is a cap on the number of applications and what that limit is. *Applicant eligibility public (dummy)* outlines criteria for public eligibility, while *applicant eligibility internal (dummy)* and *applicant eligibility status (dummy)* describe eligibility for internal candidates and based on applicant status. Other variables include *applicant_eligibility usajobs (dummy)*, which reflects eligibility criteria from USAJobs.gov, and *vacancy eligibility list*, detailing eligibility conditions.

The fourth category, Assessment and Selection, contains information on the methods used for assessing candidates and the selection outcomes. Variables include *assessment used (dummy)*, indicating whether assessments were applied, *off the shelf assessment (dummy)*, which records the use of standardized assessments, and *assessment questionnaires used (dummy)*, specifying the use of assessment questionnaires. Other variables, such as *other assessment type used* and *additional manual assessment use*, detail the types of assessments applied. The category also includes *selection (dummy)* and *selection number*, which indicate if selections were made from the announcement and the number of such selections, respectively. *Scoring option* records the scoring methods applied in assessments.

The fifth category, Appointment Details, addresses variables such as source, which identifies the source of the announcement or candidate, *appointment type*, which specifies the type of appointment (e.g., permanent or temporary), and *appointment type list* and *appointment type groupings*, which detail and group appointment types. Additionally, *appointing authority list* records the legal authorities under which appointments are made.

The sixth category, Work Flexibility, captures data on job flexibility options. Variables include *location flexibility*, indicating flexibility (both remote and telework) in job location, and *remote work available* and *telework available*, which specify the availability of remote and telework options.

The seventh category, Applicant Information, contains metrics on applicant demographics and application outcomes (currently limited to data from the US OPM). Variables include applicant race information, capturing data on racial groups such as *White*, *Two or More Races*, *Asian*, *Black or African American*, *Hispanic or Latino*, *American Indian or Alaskan Native*, *Hawaiian or Pacific Islander*, and applicants with *unspecified ethnicity*. Additional metrics include *total applications*, *veteran applicants*, *eligible applications*, *total completed application*, and *total selectees*. This category also includes gender-based information, covering the number of *female* and *male* applicants, as well as those with *unspecified gender*. We plan to request additional data for other agencies through the Freedom of Information Act (FOIA).

Detailed job posting table. The Detailed Job Posting Table builds upon the Job Posting Table. In the original Job Posting Table, each job posting is identified by the job posting's unique ID ("*usajobs control number*"), which serves as a unique identifier. However, this variable cannot function as a primary key in the expanded dataset because each job posting often includes multiple announcement locations (e.g., Chicago, IL; Denver, CO) and multiple job series (e.g., 3100; 3102) listed together in a single row. To address this, the detailed job posting table separates the data so that each row represents a specific combination of announcement location and job series associated with a job posting. This expanded structure allows for more in-depth analysis. By transforming each unique location and job series combination as a distinct row, the table allows researchers to examine patterns such as which agencies are posting specific job series in different states. For instance, it becomes possible to analyze whether certain agencies are more active in particular geographic regions or whether specific job series are concentrated in certain states. This level of detail provides valuable insights into the spatial and organizational distribution of federal job postings. Moreover, it allows researchers to assess recruitment strategies and workforce needs at a granular level.

In developing our database based on this expansion, we created a unique ID for the Detailed Job Posting Table ("DJPT_Unique_ID"), which acts as the primary key of the table. Furthermore, we created additional variables based on job series and location. A binary dummy variable, *Health*, identifies health-related occupations within specific job families. Another binary variable, *STEM*, marks positions related to Science, Technology, Engineering, and Mathematics, using defined keywords and occupational series extracted from job titles and job series numbers.

We also included the *FIPS Code*, a numeric variable that corresponds to State, County, and Place identifiers, enabling precise geographic analysis (U.S. Census Bureau, 2020). In addition, we created two variables named *Regional Federal Executive Boards (RFEB)* and *Metropolitan Federal Executive Boards (MFEB)*, based on categorizations provided by the U.S. Office of Personnel Management (OPM). The Regional FEB variable divides the United States into four distinct regions: Central, Eastern, Southern, and Western. Within each Regional FEB, the Metropolitan FEB variable identifies specific metropolitan areas. While functional Federal Executive Boards have since been eliminated by the current Trump administration, the designations continue to provide coherent regional and metropolitan parameters for useful analysis.

These variables, structured to align with the specifics of agency-level data, support detailed analysis of employment trends and workforce requirements in the public sector. To facilitate this connection, we created an agency unique ID ("AT_Unique_ID") by grouping data by *agency*, *state*, and *year*. This unique ID serves as a foreign key, linking the Detailed Job Posting Table to the Agency Table. This linkage enables the integration of temporally and locationally variant agency-level traits, such as political ideology, independence, and locational workforce statistics, with the job posting data. This relational structure supports in-depth analysis of how agency-specific factors influence recruitment strategies and workforce distribution.

