



June 2023

DATA SCIENCE

NIH Needs to Implement Key Workforce Planning Activities

Why GAO Did This Study

NIH, the federal government’s leader in supporting biomedical research, faces a shortage of employees with data science expertise needed to, among other things, analyze and extract insights from increasingly large and complex sets of data. In June 2018, NIH developed a Strategic Plan for Data Science, which included an objective to enhance its data science workforce that addresses this need.

The explanatory statement accompanying the Further Consolidated Appropriations Act, 2020, contained a provision for GAO to review NIH’s data science workforce planning. This report, among other things, determines the extent to which 1) NIH has conducted data science workforce strategic planning in accordance with key practices and 2) NIH’s data management and sharing policy and guidance are consistent with federal guidance.

To do so, GAO assessed agency documentation against key workforce planning practices identified in prior GAO work. It also compared NIH’s data management and sharing policy and plans to relevant federal requirements, and interviewed NIH officials.

What GAO Recommends

GAO is making 11 recommendations to NIH to fully implement key workforce planning activities and finalize data management and sharing guidance. NIH concurred with nine of the recommendations and stated it had implemented the other two. However, the agency did not provide sufficient evidence of the implementation. As a result, GAO continues to believe the recommendations are appropriate. View [GAO-23-105594](#). For more information, contact David B. Hinchman at 214-777-5719 or HinchmanD@gao.gov.

DATA SCIENCE

NIH Needs to Implement Key Workforce Planning Activities

What GAO Found

While the National Institutes of Health (NIH) included a data science workforce goal in its June 2018 Strategic Plan for Data Science, the agency has not fully implemented the key workforce planning activities established by federal guidance (see table). For example, NIH developed and implemented plans to enhance its data science workforce; however, these plans were not linked to gaps in its data science workforce. Near the conclusion of GAO’s review, officials said that an agency-wide Data Science Workforce Working Group had been established to address priority hiring and retention needs. However, they did not provide documentation supporting the group’s activities. Fully addressing the workforce planning activities would help ensure that NIH has the data science workforce it needs to effectively meet its mission.

National Institutes of Health’s Implementation of Key Activities for Data Science Workforce Planning

Key workforce planning practices and supporting activities	Rating
Set the strategic direction for workforce planning	
Establish and maintain a workforce planning process	Partially implemented
Develop competency and staffing requirements	Partially implemented
Analyze the workforce to identify skill gaps	
Reassess competency and staffing needs regularly	Not implemented
Determine gaps in competencies and staffing regularly	Not implemented
Develop and implement strategies to address skill gaps	
Develop strategies and plans to address gaps in competencies and staffing	Partially implemented
Implement activities that address gaps	Partially implemented
Monitor and report progress in addressing skill gaps	
Monitor the agency’s progress in addressing competency and staffing gaps	Not implemented
Report to agency leadership on progress in addressing competency and staffing gaps	Not implemented

Legend: Fully implemented: NIH provided evidence that addressed the activity; partially implemented: NIH provided evidence that it had addressed some, but not all of the activity; not implemented: NIH did not provide evidence that it had addressed any of the activity.

Source: GAO analysis of NIH documentation. | GAO-23-105594

NIH’s data management and sharing policy, effective January 2023, is consistent with relevant Office of Science and Technology Policy data sharing requirements. However, NIH had not finalized the guidance its staff needs to evaluate the data management and sharing plans and determine researchers’ compliance with them. In addition, officials stated several times during the course of GAO’s review that they had revised their time frames for doing so. The officials said they were delayed in completing the guidance because they were focused on informing the public about the new policy. They also anticipated releasing the guidance by June 2023 in time to assess the first round of plans. However, NIH did not document this new time frame. Documenting the new time frame and monitoring progress against it would ensure NIH’s accountability for finalizing the guidance on time. In addition, until the agency finalizes and implements the guidance, its staff are less likely to consistently assess data sharing plans. This, in turn, would limit NIH’s goal of maximizing appropriate sharing of scientific data generated from federally funded research.

Contents

Letter		1
	Background	3
	NIH Has Not Fully Implemented Key Data Science Workforce Planning Activities	8
	NIH Has a Defined Process for Funding Computational Talent in Its Grant Awards	14
	NIH Established Data Sharing Policy That Addresses Requirements but Has Not Finalized Supporting Guidance	20
	Conclusions	22
	Recommendations for Executive Action	23
	Agency Comments and Our Evaluation	24
Appendix I	Objectives, Scope, and Methodology	26
Appendix II	Comments from the Department of Health and Human Services	31
Appendix III	GAO Contact and Staff Acknowledgments	36
Tables		
	Table 1: Key Workforce Planning Practices and Activities	7
	Table 2: National Institutes of Health’s (NIH) Implementation of Key Activities for Data Science Workforce Planning	9
	Table 3: Key Workforce Planning Practices and Activities	26
Figures		
	Figure 1: Overview of the National Institutes of Health’s Grant Application and Peer Review Process	18

Abbreviations

IC	Institute/Center
NIH	National Institutes of Health
OPM	Office of Personnel Management
OSTP	Office of Science and Technology Policy
SRG	Scientific Review Group

This is a work of the U.S. government and is not subject to copyright protection in the United States. The published product may be reproduced and distributed in its entirety without further permission from GAO. However, because this work may contain copyrighted images or other material, permission from the copyright holder may be necessary if you wish to reproduce this material separately.



June 22, 2023

The Honorable Tammy Baldwin
Chair
The Honorable Shelley Moore Capito
Ranking Member
Subcommittee on Labor, Health and Human Services, Education, and
Related Agencies
Committee on Appropriations
United States Senate

The Honorable Robert Aderholt
Chair
The Honorable Rosa DeLauro
Ranking Member
Subcommittee on Labor, Health and Human Services, Education, and
Related Agencies
Committee on Appropriations
House of Representatives

A talented and diverse cadre of digital-ready, tech-savvy federal employees is critical to federal agencies as they carry out their missions and address challenges facing the United States. However, agencies face a shortage of staff in fields such as artificial intelligence, data science, and computational biology expertise. For example, the National Institutes of Health (NIH), the federal government’s leader in supporting biomedical research, faces a shortage of data scientists. Since 2001, GAO has identified mission-critical gaps in federal workforce skills and expertise in fields such as science, technology, engineering, and mathematics as high-risk areas.¹

The explanatory statement accompanying the Further Consolidated Appropriations Act, 2020, contained a provision for GAO to review NIH’s efforts to acquire data scientists for its internal workforce and how NIH

¹GAO, *High-Risk Series: Efforts Made to Achieve Progress Need to Be Maintained and Expanded to Fully Address All Areas*, [GAO-23-106203](#) (Washington, D.C.: April 20, 2023).

funds computational talent (e.g., data scientists) in its grant awards.² Our objectives were to (1) determine the extent to which NIH has conducted data science workforce strategic planning in accordance with key practices; (2) describe how NIH funds computational talent in its grant awards; and (3) determine the extent to which NIH's data management and sharing policy and guidance are consistent with federal guidance.

To address the first objective, we adjusted GAO's IT workforce planning framework³ to reflect a general workforce, including the data science workforce. We validated the revised framework by confirming that it is supported by federal guidance and prior GAO work and seeking input from internal subject matter experts. We compared NIH's data science workforce planning documentation to practices identified in our revised workforce planning framework.

We reviewed NIH's Strategic Plan for Data Science, which includes an objective to enhance the NIH data science workforce, and related 2019 implementation plans; the 2018 State of *Data Science Workforce Development* report; and data science position description and job analysis documents. We focused our review at the agency level. We also selected three of 21 institutes to verify NIH officials' claims that each institute determines its need for data science expertise. We selected these institutes based on NIH officials identifying them as having key data science responsibilities. The selected institutes are the National Library of Medicine, the National Human Genome Research Institute, and the National Institute of Child Health and Human Development. In addition, we interviewed officials from NIH's Office of Data Science Strategy and Office of Human Resources. Because we selected the institutes to review based on NIH officials identifying them as having key data science responsibilities, our findings about the institutes' workforce planning cannot be used to make inferences about other NIH institutes.