(Federal) Agency Table. The Agency Table captures detailed information on federal agencies, which in this paper and data refers to subagency-level entities within federal Cabinet departments or independent agencies. This table includes data on workforce statistics, such as separations, accessions, and employment, along with measures of agency ideology, independence, reputation, and structural and procedural characteristics of federal agencies. By incorporating political and administrative traits from various scholarly and institutional sources, the Agency Table provides a comprehensive view of the organizational dynamics and governance structures within the federal bureaucracy.

First, we incorporate workforce statistics from the U.S. Office of Personnel Management's FedScope (2023), focusing on separations, accessions, and employment data at the state level. These datasets provide insights into workforce dynamics within federal agencies and their geographic and temporal variations. The separation dataset captures state-level data on employees leaving federal service on a quarterly basis. This dataset reflects trends in voluntary resignations, retirements, and other types of separations and offers a snapshot of workforce turnover and potential challenges in retention. The accession dataset records state-level data on employees entering federal service on a quarterly basis. It highlights recruitment trends such as new hires and interagency transfers and reveals patterns in workforce expansion or redistribution. The employment dataset provides state-level data on the total number of employees in federal service but is recorded on a yearly basis due to minimal fluctuations within shorter timeframes. This dataset establishes a foundation for workforce size and distribution across states and acts as a baseline to analyze trends in hiring and separations.

Second, we also incorporate measures developed by scholars in public administration and political science to examine the political environment surrounding federal agencies. *Chen's Agency Ideological Score* uses the Poole and Rosenthal Nominate Common Space scale to estimate the ideological leanings of 74 U.S. federal agencies. This measure analyzes campaign contributions from bureaucratic employees and provides a systematic view of agency ideology (where negative values represent more liberal orientations, while positive values represent more conservative orientations)³⁰. Seli's *Agency Independence Score* quantifies structural independence through a dataset that includes 50 structural characteristics of 321 federal agencies. A Bayesian latent variable model calculates this score by assessing the appointment limitations of key decision-makers and the extent of political oversight in agency policy decisions³¹. Richardson, Clinton, and Lewis's team expand the scope of ideological assessments with a survey of federal executives. Their *Agency Ideological Score* doubles the number of agencies analyzed compared to earlier studies and aligns closely with existing measures³². Bellodi's *U.S. Federal Agency reputation measure* introduces a method to assess bureaucratic reputation through natural language processing of congressional speeches³³. This measure evaluates how lawmakers perceive agencies by calculating cosine similarity between agency embeddings and a positivity vector. However, Bellodi's dataset ends in 2016, and our research covers 2017 to 2023. Thus, we expand and create a new dataset that follows Bellodi's method by using large language models to analyze legislative speeches from 2017 to the most recent data available in both chambers of the U.S. Congress. Finally, *Richardson's measure of partisan distance* captures the magnitude of partisan disagreement faced by 93 agencies. This measure illustrates variation in partisan conflict and highlights agency interactions within a divided political environment³⁴.

Lastly, we also include data from the Administrative Conference of the United States (ACUS) Sourcebook of Federal Administrative Agencies. This sourcebook provides detailed information on the structural and procedural characteristics of federal agencies, offering a foundational resource for understanding administrative governance⁴⁴. It covers key elements such as the presence of *multi member leadership*, *quorum requirements*, and *term lengths for decision-making authorities*. The sourcebook also identifies whether agencies have independent *litigating authority*, whether they operate under budgetary review by the Office of Management and Budget, and whether they use formal rulemaking procedures.

Different datasets vary in their coverage of federal agencies. Some, such as those focused on CFO Act agencies, are limited in scope, while others provide broader coverage but remain incomplete. To ensure comparability, all agency identifiers were standardized to the U.S. Office of Personnel Management (OPM) agency code system. This standardization allowed us to align agencies consistently, even when sources used different acronyms, naming conventions, or aggregation levels. When a dataset did not provide a direct match to an OPM code, we cross-checked and manually reconciled the agency identifier to maintain consistency.

In handling missing or inconsistent data, we adopted a conservative approach. If an agency, year, or variable could not be reliably matched across datasets, the observation was flagged as missing or excluded from integration, rather than being filled with assumptions. This approach ensures that the integrated dataset reflects only information that can be validated across sources, providing the most reliable foundation for subsequent analyses.

Locational labor market table. At a broader scale, the Locational Labor Market Table captures state-level labor market characteristics. We created a state-year variable as the unique ID ("LLMT_Unique_ID"), which functions as the primary key for this table and links to the Agency Table through the same variable as a foreign key. This linkage allows an analysis of how state-level labor market conditions, such as *wage levels*, the total number of employees within the occupational cluster, and occupational distributions, affect agency-level workforce dynamics. By connecting these tables, we believe that the database provides insights into relationships between federal agency operations and broader economic contexts. Moreover, we aim to show patterns such as regional variations in recruitment or retention and the impact of local labor market trends on federal employment practices.

In the table, an *occupational cluster* variable was created to group job categories into broad clusters. These clusters were derived from lists provided by two key sources: the Occupational Handbook by the Office of Personnel Management (OPM) and the Occupational Employment and Wage Statistics (OEWS) Survey by the U.S. Bureau of Labor Statistics (US BLS)³⁷. The inclusion of US BLS data allowed the development of additional metrics, including aggregated counts of individuals employed within specific clusters, summing figures annually for similar clusters within each state. Metrics such as *the Employment Percent Relative Standard Error (PRSE)*, *total number of employees in each occupational cluster*, *location quotients for each cluster*, and *average hourly and annual mean wages* were also calculated. The annual salary of each occupation in federal agencies is also calculated in the Agency Table to allow an understanding of how each public sector occupational cluster is positioned in terms of salary within a given state.