We assessed NIH's implementation of each of the workforce planning activities as

- fully implemented—the agency provided evidence that it fully implemented the activity;

²The joint explanatory statement of conference, 165 Cong. Rec. H11061, H11072 (daily ed. Dec. 17, 2019) (statement of Chairwoman Lowey), specifically referenced in § 4 of the Further Consolidated Appropriations Act, 2020, Pub. L. No. 116-94, § 4, 133 Stat. 2534, 2536 (2019).

³GAO, *IT Workforce: Key Practices Help Ensure Strong Integrated Program Teams; Selected Departments Need to Assess Skill Gaps*, [GAO-17-8](#) (Washington, D.C.: Nov. 30, 2016).

-
- partially implemented—the agency provided evidence that it had addressed some, but not all, of the activity; or
 - not implemented—the agency did not provide any evidence that it implemented the activity.

To address the second objective, we reviewed NIH documentation on how the agency funds grants, including grants supporting computational work, and interviewed NIH officials about the process. We also interviewed officials and representatives from research organizations and associations of computational experts who represent grant applicants to obtain their perspectives on the grant application and funding process. We selected the organizations and associations based on being included in a prior relevant GAO report and recommendations from those we interviewed. In addition, we reviewed relevant reports and studies identified by these organizations and through a literature search to understand the grant application and funding process.

To address the third objective, we identified relevant requirements in the Office of Science and Technology Policy’s (OSTP) memorandum on increasing access to the results of federally funded scientific research.⁴ Specifically, according to the memo, agencies investing over \$100 million annually in research and development should create a public access plan that ensures that researchers develop data management plans, the plans are appropriately evaluated, and researchers comply with them. We then compared NIH’s data management and sharing policy and plans for developing associated guidance to the OSTP requirements. Additional details on our objectives, scope, and methodology can be found in appendix I.

We conducted this performance audit from December 2021 to June 2023 in accordance with generally accepted government auditing standards. Those standards require that we plan and perform the audit to obtain sufficient, appropriate evidence to provide a reasonable basis for our findings and conclusions based on our audit objectives. We believe that the evidence obtained provides a reasonable basis for our findings and conclusions based on our audit objectives.

Background

NIH’s mission is to “Turn Discovery into Health” by seeking fundamental knowledge about the nature and behavior of living systems and to use that knowledge to enhance health, lengthen life, and reduce illness and disability. To achieve this mission, NIH works to support research aimed

⁴Office of Science and Technology Policy, *Increasing Access to the Results of Federally Funded Scientific Research* (Washington, D.C., Feb. 22, 2013).

at protecting and improving human health; train the biomedical research workforce; and develop scientific infrastructure. NIH also works to contribute to the nation's economic growth by expanding the biomedical knowledge base and promote integrity, public accountability, and societal responsibility in scientific research.

NIH is made up of 28 components: the Office of the Director, 21 institutes, and six centers. The Office of the Director operates as NIH's central managing office, and has responsibility for setting policy, and for planning, managing, and coordinating overall NIH programs and activities. Each institute has a specific research agenda that often focuses on particular diseases or body systems. For example, the National Eye Institute's mission is to conduct and support research, training, health information dissemination, and other programs with respect to blinding eye diseases, visual disorders, preservation of sight, and the special health problems and requirements of the blind. The centers vary in function, to include research, program support, patient care, and other NIH-wide services. The six centers include, for example, the NIH Clinical Center, Center for Information Technology, and Center for Scientific Review.

For fiscal year 2022, NIH received an appropriation of about \$45.2 billion. For fiscal year 2023, the agency received about \$47.5 billion. About 84 percent of NIH's funding (for example, about \$38 billion in fiscal year 2022) is passed on to researchers and research institutions around the country—the extramural research community. About 10 percent supports intramural projects conducted by scientists in its own laboratories. The remaining six percent covers research support, administrative, and facility costs.

NIH reported that each year it receives about 54,000 research project grant applications and funds almost 50,000 new and continuing grants. The grants support about 300,000 researchers, including more than 43,000 principal investigators at approximately 2,500 universities, medical schools, and other research institutions in every state of the U.S. and around the world.

Data Science Is Important for Biomedical Research

Data science is a growing field due to the rapidly increasing volume of complex data. According to the National Academies of Sciences, Engineering and Medicine, sudden orders-of-magnitude increases in data

collection have moved biomedical research into the realm of “big data.”⁵ Also, given recent advances in genetics and genomics research, biomedical research will continue to experience tremendous growth that likely will add to increasing volumes of data.

In June 2018, NIH developed a Strategic Plan for Data Science to address storing data efficiently and securely, making data usable to as many people as possible, and developing a workforce capable of taking advantage of advances in data science and information technology. In the plan, NIH defines data science as the interdisciplinary field of inquiry in which quantitative and analytical approaches, processes, and systems are developed and used to extract knowledge and insights from increasingly large and/or complex sets of data.

One of the goals in NIH’s plan is to enhance workforce development for biomedical data science. Associated with this goal, the plan identifies an objective to enhance the NIH data science workforce. The plan states that given the importance of data science for biomedical research, NIH needs an internal workforce that is increasingly skilled in this area. This includes ensuring that NIH program and review staff who administer and manage grants and coordinate the evaluation of applications have sufficient experience with and knowledge of data science.⁶

⁵According to the National Institute on Standards and Technology, “big data” is a term used to describe the large amount of data in the networked, digitized, sensor-laden, information-driven world. The data can overwhelm traditional technical approaches, and the growth of data is outpacing scientific and technological advances in data analytics. In the NIH context, big data are generally associated with biomedical research fields, such as genomics, where petabyte-sized datasets, i.e., datasets measuring quadrillions of bytes, are common.

⁶Also associated with this goal is an objective to expand the national research workforce. In its plan, NIH says that modern biomedical research is becoming increasingly quantitative and it is essential that the next generation of researchers be equipped with the skills needed to take advantage of the growing promise of data science for advancing human health. NIH says that it will work to ensure that NIH-funded training and fellowship programs emphasize teaching of quantitative and computational skills and integrate training in data science approaches throughout their curricula and during mentored research.

The Office of Personnel Management Has Established a New Occupational Series for Data Science

In December 2021, the Office of Personnel Management (OPM) established an occupational series for data science.⁷ According to OPM, data scientists use scientific methodology, processes, algorithms, and systems to extract insights from structured and unstructured data, and to provide guidance for data-driven decision making. Further, they use powerful technology (e.g., machine learning and artificial intelligence) to manage enormous data sets and work with complex algorithms. The work requires expertise in coding, prototyping, and integration with complex data systems.

In establishing the series, OPM determined that data science work may be found in various occupational series, including, for example, the Epidemiology–Medical and Health Care Series and the Statistician Series. According to OPM guidance, agencies may use a parenthetical related to data science with the occupational title for positions that perform data science work as a major portion of the job. For example, NIH has a position title, which is Health Scientist (Data Science).

Strategic Workforce Planning Can Help Agencies Identify Gaps in Scientific and Technical Expertise

We previously reported that identifying the skills needed to achieve their mission and to close any gaps in their current workforce helps agencies to select the right human capital strategies to address those needs. Agency efforts to identify skill gaps and future needs in the expertise of their scientific and technical staff through strategic workforce planning can help ensure they are better positioned to implement their missions.⁸

In November 2016, we issued an evaluation framework, which identifies four steps and eight supporting activities, for assessing federal agencies' IT workforce planning efforts.⁹ We used the framework to evaluate selected agencies' strategic IT workforce planning efforts in 2016 and 2019.¹⁰

⁷OPM requires agencies to prepare and submit human resources, payroll, and training data files to its Enterprise Human Resources Integration data warehouse. The data warehouse system collects, integrates, and publishes data about executive branch employees, supporting agency and government-wide data analytics. Among the data collected about each employee is their occupational series. The code for the data science occupational series is 1560.