The clustering process was validated through a collaborative effort involving three graduate-level students affiliated with the CLEAR Initiative. They conducted three discussion sessions to refine the clustering approach and ensure the alignment and validity of occupational categories across different sectors. To enhance precision and reliability, the team employed generative AI tools and prompt engineering with ChatGPT for additional validation.

Table 1 illustrates occupational clusters developed by the CLEAR Initiative. The clusters integrate federal job categories from the U.S. Office of Personnel Management (OPM) Occupational Handbook (second column) with private-sector occupational groupings from the Occupational Employment and Wage Statistics (OEWS) Survey by the U.S. Bureau of Labor Statistics (third column). Each row aligns federal civil service occupations with corresponding private-sector occupational groups, organized into ten comprehensive clusters (first column).

| Given Cluster Name | US OPM Occupational Job Family | US BLS Occupational Groups |
|---------------------------------------------|---------------------------------------------------------------------|--------------------------------------------------------------------|
| Health and Social Welfare | 0600 – Medical, Hospital, Dental, and Public Health Group | 21-0000 Community and Social Service Occupations |
| | 0700 – Veterinary Medical Science Group | 29-0000 Healthcare Practitioners and Technical Occupations |
| | 0100 – Social Science, Psychology, and Welfare Group | 31-0000 Healthcare Support Occupations |
| Engineering, Architecture, and IT | 0800 – Engineering and Architecture Group | 17-0000 Architecture and Engineering Occupations |
| | 2200 – Information Technology Group | 47-0000 Construction and Extraction Occupations |
| Legal, Compliance, and Public Safety | 0900 – Legal and Kindred Group | 23-0000 Legal Occupations |
| | 1800 – Inspection, Investigation, Enforcement, and Compliance Group | 33-0000 Protective Service Occupations |
| | 1200 – Copyright, Patent, and Trademark Group | |
| | 1910 – Quality Assurance | |
| Arts, Communication, and Education | 1000 – Information and Arts Group | 27-0000 Arts, Design, Entertainment, Sports, and Media Occupations |
| | 1400 – Library and Archives Group | 25-0000 Educational Instruction and Library Occupations |
| | 1700 – Education Group | |
| Administration and Office Support | 0300 – General Administrative, Clerical, and Office Services Group | 43-0000 Office and Administrative Support Occupations |
| | 0200 – Human Resources Management Group | |
| Mathematics, Statistics, and Analytics | 1500 – Mathematics and Statistics Group | 15-0000 Computer and Mathematical Occupations |
| Science and Environmental Management | 0400 – Natural Resources Management and Biological Sciences Group | 19-0000 Life, Physical, and Social Science Occupations |
| | 1300 – Physical Sciences Group | 45-0000 Farming, Fishing, and Forestry Occupations |
| Service, Operational Support, and Logistics | 1600 – Equipment, Facilities, and Services Group | 51-0000 Production Occupations |
| | 2000 – Supply Group | 35-0000 Food Preparation and Serving Related Occupations |
| | | 49-0000 Installation, Maintenance, and Repair Occupations |
| | | 39-0000 Personal Care and Service Occupations |
| | | 37-0000 Building and Grounds Cleaning and Maintenance Occupations |
| | | |
| Business, Financial, and Management | 1100 – Business and Industry Group | 13-0000 Business and Financial Operations Occupations |
| | 0500 – Accounting and Budget Group | 11-0000 Management Occupations |
| | | 41-0000 Sales and Related Occupations |
| Transportation | 2100 – Transportation Group | 53-0000 Transportation and Material Moving Occupations |

Table 1. U.S. OPM-BLS Occupational Clustering Result.

AI Occupational impact Table. AI Occupational Impact Table provides occupation-level assessments of how generative AI may reshape public sector work over the next five years⁴⁵. Using retrieval-augmented generation (“RAG”) is a method that enhances language model responses by retrieving relevant information from an external knowledge base, allowing for more accurate and grounded outputs⁴⁶), techniques and expert-validated prompts, each occupation’s knowledge, skills, and abilities (KSAs) are scored across three dimensions of how these emerging technologies might impact occupational labor: complementarity, augmentation, and substitutivity. The metrics are designed to quantify whether AI is more likely to enhance, transform, or replace specific federal job functions. These metrics were standardized to OPM job series codes, ensuring consistency and comparability across occupational categories within our database.

Understanding emerging technologies such as AI’s effect on the federal workforce is critical. By quantifying potential shifts in competency structures, the AI Impact Table helps anticipate which occupations are most likely to evolve with technological change. This not only informs workforce planning and training but also ensures that government agencies can adapt effectively to technological disruption while maintaining service delivery and institutional capacity.

Political environment table. The Political Environment Table captures information on the U.S. political environment from the 107th to the 118th congressional term, covering the years included within the scope of this study. Each record represents a state-year observation, which can also be extended to an election cycle if a researcher wishes to examine political changes within a two-year congressional period. The table uses the unique identifier (“PET_Unique_ID”), representing a combination of state and year, as the primary key, as shown in Fig. 1.

The table focuses on the federal political composition, capturing how each state’s congressional delegation and executive leadership align with the sitting presidential administration. The Senator Representation Alignment variable indicates whether a state’s two U.S. Senators are divided, partially aligned, or fully aligned by party in relation to the president (0 = divided, 1 = half, 2 = aligned). The New Senator Indicator identifies whether a state elected a new senator compared with the previous congressional term (1 = change, 0 = no

change). The Party of Governor variable records the partisan affiliation of each state's governor (1 = Republican, 2 = Democrat, 3 = Independent). The President's Party variable identifies the political party of the U.S. president for that year (1 = Republican, 2 = Democrat). To capture the partisan composition of the legislative branch, the table also includes variables describing the U.S. House of Representatives by state. These include the Number of Republican Representatives, Number of Democratic Representatives, and Number of Other-Party Representatives, along with their corresponding percentage shares, Total House Seats by State, and Total House Seats Nationwide.

The data were compiled from multiple official sources. Information on members of Congress and their partisan affiliation derives from the U.S. Government Publishing Office's Congressional Pictorial Directory (GovInfo, 2025)⁴⁷. State governor affiliation data are drawn from the National Conference of State Legislatures (NCSL, 2025)⁴⁸. Presidential party information is obtained from the White House Historical Association (WHHA, 2025)⁴⁹. The Political Environment Table can be linked to the Locational Labor Market Table and the Agency Table through state and year, which serve as foreign keys. This structure allows researchers to analyze how changes in congressional and executive alignment correspond with agency behavior and workforce outcomes over time.

Data Records

Repository and folder structure. Each table is hosted in its own figshare repository. Within each repository:

- The root folder contains a Stata.dta file, a CSV with the same schema, and a Stata do-file that generates the.dta and documents variable types and labels.
- A raw/ subfolder contains the original source extracts referenced in the “Method — Database Development” section for the key variables used to build the table.

Job posting table. The dataset is available at: (Resh, William; Lee, Keunyoung; Ming, Yi (2025). U.S. Federal Civil Position Job Postings (2018–2023). *figshare*. Dataset. <https://doi.org/10.6084/m9.figshare.28509314.v5>). This file contains one observation per USAJOBS posting. Core fields include usajobs_control_number (primary key), posting and closing dates, pay plan, job series, grade range, vacancy title, announcement type, applicant-eligibility flags, and application counts. Use usajobs_control_number to join to the Detailed Job Posting Table, where it serves as a foreign key.

Detailed job posting table. The dataset is available at: (Resh, William; Lee, Keunyoung; Ming, Yi (2025). Detailed U.S. Federal Civil Position Postings (2018–2023). *figshare*. Dataset. <https://doi.org/10.6084/m9.figshare.28509311.v5>). This file expands each posting into distinct rows by announcement location and job series. The primary key is a numeric unique_id. It carries usajobs_control_number as a foreign key to the Job Posting Table and AT_unique_ID as a foreign key to the Agency Table. It also includes state_fips, year, and job_series_number for downstream linkage. Use usajobs_control_number to rejoin to posting-level records and AT_unique_ID to attach agency attributes.

Agency table. The dataset is available at: (Resh, William; Lee, Keunyoung; Ming, Yi (2025). U.S. Federal Civil Agencies & Departments (2018–2023). *figshare*. Dataset. <https://doi.org/10.6084/m9.figshare.28509341.v4>). This file aggregates agency attributes by agency, state, and quarter. The primary key is AT_unique_ID. Key columns include agysub (US OPM agency code), state_fips, qdate, and agency-level measures such as employment composition, separation and accession rates, and available reputation or ideology variables. The table includes LLMT_unique_ID as a foreign key to the Locational Labor Market Table. Join from the Detailed Job Posting Table using AT_unique_ID.

Locational labor market table. The dataset is available at: (Resh, William; Lee, Keunyoung; Ming, Yi (2025). Locational Labor Market Data for U.S. Federal Civil Positions (2018–2023). *figshare*. Dataset. <https://doi.org/10.6084/m9.figshare.28509347.v3>). This file reports state-year labor market indicators derived from US BLS OEWS and related sources. The primary key is LLMT_unique_ID. Core fields include state_fips, year, and occupational-cluster measures such as *bls_hsw_totalemp*, *bls_hsw_avg_empprse*, *bls_hsw_avg_jobs1000*, *bls_hsw_avg_locquot*, *bls_hsw_mean_hmean*, and *bls_hsw_mean_amean* (and analogous variables for the other clusters). Join the Agency Table to this file using LLMT_unique_ID, or align by state_fips and year for descriptive comparisons when appropriate.