⁸GAO, *Science and Technology: Strengthening and Sustaining the Federal Science and Technology Workforce*, [GAO-21-461T](#) (Washington, D.C.: Mar. 17, 2021).

⁹[GAO-17-8](#).

¹⁰[GAO-17-8](#) and GAO, *Information Technology: Agencies Need to Fully Implement Key Workforce Planning Activities*, [GAO-20-129](#) (Washington, D.C.: Oct. 30, 2019).

While the framework was developed for an IT workforce, it identifies fundamental and sound workforce planning practices that are relevant to a data science workforce. A general version of the workforce planning framework is shown in table 1. It is based on federal guidance, including the OPM Workforce Planning Model and prior GAO reports.¹¹ It includes practices and activities that are applicable to a data science workforce.

Table 1: Key Workforce Planning Practices and Activities

Set the strategic direction for workforce planning
Establish and maintain a workforce planning process
Develop competency and staffing requirements
Analyze the workforce to identify skill gaps
Reassess competency and staffing needs regularly
Determine gaps in competencies and staffing regularly
Develop strategies and implement activities to address skill gaps
Develop strategies and plans to address gaps in competencies and staffing
Implement activities that address gaps
Monitor and report progress in addressing skill gaps
Monitor the agency's progress in addressing competency and staffing gaps
Report to agency leadership on progress in addressing competency and staffing gaps

Source: GAO analysis of federal guidance. | GAO-23-105594

The Office of Science and Technology Policy Required Agencies to Increase Access to Federally Funded Research Data

In 2013, the Office of Science and Technology Policy (OSTP) released its Memorandum on *Increasing Access to the Results of Federally Funded Scientific Research*. The memo states that federal agencies must have clear and coordinated policies for increasing access to federally funded digital scientific data.¹² It then requires that agencies investing over \$100 million annually in research and development create a plan to support increased public access to the results of research funded by the federal

¹¹*Human Capital: Key Principles for Effective Strategic Workforce Planning*, GAO-04-39 (Washington, D.C.: Dec. 11, 2003) and *Standards for Internal Control in the Federal Government*, GAO-04-704G (Washington, D.C.: Sept. 10, 2014).

¹²The OSTP memo defines data as the digital recorded factual material commonly accepted in the scientific community as necessary to validate research findings. It includes data sets used to support scholarly publications, but does not include laboratory notebooks, preliminary analyses, drafts of scientific papers, plans for future research, peer review reports, communications with colleagues, or physical objects, such as laboratory specimens.

NIH Has Not Fully Implemented Key Data Science Workforce Planning Activities

government. The memo states that each public access plan shall ensure that

- all researchers receiving federal grants and contracts for scientific research develop data management plans, as appropriate. The plans should describe how the researchers will provide for long-term preservation of, and access to, scientific data in digital formats resulting from federally funded research.¹³
- the merits of data management plans are evaluated appropriately.
- researchers comply with data management plans and policies.

According to OSTP, policies that mobilize data for re-use through preservation and broader public access also maximize the impact and accountability of the federal research investment. Further, according to OSTP, access to digital data sets resulting from federally funded research allows companies to focus resources and efforts on understanding and exploiting discoveries. For example, making genome sequences publicly available has spawned many biotechnology innovations.

NIH partially implemented four of the activities for its data science workforce that GAO identified are needed for effective workforce planning, and did not implement the other four. NIH's implementation of the activities are identified in table 2.

¹³We are referring to these plans as data management and sharing plans because they are required to describe how researchers will provide for access to scientific data. If researchers believe long-term preservation and access cannot be justified, they are to explain why.

Table 2: National Institutes of Health’s (NIH) Implementation of Key Activities for Data Science Workforce Planning

Key workforce planning activity	Description	Rating
<i>Set the strategic direction for workforce planning</i>		
Establish and maintain a workforce planning process	The agency should have a documented data science workforce planning process that describes how the agency will implement key workforce planning activities, including those identified in our workforce planning framework. The workforce planning process should define roles and responsibilities for implementing the activities; align with mission goals and objectives; and address both the agency-level and component-level workforce, including how the agency is to maintain visibility and oversight into component-level workforce planning efforts. In addition, the agency should periodically update the process.	Partially implemented
Develop competency and staffing requirements	The agency should develop a set of competency (e.g., knowledge, skills, and abilities) requirements for its data science workforce. In addition, the agency should develop staffing requirements, which include projections of future staffing needs over several years.	Partially implemented
<i>Analyze the workforce to identify skill gaps</i>		
Reassess competency and staffing needs regularly	The agency should periodically assess competency and staffing needs.	Not implemented
Determine gaps in competencies and staffing regularly	The agency should periodically analyze its workforce to determine gaps in data science competencies. In addition, the agency should periodically determine gaps in staffing for its data science workforce.	Not implemented
<i>Develop strategies and implement activities to address skill gaps</i>		
Develop strategies and plans to address gaps in competencies and staffing	The agency should develop strategies and plans to address identified competency gaps, including specific actions and milestones that are linked to a gap. In addition, the agency should develop strategies and plans to address identified staffing gaps, including specific actions and milestones that are linked to a gap.	Partially implemented
Implement activities that address gaps	The agency should execute its strategies and plans to address identified gaps in competencies and staffing.	Partially implemented
<i>Monitor and report progress in addressing skill gaps</i>		
Monitor the agency’s progress in addressing competency and staffing gaps	The agency should track progress in implementing strategies and plans to address competency gaps. In addition, the agency should track progress in implementing strategies and plans to address staffing gaps.	Not implemented
Report to agency leadership on progress in addressing competency and staffing gaps	The agency should periodically report to agency leadership on progress in implementing strategies and plans to address gaps in competencies. In addition, the agency should track progress in implementing strategies and plans to address gaps in staffing.	Not implemented

Source: GAO analysis of NIH documentation. | GAO-23-105594

Legend: Fully implemented: NIH provided evidence that addressed the activity; partially implemented: NIH provided evidence that it had addressed some, but not all of the activity; not implemented: NIH did not provide evidence that it had addressed any of the activity.

NIH partially established a data science workforce planning process.

It developed plans to enhance its data science workforce through training and a data fellows program. Specifically, NIH documented a process to determine its data science competency needs. To do this, it planned to collect information from various NIH audiences through surveys, interviews, and focus groups to determine and document the levels of data science expertise needed. The types of expertise needed might range from general data literacy for non-computational researchers and program staff to higher-level data science techniques for data scientists.

However, NIH does not have a fully documented planning process for its data science workforce. Specifically, the agency has not documented a process for developing data science staffing requirements and reassessing competency and staffing needs regularly. It also has not established a process for conducting an analysis of its workforce to determine its data science competency and staffing gaps. In addition, it has not documented a process for monitoring and periodically reporting to agency leadership on progress in addressing competency and staffing gaps.

Further, it has not defined roles and responsibilities at the agency- and component-levels. In addition, NIH has not documented how it will maintain visibility and oversight into component-level data science workforce planning efforts.

Until NIH fully documents a data science workforce planning process that includes all elements and addresses all the activities in our framework, the agency will likely not have the staff with the necessary knowledge, skills, and abilities to support its mission and goals.

NIH developed data science competency requirements, but not staffing requirements. Specifically, in 2020, the Office of Human Resources created standardized position descriptions and job analysis documents for data scientists that hiring managers can tailor to their needs. These documents describe the knowledge, competencies, and skills required for NIH data scientists, such as statistical methods and techniques, technology application, and data management.