AI Occupational impact table. The dataset is available at: (Resh, William; Lee, Keunyoung; Ming, Yi (2025). AI Occupational Impacts Table. *figshare*. Dataset. <https://doi.org/10.6084/m9.figshare.28665347>) The AI Occupational Impacts Table provides occupation-level assessments of how generative AI may influence public sector work over the next five years. Covering over 400 federal job series, the dataset scores each occupation's knowledge, skills, and abilities (KSAs) across three dimensions: *complementarity* (AI enhancing existing tasks), *augmentation* (AI transforming task performance), and *substitutivity* (AI replacing tasks). Integrated with OPM job series codes, this table can be linked to hiring patterns and workforce planning, supporting analyses of how emerging technologies reshape federal work.

Political environment table. The dataset is available at: (Resh, William; Lee, Keunyoung; Ming, Yi (2025). U.S. Political Environment Table (2018–2023). *figshare*. Dataset. <https://doi.org/10.6084/m9.figshare.30491783>).

| Category Name | Accuracy |
|---------------------------------|----------|
| <i>Open & closing dates</i> | 98.93% |
| <i>Pay scale & grade</i> | 99.99% |
| <i>Job family (Series)</i> | 98.97% |
| <i>Department</i> | 100% |
| <i>Location</i> | 99.87% |

Table 2. Scraped Job Attribute-level Accuracy Result.

The U.S. Political Environment Table reports annual indicators describing the partisan composition of Congress, gubernatorial affiliation, and presidential administration alignment. The primary key is PET_Unique_ID, representing a unique combination of state_fips and year. Core fields include state_senator_repre, state_new_senator, term_party_of_gov, president, house_rep_count, house_rep_perc, house_dem_count, house_dem_perc, house_others_count, house_others_perc, house_state_total, and house_total. These variables capture the alignment of U.S. Senators with the president, new senator entries between congressional terms, the partisan affiliation of each state's governor, the president's party, and the composition of the U.S. House of Representatives by party and state. The table can be joined with the Agency Table and Locational Labor Market Table using state_fips and year (or congressional term (congress)) if a researcher prefers to analyze political dynamics across the two-year legislative cycle) to examine how variations in federal political conditions correspond with agency-level operations and broader labor market outcomes.

Technical Validation

The foundational data source for INSIGHT + is the Federal Hiring Assessments and Selection Outcome Dataset curated by the U.S. General Services Administration (2023). This dataset compiles hiring process and assessment information from 24 Chief Financial Officers (CFO) Act agencies, including selection outcomes and recruitment assessments drawn from platforms such as USA Staffing, Monster, and Aviator (FAA). We adopted this dataset as a foundation of INSIGHT + and further expanded it to the current database with other important metrics (See Fig. 1), which formed a comprehensive resource platform for public sector labor market analysis across multiple occupational and government agency contexts. Accordingly, we implemented two validation approaches to ensure the reliability of this foundational data: the first is on verifying the internal accuracy of data extraction and integration processes, and the second is on assessing consistency with external benchmark data sources.

Validation #1: Web-scraping accuracy. To assess the accuracy of our data collection and integration process, we first conducted a validation using the job control numbers provided in the GSA dataset to retrieve the corresponding job announcements from the official federal hiring website, USAJOBS.gov. We randomly sampled around one-tenth of the job announcements (370,269 out of more than two million). Our results indicate that more than 99% of job control numbers in our database successfully matched with actual job announcements, suggesting a high degree of consistency in data capture and identifier integrity.

Further, we constructed an **attributes-level validation** metric to quantify the accuracy of web-scraped job attributes. For each matched job, we compared key fields—including *Open & closing dates*, *Pay scale & grade*, *Job family (Series)*, *Department*, *open and closing dates*, and *Location*—between the INSIGHT + database and the original USAJOBS posting. Each attribute match was coded as 1 and a mismatch as 0. More precisely, we compute the accuracy score for all categories c in job postings j by percentage using this formula:

$$Accuracy_{j \in k} = \frac{N_j^c}{N_j} \quad (1)$$

Where N_j^c is the number of correctly matched categories for job posting j in the random sample. N_j is the total number of categories assessed for job posting j in the random sample. To calculate the overall accuracy for all job postings in the random sample:

$$Accuracy_{sample} = \frac{\sum_{j \in k} N_j^c}{\sum_{j \in k} N_j} \quad (2)$$

This formula provides the aggregated accuracy by computing the ratio of correctly matched categories to the total number of categories assessed for all job postings j within the sample k . We randomly sampled 370,269 job postings from 2 million job postings and the attribute-level match rate across all jobs was 97.78%, supporting the reliability of the job attributes extraction and integration procedures.

Table 2 shows specific accuracy scores for each attribute. Accuracy scores represent the percentage of correctly matched categories for each attribute: Open & Closing Dates (98.93%), Pay Scale & Grade (99.99%), Job Family Series (98.97%), Department (100%), and Location (99.87%).