However, NIH has not developed staffing requirements. NIH officials stated that Office of Human Resources specialists meet with hiring managers and institute and center officials on a regular basis to address staffing requirements and to communicate recruitment and hiring goals. However, the three institutes NIH officials identified as having key data science responsibilities did not have documentation supporting these activities. Specifically,

-
- According to National Institute of Child Health and Human Development plans, in April 2023 through June 2023, the institute plans to conduct a current state analysis of all staff, including identifying critical skills and competencies of the workforce based on projected scientific and administrative needs. However, the institute has not established plans for identifying data science staffing requirements.
 - NIH officials said that the National Library of Medicine’s data science staffing requirements take a variety of forms, from biomedical informatics experts to technical information specialists who work with data, to administrative staff who make decisions based on data. The officials also said that the institute had recently hired two principal investigators who apply computational, data science approaches to medical imaging and electronic health record data. However, the officials did not provide documented data science staffing requirements.
 - The National Human Genome Research Institute stated that it plans to do a needs assessment over the next 3 to 5 years. According to officials, implementation of NIH’s Data Management and Sharing Policy, effective in January 2023, will likely raise the need for additional data science expertise in the institute to review submitted data management and sharing plans, make recommendations about data repositories, and provide guidance to investigators.

Until NIH conducts an analysis to fully determine its data science staffing needs, the agency lacks assurance that it is appropriately identifying the number of data science staff it needs to meet its mission and programmatic goals.

NIH has not reassessed its data science competency and staffing needs. While the agency determined its data science competency needs in 2020, it has not reassessed competency needs since then. In addition, as previously noted, the agency has not determined its data science staffing needs.

Until it reassesses data science competency and staffing needs, and establishes plans to regularly reassess them, NIH lacks assurance that it has the appropriate number of staff and that the staff have the necessary knowledge and skills.

NIH has not determined gaps in its data science competencies and staffing. In response to our request for NIH’s determination of gaps in data science competencies and staffing, NIH provided the 2018 National Library of Medicine State of Data Science Workforce Development report.

It also provided data scientist position description and job analysis documents. However, the documents did not include a gap analysis.

NIH officials also referred us to institutes and centers, saying that each determines its need for data science expertise. However, none of the three institutes we reviewed had analyzed their workforce to determine what gaps in data science competencies and staffing they may have.

Until NIH analyzes its workforce to identify its data science competency and staffing gaps, the agency will lack assurance that it has the data science workforce it needs to effectively meet its mission.

NIH has developed plans to enhance its data science workforce, but the plans are not linked to gaps. NIH's 2018 Strategic Plan for Data Science includes developing data science training programs for NIH staff and the launch of the NIH Data Fellows program. In addition, the supporting February 2019 implementation plan includes determining and documenting the levels of data science expertise needed; providing coordination and collaboration for data science training efforts for NIH staff; and establishing formal and informal mentoring opportunities to connect data science learners with data science expertise.

The implementation plan also includes steps for launching the NIH Data Fellows program. These steps are recruit and hire its first cohort via a funding announcement; recruit and place subsequent cohorts of fellows; and develop program evaluations.

In addition, the report on the National Library of Medicine's 2018 data science workshop included actions NIH could take to incentivize and attract data scientists who were not currently working with biomedical data. These actions included establishing a NIH webpage with data-science related items; communicating the availability of funding opportunities that allow for data scientists and subject matter experts to serve as equal partners to lead research projects or training efforts; and creating multiple pathways for discovering funding opportunities (e.g., discipline-specific listservs).

However, NIH has not developed strategies and plans linked to gaps because, as previously stated, it has not determined the gaps. Until NIH develops strategies and plans that are linked to gaps, the agency will be limited in its ability to acquire the data science workforce it needs to effectively meet its mission.

NIH implemented activities to enhance its data science workforce, but the activities are not linked to gaps. For example, NIH established a "Data Science at NIH" webpage that provides links to training resources

and information related to data science. In addition, NIH established a Data and Technology Advancement (DATA) Scholar program (i.e., its planned data fellows program, mentioned above), which provides one- to two-year positions in which scholars address challenging biomedical data problems with the potential for substantial public health impact.

However, NIH has not implemented activities to address gaps in data science competencies and staffing because, as previously stated, it has not determined the gaps. Until NIH determines its data science competencies and staffing gaps and implements activities to address the gaps, the agency will be limited in its ability to acquire the data science workforce it needs to effectively meet its mission.

NIH has not monitored progress in addressing data science competency and staffing gaps. It has not done this because, as previously stated, it has not determined the gaps.

Further, NIH officials stated that they do not track data science staff. Specifically, they stated that the agency does not have a tracking system or centralized process for identifying employees who are referred to as or may be considered data science staff. The officials said that this is because the system that NIH uses to process personnel actions is built and maintained by the Department of Health and Human Services, and is designed around OPM's Data Standards and Guide to Processing Personnel Actions, which do not call for elements relating to data science.¹⁴ However, since August 2021, OPM data standards have included the code for the data scientist occupational series, which NIH could use to track data science staff.

NIH officials said that although there is a new data science occupational series, they classify positions based on the paramount knowledge required. For example, the officials said that if a position requires mastery level knowledge of the biological sciences and performs data science work, the position is classified in the natural resources management and biological science occupational series. While NIH's position classification is consistent with OPM guidance, establishing a process to track the competencies and staff associated with its data science workforce would

¹⁴Office of Personnel Management, "The Guide to Processing Personnel Actions" (Washington, D.C.), accessed Feb. 13, 2023, <https://www.opm.gov/policy-data-oversight/data-analysis-documentation/personnel-documentation/#url=Personnel-Actions> and "Data Standards" (Washington, D.C.), accessed Aug. 10, 2022, <https://dw.opm.gov/datastandards/referenceData/1490/current?index=O&category=&d-5590585-p=1>.

help the agency determine whether it is meeting its goal to acquire the data science workforce it needs to effectively meet its mission.

Until NIH analyzes its workforce to determine its data science competencies and staffing gaps and monitors progress in addressing the gaps, the agency will be unable to ensure that any strategies and plans it implements will effectively address gaps. In addition, until it establishes a process for tracking data science staff it will be limited in its ability to monitor its progress in acquiring a data science workforce.

NIH has not reported progress in addressing data science competency and staffing gaps. It has not done this because, as previously stated, it has not determined the gaps. Until NIH analyzes its workforce to determine data science competencies and staffing gaps and agency leadership receives reports on progress addressing gaps, NIH's leadership will lack the information necessary to effectively address the gaps.

Officials did not explain why NIH had not fully implemented the workforce planning activities. However, near the conclusion of our review, they said that the agency established a Data Science Workforce Working Group, composed of experts from each NIH component. They said that the group is charged with providing the agency an implementation strategy and executing on priority hiring and retention needs. However, they did not provide documentation supporting the establishment of the group or its activities.

NIH Has a Defined Process for Funding Computational Talent in Its Grant Awards

NIH funds computational talent in its grant awards in the same way it funds other researchers. In general, the process for obtaining a grant from NIH is

- NIH announces opportunities for grant funding
- Researchers submit applications, including a budget
- Applications undergo two levels of peer review
- The institute or center director makes the final funding decision

Salaries for personnel funded by NIH grants are limited by statutory restriction.¹⁵

¹⁵See, e.g., Consolidated Appropriations Act, 2023, Pub. L. No. 117-328, division H, title II, § 202, 136 Stat. 4459 (2022); Consolidated Appropriations Act, 2022, Pub. L. No. 117-103, division H, title II, § 202, 136 Stat. 49, 466 (2022).

NIH Announces Grant Funding Opportunities

NIH advertises opportunities for grants through funding opportunity announcements on its website. The three primary types of funding announcements are:

- parent announcements, which are broad and allow applicants to submit investigator-initiated applications for specific activity codes;
- program announcements, which are issued by one or more institute or center to highlight areas of scientific interest; and
- requests for applications, which are issued by one or more institute or center to highlight well-defined areas of scientific interest to accomplish specific program objectives.

Researchers Submit Applications, Including Budgets

To pursue a grant funding opportunity, applicants submit grant applications. These applications are to include, among other things, a budget that considers the cost of personnel, such as computational experts, who would work on a project. According to NIH guidance, applicants should review funding opportunity announcements for budget criteria, which can include limits on the types of expenses (e.g., no construction allowed), caps on certain expenses (e.g., salaries), and overall funding limits.

Applicants can develop one of two budget submissions—modular or detailed—depending on the total of direct costs requested and the activity code.¹⁶ Modular budgets require less detail and can be used when applications meet certain criteria. A modular budget is used to request up to a total of \$250,000 in direct costs per year in modules of \$25,000. These budgets are to include, among other things, the name, role, and number of person-months for all individuals on the project. The modular budget does not need to include salary rates, but it should consider the statutory salary cap (discussed below).

The detailed budget requires that all personnel from the applicant organization who are dedicating effort to the project be listed with their base salary and effort in person-months, even if they are not requesting salary support. NIH instructs applicants to base their personnel budget on actual institutional base salaries (not the cap) so that NIH staff have the most current information and can apply the appropriate cap at the time of award.

¹⁶NIH uses 246 activity codes to differentiate its research programs. For example, R series codes (e.g., R01) are for research grants, K series codes (e.g., K01) are for career development awards, and T series codes are for research training (e.g., T32), among others.

Applications Undergo Two Levels of Peer Review

Federal law requires two levels of peer review for applications submitted to NIH.¹⁷ According to NIH, the peer review policy is intended to ensure that applications are evaluated using a process that is fair, equitable, timely, and balanced. The peer review system is based on two sequential levels of review for each application—first by a scientific review group and then by the advisory council or board of the funding institute or center.

Scientific Review Group. A scientific review group is primarily composed of 12 to 22 non-federal scientists with expertise in the relevant field of research. When NIH receives an application, the Division of Receipt and Referral, within the Center for Scientific Review, assigns the application to the appropriate institute or center. The referral officer at the institute or center then assigns the application to the appropriate scientific review group (also known as a study section). The assignment is based on many factors, including the scientific area of research, expertise needed, applicant requests, and assignments of previous applications. While NIH has standing review groups with focus on various scientific areas, a special emphasis panel may be formed to review applications requiring special expertise. NIH officials stated that reviewers are recruited based on the expertise needed and the subject matter of applications received. Research organization officials and representatives stressed the importance of including computational experts in the review groups to ensure a fair evaluation of computational work proposed in the grant applications.

Once assigned, the scientific review group follows a defined review process to assess the scientific and technical merit of the applications and determine overall impact scores for them, based on review criteria specified in the relevant funding announcement or request for application. In some cases, the scientific review group also gives the application a percentile rank.

Following the initial review, the scientific review officer prepares a summary statement, which is used by the National Advisory Council or Board of the Institute/Center for the next level of review.¹⁸ The statement reflects the scientific review group's assessment, including the reviewers' written comments, and, for scored applications, a summary of the discussion and the impact score.

¹⁷42 U.S.C. §289(a).

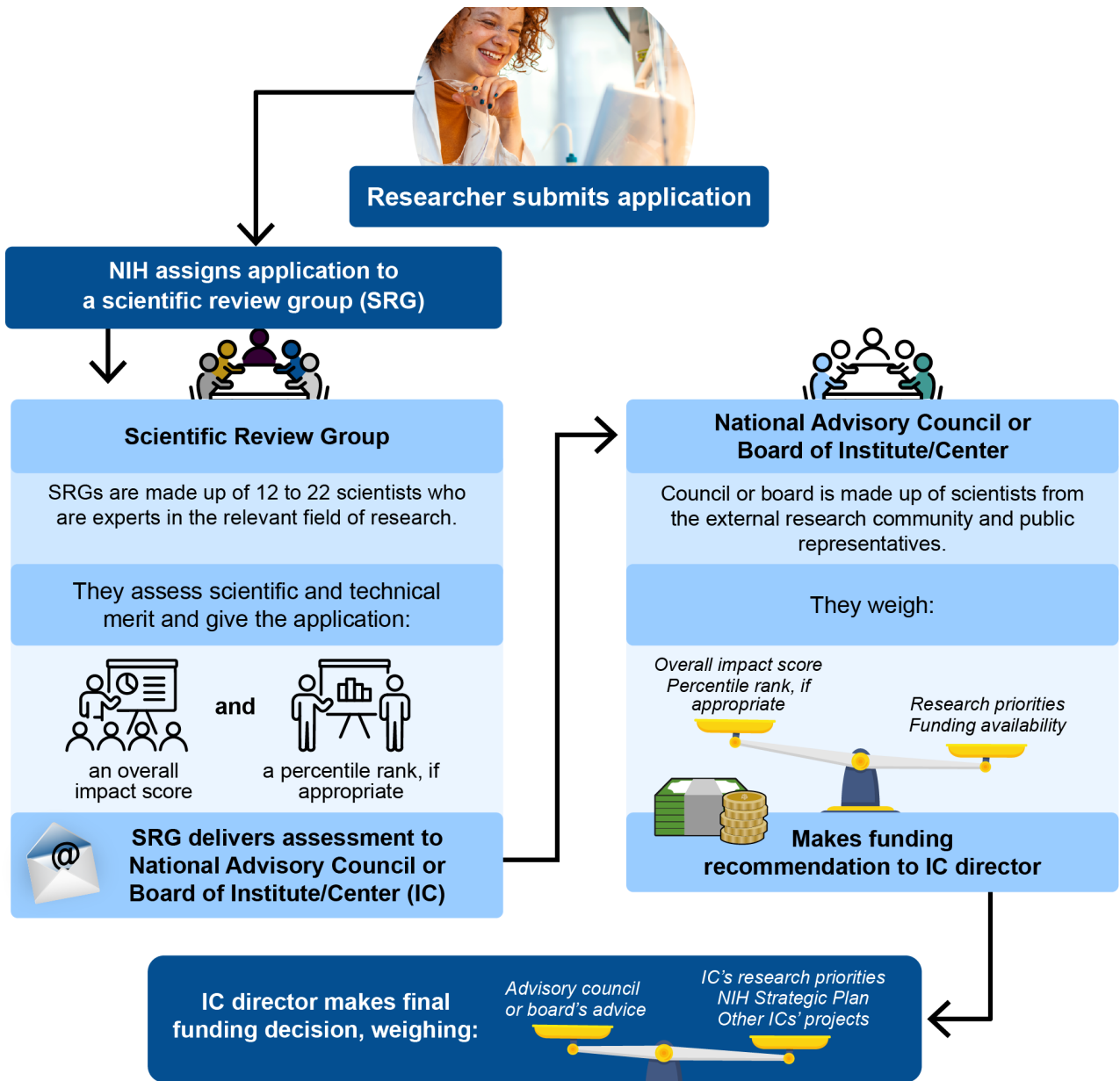
¹⁸A scientific review officer is responsible for managing the peer review meeting, the procedures for evaluating the applications assigned to the scientific review group, and determinations and management of conflicts of interest.

National Advisory Council or Board of the Institute/Center. The national advisory council or board is composed of scientists from the external research community and public representatives chosen by the relevant institute or center and approved by the Department of Health and Human Services. The council or board weighs the application's scientific and technical merit (i.e., overall impact score) and percentile rank, if appropriate, against research priorities and funding availability, and advises the institute/center director on funding decisions.