Validation #2: Hiring accession trend. We additionally benchmarked our database against official hiring outcome data from FedScope, US OPM's official workforce statistics platform. We used the count of selections of job announcements in the GSA dataset as a proxy measure for hiring outcome and aggregated at the agency-year level. Then compared it with accession trends reported in FedScope. It is important to note that perfect alignment

is not expected, as the GSA dataset reflects hiring assessments and selection outcomes, while FedScope reports formal personnel accessions based on post-onboarding administrative records. Discrepancies may arise due to attrition between selection and actual hire, differences in data reporting scope and timing, varying definitions of hire (direct hire vs. transfer-in), and inconsistent integration between HR systems across different government agencies. Additionally, the GSA dataset may exclude certain agencies, suppress small-sample data for privacy reasons, or reflect preliminary selection outcomes that do not result in finalized employment, all of which contribute to potential mismatches with US OPM's aggregated workforce counts. Nonetheless, our validation shows a strong positive correlation ($r = 0.92$) between selection trends in the INSIGHT + dataset and hiring accession counts in FedScope, demonstrating that the GSA dataset offers a robust and predictive signal for broader hiring patterns. This finding supports the utility of the GSA dataset for labor market analysis, even if exact parity with official workforce totals is not expected.

Usage Notes

In addition to direct use of the static dataset, researchers may find value in linking the database to external contextual information. For example, the dataset can be combined with Political Environment Metrics, which capture election outcomes, partisan dynamics, public opinion indicators, and related state-level political conditions. By connecting these variables through the shared state–election cycle identifiers, users can analyze how changes in the political environment shape agency operations, hiring priorities, and labor market outcomes. Furthermore, while outside the scope of this Data Descriptor, we are developing an automated data collection pipeline to facilitate recurring updates of job postings and metadata from USAJOBS.gov. This infrastructure is intended to minimize data latency and provide users with more timely snapshots of federal workforce dynamics, complementing the static dataset described here.

Data availability

All materials supporting this study are publicly available through the figshare repository at https://figshare.com/projects/_NEW_Integrated_Network_Solutions_in_Federal_Government_Hiring_Trends_INSIGHT_/239336. The relational data infrastructure includes the following tables:

1. U.S. Political Environment Table (2018–2024). <https://doi.org/10.6084/m9.figshare.30491783.v9>⁵⁰
2. U.S. Federal Civil Agencies & Departments (2018–2025). <https://doi.org/10.6084/m9.figshare.28509341.v10>⁵¹
3. AI Occupational Impacts Table. <https://doi.org/10.6084/m9.figshare.28665347.v4>⁵²
4. Detailed U.S. Federal Civil Position Postings (2018–2023). <https://doi.org/10.6084/m9.figshare.28509311.v5>⁵³
5. U.S. Federal Civil Position Job Postings (2018–2023). <https://doi.org/10.6084/m9.figshare.28509314.v5>⁵⁴
6. Locational Labor Market Data for U.S. Federal Civil Positions (2018–2023). <https://doi.org/10.6084/m9.figshare.28509347.v3>⁵⁵

A Data Dictionary for the entire data infrastructure is available in a zip file at <https://doi.org/10.6084/m9.figshare.28029611.v14>. As well, results from our Technical Validation are available at <https://doi.org/10.6084/m9.figshare.28680110.v1>^{56,57}.

Code availability

We used Stata 14 to extract, transform, and load the data, and all processing code is available in a Stata do-file. Researchers can use Stata to access and manipulate the database directly. Furthermore, for creating the FIPS Code variable and expanding the dataset, we used Python to process the data. Expanding the dataset addressed instances where multiple job series and locations were listed within a single cell by separating and restructuring the data into individual rows. Moreover, the *FIPS Code* was created using data from the US Census Bureau⁵⁸ and matched to the *announcement location* of each job posting to ensure geographic consistency and accuracy. The python programs were used to collect and validate public sector job postings. They include two Python scripts: “supp_scrapper.py”, which extracts structured job data from publicly accessible federal hiring platforms, and “tech_valid.py”, which performs keyword-based validation to classify postings related to technical occupations. Both scripts were developed and tested using Python 3.10 and rely on standard libraries such as *requests*, *BeautifulSoup4*, and *pandas*. The database and code are available at the following link: https://figshare.com/projects/_NEW_Integrated_Network_Solutions_in_Federal_Government_Hiring_Trends_INSIGHT_/239336.

Received: 31 March 2025; Accepted: 4 February 2026;

Published online: 19 February 2026

References

1. U.S. Office of Personnel Management. *Federal workforce size and composition*. <https://data.opm.gov/explore-data/analytics/workforce-size-and-composition> (2026)
2. U.S. Bureau of Labor Statistics. Employment continues to expand in 2023, though at a slower pace than in the previous 2 years. *BLS* <https://www.bls.gov/opub/mlr/2024/article/employment-continues-to-expand-in-2023-though-at-a-slower-pace-than-in-the-previous-2-years.htm> (2024).
3. Megerian, C. The Trump administration sets the stage for large-scale federal worker layoffs in a new memo. *The Associated Press*. <https://apnews.com/article/trump-elon-musk-federal-workers-layoffs-d295d4bb2cdd5023c27d9cb03754e81b> (2025).
4. U.S. Office of Personnel Management. *Federal workforce changes*. <https://data.opm.gov/explore-data/analytics/workforce-changes> (2026).
5. The Economic Times, US Education Department to cut half its staff as Trump eyes elimination *The Economic Times*. Available at: https://economictimes.indiatimes.com/news/international/global-trends/us-education-department-to-cut-half-its-staff-as-trump-eyes-elimination/articleshow/118911626.cms?utm_source=contentofinterest&utm_medium=text&utm_campaign=cppst (2025).
6. Light, P. C. *The True Size of Government*. (Brookings Institution Press, 1999).
7. Lewis, D. E. & Selin, J. L. Political control and the forms of agency independence. *George Wash. Law Rev.* **80**, 619–645 (2012).

8. Joyce, P. G. *The Congressional Budget Office: Honest Numbers, Power, and Policy Making* (Georgetown Univ. Press, 2011).
9. Rockman, B. A. *The Oxford Handbook of Political Institutions* (Oxford Univ. Press, 2019).
10. Keynes, J. M. *The General Theory of Employment, Interest and Money* (Macmillan, 1936).
11. Garrett, T. A. & Rhine, R. M. On the size and growth of government. *Fed. Reserve Bank St. Louis Rev.* **93**, 13–30 (2011).
12. Selden, S. C. & Moynihan, D. P. A model of voluntary turnover in state government. *Rev. Public Pers. Adm.* **20**, 63–74 (2000).
13. U.S. Office of Personnel Management. Hiring authorities. *U.S. Office of Personnel Management* (2021).
14. Falk, J. Comparing benefits and total compensation in the federal government and the private sector. *The BE Journal of Economic Policy and Analysis* **12**, Article 48 (2012).
15. Light, P. C. *A Government Ill Executed: The Decline of the Federal Service and How to Reverse It* (Harvard Univ. Press, 2008).
16. Wilson, J. Q. *Bureaucracy: What Government Agencies Do and Why They Do It* (Basic Books, 1989).
17. Aghion, P., Antonin, C. & Bunel, S. Artificial intelligence, growth and employment: The role of policy. *Econ. Stat.* **510–512**, 150–164 (2019).
18. Moynihan, D. P. & Pandey, S. K. Finding workable levers over work motivation: Comparing job satisfaction, job involvement, and organizational commitment. *Adm. Soc.* **39**, 803–832 (2007).
19. Deschenes, O. Climate policy and labor markets. *Natl. Bur. Econ. Res.* <https://doi.org/10.3386/w16111> (2010).
20. Joyce, P. G. The Obama administration and PBB: Building on the legacy of federal performance-informed budgeting? *Public Administration Review* **71**(3), 356–367 (2011).
21. Kettl, D. F. *The Global Public Management Revolution* (Brookings Institution Press, 2005).
22. Bureau of Labor Statistics. Union members summary. U.S. Department of Labor (2020).
23. Freeman, R. B. & Han, E. The war against public sector collective bargaining in the U.S. *J. Ind. Relat.* **54**, 386–408 (2012).
24. Light, P. C. *The True Size of Government* (Brookings Institution Press, 1999).
25. Perry, J. L. Measuring public service motivation: An assessment of construct reliability and validity. *J. Public Adm. Res. Theory* **6**, 5–22 (1996).
26. Lewis, G. B. & Stewart, L. M. *Men and Women of the State: Public Administrators at the State Level* (M.E. Sharpe, 1996).
27. Dunleavy, P., Margetts, H., Bastow, S. & Tinkler, J. *Digital Era Governance: IT Corporations, the State, and e-Government* (Oxford Univ. Press, 2006).
28. Deloitte. Government workforce rebound: Navigating changes in public-sector employment. *Deloitte Insights*. Retrieved from <https://www2.deloitte.com/us/en/insights/economy/spotlight/government-workforce-rebound.html> (2025).
29. Neogov. The quiet crisis in the public sector. *Neogov* <https://info.neogov.com/resources/quiet-crisis-report> (2022).
30. Chen, J. & Johnson, T. Federal employee unionization and presidential control of the bureaucracy: Estimating and explaining ideological change in executive agencies. *Journal of Theoretical Politics* <https://doi.org/10.1177/0951629813518126> (2015).
31. Selin, J. L. What Makes an Agency Independent? *American Journal of Political Science* <https://doi.org/10.1111/ajps.12161> (2015).
32. Richardson, M. D., Clinton, J. D. & Lewis, D. E. Elite perceptions of agency ideology and workforce skill. *The Journal of Politics* <https://doi.org/10.1086/694846> (2018).
33. Bellodi, L. A dynamic measure of bureaucratic reputation: New data for new theory. *American Journal of Political Science* <https://doi.org/10.1111/ajps.12695> (2023).
34. Richardson, M. D. Characterizing Agencies' Political Environments: Partisan Agreement and Disagreement in the US Executive Branch. *The Journal of Politics* <https://doi.org/10.1086/729954> (2024).
35. Bouchard, M. J. *et al.* A relational database to understand social innovation and its impact on social transformation. *New Frontiers Social Innov. Res.* 69–85 (2015).
36. Federal Administrative Procedure Sourcebook. https://sourcebook.acus.gov/wiki/Federal_Administrative_Procedure_Sourcebook (2019).
37. U.S. Bureau of Labor Statistics. Occupational Employment and Wage Statistics (OEWS). *BLS* <https://www.bls.gov/oes> (2024).
38. Date, C. J. *An Introduction to Database Systems* 8th edn (Pearson Education, 2004).
39. Elmasri, R. & Navathe, S. B. *Fundamentals of Database Systems* 7th edn (Pearson, 2015).