**Institute/Center Director
Makes the Final Funding
Decision**

The institute/center director makes the final award decision, including the funding level, from among those applications receiving a favorable initial review and advisory council recommendation. The director is to weigh the institute/center's mission and research priorities, NIH-wide Strategic Plan, and other institutes' and centers' projects on similar topics. NIH advises the applicant of the decision to award or not award a grant. NIH also advises grant applicants that funding for a project may be reduced after an award has been granted. For example, NIH may reduce a project's budget if sufficient funds are not available to support it. See figure 1 for an overview of the grant application and peer review process.

Figure 1: Overview of the National Institutes of Health’s Grant Application and Peer Review Process



Source: NIH information; images: zaie/stock.adobe.com, stas111/stock.adobe.com, lightpoet/stock.adobe.com. | GAO-23-105594

Salaries Funded by Grants Are Limited by Federal Law

Grants awarded by NIH provide for reimbursement of actual, allowable costs incurred, including for salaries and wages, within certain limits. According to cost principles for NIH awards, the cost of salaries and

wages is allowable for reimbursement if, among other things, it is reasonable for the services rendered. A cost is reasonable if it does not exceed what would be incurred by a prudent person under the circumstances prevailing at the time the decision was made to incur the cost. In determining the reasonableness of a given cost, consideration is to be given to market prices for comparable services for the geographic area.¹⁹ Some of the officials and representatives from the research organizations and associations we interviewed stated that they sometimes found it challenging to compete with the private sector to compensate computational talent. They described actions they took to address this challenge, including recruiting overseas or in the research field, where lower salaries may be accepted.

Since fiscal year 1990, federal law has limited the direct salary that individuals being funded by NIH grants could receive.²⁰ The restriction is in the annual appropriations act for NIH.²¹ Starting in fiscal year 1999, the salary cap was tied to the Federal Executive Level pay scale.

Over the years, the level at which the cap was tied has changed. In fiscal year 1999, the cap was Executive Level III of the Federal Executive pay scale. In fiscal year 2000, the cap was increased to Executive Level II, and in fiscal year 2001, it was further increased to Executive Level I. In fiscal year 2012, the cap was lowered to Executive Level II, and has remained at this level since then.

The salary cap for grants awarded in January 2023 through September 2023 is \$212,100. The salary cap for grants awarded in 2022 was \$203,700.

According to some researchers we interviewed, the salary cap most affects personnel who have high salaries, such as data scientists, and therefore a larger gap exists between their salary and the NIH cap.

¹⁹Code of Federal Regulations, Title 45 – Public Welfare, Subtitle A – Department of Health and Human Services, Subchapter A – General Administration, Part 75 – Uniform Administrative Requirements, Cost Principles, and Audit Requirements for HHS Awards, Subpart E Cost Principles; and U.S. Department of Health and Human Services, National Institutes of Health, NIH Grants Policy Statement (December 2022).

²⁰Departments of Labor, Health and Human Services and Education, and Related Agencies Appropriations Act, 1990, Pub. L. No. 101-166, title II, § 217, 103 Stat. 1159, 1178 (1989).

²¹See, e.g., Consolidated Appropriations Act, 2023, Pub. L. No. 117-328, division H, title II, § 202, 136 Stat. 4459 (2022); Consolidated Appropriations Act, 2022, Pub. L. No. 117-103, division H, title II, § 202, 136 Stat. 49, 466 (2022).

NIH Established Data Sharing Policy That Addresses Requirements but Has Not Finalized Supporting Guidance

NIH issued a policy, effective in January 2023, which addresses OSTP's requirement to ensure that all researchers receiving NIH federal grants and contracts for scientific research develop data management and sharing plans, as appropriate. However, as of February 2023, the agency had not finalized guidance and tools for staff to assess submitted plans. It also had not finalized guidance for staff to determine compliance with approved plans. In addition, NIH had not documented its updated time frames for doing so.²²

NIH's Policy Requires Data Management and Sharing Plans

As described earlier in this report, OSTP requires each agency investing over \$100 million annually in research and development to create a public access plan to ensure that all researchers receiving federal grants and contracts for scientific research develop data management and sharing plans, as appropriate. Consistent with OSTP's requirement, in October 2020, NIH issued a new policy effective as of January 25, 2023, that requires all grant applications, as appropriate, to include a data management and sharing plan.²³ The new policy replaces NIH's 2003 Data Sharing Policy and establishes the expectation for maximizing the appropriate sharing of scientific data generated from NIH-funded or conducted research.

NIH's new policy requires applicants competing for a grant to submit a plan that outlines how scientific data and any accompanying metadata will be managed and shared, taking into account any potential restrictions or limitations. The policy also requires awardees to comply with the plan as approved by the institute or center.

NIH Has Not Finalized Guidance for Evaluating Data Sharing Plans and Determining Compliance with Them

According to OSTP, agencies should ensure that data management and sharing plans are evaluated appropriately. Agencies should also ensure that researchers comply with data management and sharing plans and policies.

To address OSTP's requirements, NIH's policy states that the funding institute or center is to assess submitted data management and sharing

²²As of February 23, 2023, NIH officials stated that the agency had not finalized guidance for staff to assess submitted plans and researchers compliance with them. In reviewing a draft of this report in April 2023, an official stated via email that the agency had disseminated guidance at the end of February 2023 and provided us the guidance. We discuss the guidance and what remains to be done in the agency comments section of this report.

²³The policy applies to all research, funded or conducted in whole or in part by NIH that results in the generation of scientific data, regardless of funding level or funding mechanism.

plans. In addition, according to the policy, the funding institute or center is to determine compliance with approved plans.

In February 2021, NIH developed a plan to assist institute and center program staff in evaluating data management and sharing plans and determining researchers' compliance with them. The plan included activities, with associated dates and deliverables, for developing guidance for the staff. According to the plan, NIH's Office of Science Policy and Office of Extramural Research were to determine the process for assessing a submitted data management and sharing plan by July 2021. The offices were to consider timing, roles and responsibilities, tools (e.g., checklists), and processes. In addition, the Office of Extramural Research was to lead an activity to develop guidance that program and grants management staff are to use to determine and document compliance checks that are consistent with the policy.

However, NIH did not meet the July 2021 deadline and has since pushed its time frame out several times. It currently estimates releasing the final guidance by June 2023 in time for its staff to begin assessing data management and sharing plans with the first round of applications subject to the policy.²⁴

Officials from the Office of Extramural Research explained that NIH had been delayed in completing the assessment and compliance resources because they were focused on informing the public about the new policy. Specifically, they said, NIH had prioritized development and release of materials needed for the initial stages of the application and award life cycle. For example, they said they have made updates to instructions in templates for notices of funding opportunity, which are used by NIH funding opportunity announcement writers, and materials for initial receipt of applications. In addition, they said, they have been incorporating feedback from the public into staff guidance.

NIH officials described the agency's efforts to develop guidance for staff to assess data management and sharing plans and its efforts to develop guidance for determining compliance with the plans.

- They said that the Office of Extramural Research was in the process of finalizing resources for staff to assess data management and

²⁴As of February 23, 2023, NIH officials stated that the agency had not finalized guidance for staff to assess submitted plans and researchers compliance with them. In reviewing a draft of this report in April 2023, an official stated via email that the agency had disseminated guidance at the end of February 2023 and provided us the guidance. We discuss the guidance and what remains to be done in the agency comments section of this report.

sharing plans and review particularly complex plans. According to officials, this includes an optional assessment decision tool to help program offices review data management and sharing plans. NIH documentation indicates this tool is to define specific criteria that can be used broadly across NIH institutes and centers to distinguish between plans that are acceptable and those that are not, so that program staff can clearly and consistently assess each plan. The officials also said that they were developing plans to set up a panel of experts from across NIH to provide consultation to program offices across institutes and centers in cases where assessment is particularly challenging.

- In addition, NIH officials said that the Office of Extramural Research was finalizing staff guidance on compliance oversight. They said the office was also completing updates to checklists related to applications and progress reporting to support compliance monitoring.