40. Silberschatz, A., Korth, H. F. & Sudarshan, S. *Database System Concepts* 7th edn (McGraw-Hill, 2020).
41. Vassiliadis, P. & Simitis, A. Extraction, transformation, and loading. *Encyclopedia Database Syst.* **10**, 14 (2009).
42. McBride, K., Aavik, G., Toots, M., Kalvet, T. & Krimmer, R. How does open government data driven co-creation occur? Six factors and a 'perfect storm'; insights from Chicago's food inspection forecasting model. *Gov. Inf. Q.* **36**, 88–97 (2019).
43. U.S. General Services Administration. Federal Hiring Assessments and Selection Outcome Dataset. GSA <https://catalog.data.gov/dataset/federal-hiring-assessments-and-selection-outcome-dataset> (2024).
44. Lewis, D. E. & Selin, J. L. Sourcebook of United States Executive Agencies. *Administrative Conference of the United States*. <https://acus.gov/publication/sourcebook-united-states-executive-agencies> (2012).
45. Resh, W. G. *et al.* Complementarity, Augmentation, or Substitutivity? The Impact of Generative Artificial Intelligence on the U.S. Federal Workforce (No. arXiv:2503.09637). *arXiv* <https://doi.org/10.48550/arXiv.2503.09637> (2025).
46. Lewis, P. *et al.* Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in neural information processing systems* **33**, 9459–9474. <https://arxiv.org/abs/2005.11401> (2020).
47. U.S. Government Publishing Office. Congressional Pictorial Directory. *GovInfo* <https://www.govinfo.gov/collection/congressional-pictorial-directory?path=/gpo/Congressional%20Pictorial%20Directory/118th%20Congress> (2025).
48. U.S. National Conference of State Legislatures (NCSL). State Partisan Composition. *NCSL*. <https://www.ncsl.org/about-state-legislatures/state-partisan-composition> (2025).
49. White House Historical Association (WHHA). Presidents and Their Administrations. *WHHA*. <https://www.whitehousehistory.org/> (2025).
50. Resh, W., Lee, K., & Ming, Y. U.S. Political Environment Table (2018–2024). *figshare. Dataset* <https://doi.org/10.6084/m9.figshare.30491783.v9> (2025).
51. Resh, W., Lee, K., & Ming, Y. U.S. Federal Civil Agencies & Departments (2018–2025). *figshare. Dataset*. <https://doi.org/10.6084/m9.figshare.28509341.v10> (2025).
52. Resh, W., Lee, K., & Ming, Y. AI Occupational Impacts Table. *figshare* <https://doi.org/10.6084/m9.figshare.28665347.v4> (2025).
53. Resh, W., Lee, K., & Ming, Y. Detailed U.S. Federal Civil Position Postings (2018–2023). *figshare* <https://doi.org/10.6084/m9.figshare.28509311.v5> (2025).
54. Resh, W., Lee, K., & Ming, Y. U.S. Federal Civil Position Job Postings (2018–2023). *figshare*. <https://doi.org/10.6084/m9.figshare.28509314.v5> (2025).
55. Resh, W., Lee, K., & Ming, Y. Locational Labor Market Data for U.S. Federal Civil Positions (2018–2023). *figshare* <https://doi.org/10.6084/m9.figshare.28509347.v3> (2025).
56. Resh, W., Lee, K., & Ming, Y. Data Dictionary for INSIGHT+. *figshare*. <https://doi.org/10.6084/m9.figshare.28029611.v14> (2025).
57. Resh, W., Ming, Y., & Lee, K. Technical Validation. *figshare*. <https://doi.org/10.6084/m9.figshare.28680110.v1> (2025).
58. U.S. Census Bureau. American National Standards Institute (ANSI), Federal Information Processing Series (FIPS), and Other Standardized Geographic Codes. *U.S. Census Bureau* <https://www.census.gov/programs-surveys/geography/guidance/geo-identifiers.html> (2020).

Acknowledgements

The authors thank Martino Meraviglia, Wing Yat Lam, Sukwon Choi, and Yongjin Ahn for support in data collection. The authors thank Mark Richardson, Jennifer Selin, David Lewis, and Luca Bellodi for advisement and the Volcker Alliance for its financial and substantive support.

Author contributions

William G. Resh is the principal investigator and led the conceptual framing, development, data acquisition, and writing process. Keunyoung (Eli) Lee is the project lead and was involved in all aspects including data acquisition, coding, structuring, refinement, and validation. Yi Ming contributed to most aspects and led in the development of the AI Impact Table as well as replicating Bellodi reputation scores for recent years. Xinyao (Andy) Xia contributed to both the AI Impact Table and various scraping and validation methods used for this data. Nicole Dias and Kecheng (Anderson) Liu made contributions to the acquisition and structuring of data. Darren Cao contributed to the validation of the occupational clustering process. William Huh contributed to data refinement and validation.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to W.G.R. or K.E.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026