While NIH officials stated that the agency revised its time frames for finalizing its guidance for assessing data management plans and determining compliance with those plans, it did not update its policy implementation plan accordingly. NIH officials said that the agency did not intend to update its policy implementation plan with the new time frames because the plan was meant to serve as an early, high-level road map to help prepare for implementation efforts. While the plan may have been an early, high-level road map, there is still value in documenting updated time frames and tracking progress against them. Until it does so, NIH may lack the accountability for completing the guidance in time for its staff to use it for the first set of plans in June 2023. In addition, until NIH completes and implements the guidance, NIH staff are less likely to clearly and consistently evaluate data management and sharing plans and determine researchers' compliance with them. This, in turn, impedes NIH's goal of maximizing appropriate sharing of scientific data generated from federally funded research.

Conclusions

Given the biomedical field's increasing reliance on large volumes of complex data, it is critically important for NIH to ensure that it has the data science staff it needs to meet its responsibilities for administering tens of billions of dollars in annual research grants. However, NIH has not fully addressed key workforce planning practices for its data science workforce. Until NIH has fully determined its data science staffing needs and identified its workforce gaps, the agency will lack assurance that it has the appropriately skilled staff to evaluate grant applications and to administer and manage grants. In addition, it will likely not meet its goal of enhancing its data science workforce set in its Strategic Plan for Data Science.

Also, given the need to make scientific data as broadly available as possible to maximize the impact of the federal government's investment in research, it is important that NIH fully implement its new data sharing policy. However, as of February 2023, NIH had not finalized guidance for its staff to assess researchers' data management and sharing plans required by this policy and determine researcher compliance with those plans. It also had not documented its new time frame for doing so. Documenting the new time frame and monitoring progress against it would ensure accountability for finalizing the guidance in time for staff to use it to assess the first round of plans subject to the new policy. Without the guidance, NIH staff will be limited in their ability to ensure that researchers develop and implement adequate data sharing plans.

Recommendations for Executive Action

We are making the following 11 recommendations to NIH:

The NIH Director should ensure that NIH establishes a comprehensive data science workforce planning process that addresses the shortfalls noted in this report. (Recommendation 1)

The NIH Director should ensure that NIH develops staffing requirements for the data science workforce. (Recommendation 2)

The NIH Director should ensure that NIH reassesses its data science competency and staffing needs periodically. (Recommendation 3)

The NIH Director should ensure that NIH analyzes its workforce to identify gaps in data science competencies and staffing. (Recommendation 4)

The NIH Director should ensure that NIH develops specific strategies and plans to address identified gaps in data science competencies and staffing. (Recommendation 5)

The NIH Director should ensure that NIH implements strategies and plans to address identified gaps in data science competencies and staffing. (Recommendation 6)

The NIH Director should ensure that NIH develops and tracks metrics to monitor the agency's progress in addressing data science competency and staffing gaps. (Recommendation 7)

The NIH Director should ensure that NIH develops a process to track data science staff. (Recommendation 8)

The NIH Director should ensure that NIH requires reporting to agency leadership on progress made in addressing data science competency and staffing gaps. (Recommendation 9)

The NIH Director should ensure that NIH documents new time frames to complete the guidance its staff will need to assess data management and

Agency Comments and Our Evaluation

sharing plans, and ensure that the guidance is implemented.
(Recommendation 10)

The NIH Director should ensure that NIH documents new time frames to complete the guidance its staff will need to determine researchers' compliance with their data management and sharing plans, and ensure that the guidance is implemented. (Recommendation 11)

We provided a draft of this report to the Department of Health and Human Services for comment. In its written comments, which are reproduced in appendix II, the department concurred with recommendations one through nine and stated that it would provide Congress an action plan to address them. NIH also stated that it considered recommendations 10 and 11 to be implemented and noted that it had provided GAO with guidance it had recently issued for its staff to implement the data management and sharing policy and supplemental notices. The department also provided technical comments, which we have incorporated as appropriate.

Regarding recommendation 10, we verified that NIH had issued guidance for staff to assess data management and sharing plans. However, it is not complete. Associated with the guidance is a checklist with questions that program and grants management officials are to complete. NIH officials stated in May 2023 that the checklist questions were in the process of being revised to provide additional clarity and would be reissued to staff when finalized. NIH also released a decision support tool that staff may use to inform responses to the checklist questions. However, the agency did not provide a time frame for completing the checklist questions. In addition, the agency did not provide documentation showing that it has implemented the guidance (i.e., that staff have used the guidance to assess plans). Accordingly, we believe the recommendation is still appropriate and plan to monitor NIH's efforts to implement it.

Regarding recommendation 11, our review of the documentation NIH provided showed that the agency is still in the process of completing guidance for determining compliance with data management and sharing plans. Specifically, staff guidance issued in February 2023 shows that program officials are required to assess a grant recipient's progress and adherence to the plan as part of the research progress reporting process. However, the form that grant recipients are to use to report on progress has not been updated with questions about compliance with plans. NIH officials said that they anticipate making changes to the form and related instructions by early fiscal year 2024. However, they did not provide documentation of the new time frame for completing the changes to the form and related instructions needed to complete the guidance. In

addition, the agency has not yet implemented the guidance. We therefore believe the recommendation is still appropriate and will continue to monitor NIH's efforts to implement it.

We are sending copies of this report to the appropriate congressional committees and the Director of the National Institutes of Health. In addition, the report is available at no charge on the GAO website at <http://www.gao.gov>.

If you or your staff have any questions about this report, please contact me at (214) 777-5719 or hinchmand@gao.gov. Contact points for our Offices of Congressional Relations and Public Affairs may be found on the last page of this report. GAO staff who made contributions to this report are listed in appendix III.

A handwritten signature in black ink that reads "David B Hinchman". The signature is written in a cursive, flowing style with a long horizontal flourish at the end.

David B. Hinchman
Director, Information Technology and Cybersecurity

Appendix I: Objectives, Scope, and Methodology

Our objectives were to (1) determine the extent to which the National Institutes of Health (NIH) has conducted data science workforce strategic planning in accordance with key practices; (2) describe how NIH funds computational talent in its grant awards; and (3) determine the extent to which NIH's data sharing policy and guidance are consistent with federal guidance.

To address the first objective, we relied on practices in GAO's IT workforce planning framework and related evaluation criteria established in prior work.¹ While the framework was developed for an IT workforce,² we adjusted it to reflect a general workforce, including the data science workforce. In particular, for the activity in the IT workforce planning framework that calls for agencies to implement activities that address IT skill gaps, we deleted references to activities that are required by law and Office of Management and Budget guidance specifically for IT workforces. We also deleted "IT" from the original framework. We validated the revised framework by confirming that it remains supported by underlying federal guidance and prior GAO reports. We also sought input from internal subject matter experts in workforce planning. The framework contains four practices and eight supporting key workforce planning activities that, when implemented, facilitate effective workforce planning. For the practices and activities in the framework, see table 3.

Table 3: Key Workforce Planning Practices and Activities

<i>Set the strategic direction for workforce planning</i>
Establish and maintain a workforce planning process
Develop competency and staffing requirements
<i>Analyze the workforce to identify skill gaps</i>
Reassess competency and staffing needs regularly
Determine gaps in competencies and staffing regularly
<i>Develop strategies and implement activities to address skill gaps</i>

¹GAO, *IT Workforce: Key Practices Help Ensure Strong Integrated Program Teams; Selected Departments Need to Assess Skill Gaps*, [GAO-17-8](#) (Washington, D.C.: Nov. 30, 2016). To create this framework, we determined strategic human capital planning and IT workforce planning activities from legislation; Office of Management and Budget and Office of Personnel Management guidance; and prior GAO reports, including *Human Capital: Key Principles for Effective Strategic Workforce Planning*, [GAO-04-39](#) (Washington, D.C.: Dec. 11, 2003) and *Standards for Internal Control in the Federal Government*, [GAO-04-704G](#) (Washington, D.C.: Sept. 10, 2014). We identified recommended practices and requirements from those sources, established an evaluative framework, and vetted it with internal and external stakeholders.

²[GAO-17-8](#).

Develop strategies and plans to address gaps in competencies and staffing

Implement activities that address gaps

Monitor and report progress in addressing skill gaps

Monitor the agency's progress in addressing competency and staffing gaps

Report to agency leadership on progress in addressing competency and staffing gaps

Source: GAO analysis of federal guidance. | GAO-23-105594

To determine the extent to which NIH had implemented the key workforce planning activities for its data science workforce, we requested its data science workforce planning documentation and compared it against the evaluation criteria for the activities. We reviewed, for example, NIH's Strategic Plan for Data Science, which includes an objective to enhance the NIH data science workforce, and related 2019 implementation plans; the 2018 *State of Data Science Workforce Development* report; and data science position description and job analysis documents. We also interviewed officials from NIH's Office of Data Science Strategy and Office of Human Resources.

We focused our review at the agency level. We also selected three of 21 institutes to verify NIH officials' claims that each institute and center determines its need for data science expertise. We selected these institutes based on NIH officials identifying them as having key data science responsibilities. The institutes are the National Library of Medicine, the National Human Genome Research Institute, and the National Institute of Child Health and Human Development. We requested information and documentation from the institutes on their data science workforce planning activities and evaluated it relative to the evaluation criteria. Because we selected the institutes to review based on NIH officials identifying them as having key data science responsibilities, our findings about the institutes' workforce planning cannot be used to make inferences about other NIH institutes.

To assess NIH's implementation of the key activities, we used the following evaluation criteria:

- **Establish and maintain a workforce planning process.** To fully implement this activity, the agency should have a documented data science workforce planning process that describes how the agency will implement key workforce planning activities, including those identified in our workforce planning framework. The workforce planning process should define roles and responsibilities for implementing the activities, and align with mission goals and objective. It should also address both the agency-level and component-level workforce, including how the agency is to maintain

visibility and oversight into component-level workforce planning efforts. In addition, the agency should periodically update the process.

- **Develop competency and staffing requirements.** To fully implement this activity, the agency should develop a set of competency (e.g., knowledge, skills, and abilities) requirements for its data science workforce. In addition, the agency should develop staffing requirements, which include projections of future staffing needs over several years.
- **Reassess competency and staffing needs regularly.** To fully implement this activity, the agency should periodically assess competency and staffing needs.
- **Determine gaps in competencies and staffing regularly.** To fully implement this activity, the agency should periodically analyze its workforce to determine gaps in data science competencies. In addition, the agency should periodically determine gaps in staffing for its data science workforce.
- **Develop strategies and plans to address gaps in competencies and staffing.** To fully implement this activity, the agency should develop strategies and plans to address identified competency gaps, including specific actions and milestones that are linked to a gap. In addition, the agency should develop strategies and plans to address identified staffing gaps, including specific actions and milestones that are linked to a gap.
- **Implement activities that address gaps.** To fully implement this activity, the agency should execute its strategies and plans to address identified gaps in competencies and staffing.
- **Monitor the agency's progress in addressing competency and staffing gaps.** To fully implement this activity, the agency should track progress in implementing strategies and plans to address competency and staffing gaps.
- **Report to agency leadership on progress addressing competency and staffing gaps.** To fully implement this activity, the agency should periodically report to agency leadership on progress in implementing strategies and plans to address gaps in competencies and staffing.

To determine an overall rating for each of the eight key workforce planning activities, we summarized the results of our assessments of the information NIH and the three selected institutes provided relative to the evaluation criteria, and determined whether NIH fully implemented, partially implemented, or did not implement the activity. If documentation

supported that NIH had implemented an activity, we rated it fully implemented. If documentation demonstrated that NIH had implemented some but not all of the activity, we rated it partially implemented. If NIH did not provide documentation to support that an activity had been implemented, we rated it not implemented.

To address the second objective, we reviewed NIH documentation to understand how the agency funds research grants, including grants with computational talent. The documents we analyzed included NIH's grant application guidance and forms, sample grants, budget development guidance, NIH's Grant Policy Statement, and grant application and peer review process documentation. We also interviewed NIH officials from NIH's Office of Extramural Research, Office of Science Policy, Office of Research Reporting and Analysis, and Office of Policy for Extramural Research Administration about the grant application and review process.

We also conducted semi-structured interviews of 20 officials and representatives from six research organizations and associations of computational experts who represent grant applicants, to obtain their perspectives on the grant application and funding process. We selected the organizations and associations based on being included in a prior relevant GAO report and recommendations from those we interviewed. These organizations are the Allen Institute, American Statistical Association, Association of American Medical Colleges, Council on Governmental Relations, Federation of American Societies for Experimental Biology, and International Society for Computational Biology.

Although our semi-structured interviews were not generalizable, they provided specific examples on how NIH grant applicants develop budgets; how NIH determines grant awards, including for salaries; how the amount of funding applicants receive compares to how much is requested; and flexibilities and constraints grant recipients have to supplement NIH grants. In their responses, they shared challenges and ways for overcoming them. We followed up with NIH on the researchers' experiences and incorporated NIH's responses as appropriate into our description of the process for funding personnel with NIH grant awards.

To address the third objective, we reviewed the 2013 Office of Science and Technology Policy (OSTP) Memorandum on Increasing Access to the *Results of Federally Funded Scientific Research* to understand its

requirements for public access to scientific data in digital forms.³ We identified three requirements in the OSTP memorandum that were relevant to our work. According to the memo, agencies investing over \$100 million annually in research and development should create public access plans that ensure

- researchers develop data management plans describing how they will provide for long-term preservation of, and access to, scientific data in digital formats;
- the merits of data management plans are evaluated appropriately; and
- researchers comply with approved data management plans and policies.

We compared NIH's data management and sharing policy and plans for developing associated guidance to the OSTP requirements. These documents include the NIH Policy for Data Management and Sharing, which was released in October 2020 and is effective as of January 2023, and supplemental information; February 2021 policy implementation plans and April 2022 communications plans; and July 2022 staff training on implementing NIH's data management and sharing policy. In addition, we interviewed officials from NIH's Office of Data Science Strategy, Office of Science Policy, and Office of Extramural Research to discuss their planned implementation of NIH's data sharing policy and development of related guidance.

We conducted this performance audit from December 2021 to June 2023 in accordance with generally accepted government auditing standards. Those standards require that we plan and perform the audit to obtain sufficient, appropriate evidence to provide a reasonable basis for our findings and conclusions based on our audit objectives. We believe that the evidence obtained provides a reasonable basis for our findings and conclusions based on our audit objectives.

³ Office of Science and Technology Policy, *Increasing Access to the Results of Federally Funded Scientific Research* (Washington, D.C.: Feb. 22, 2013). The OSTP memo defines data as the digital recorded factual material commonly accepted in the scientific community as necessary to validate research findings including data sets used to support scholarly publications, but does not include laboratory notebooks, preliminary analyses, drafts of scientific papers, plans for future research, peer review reports, communications with colleagues, or physical objects, such as laboratory specimens.

Appendix II: Comments from the Department of Health and Human Services



DEPARTMENT OF HEALTH & HUMAN SERVICES

OFFICE OF THE SECRETARY

Assistant Secretary for Legislation
Washington, DC 20201

June 2, 2023

David Hinchman
Director, Information Technology and Cybersecurity
U.S. Government Accountability Office
441 G Street NW
Washington, DC 20548

Dear Mr. Hinchman:

Attached are comments on the U.S. Government Accountability Office's (GAO) report entitled, "**Data Science: NIH Needs to Implement Key Workforce Planning Activities**" (GAO-23-105594).

The Department appreciates the opportunity to review this report prior to publication.

Sincerely,

Melanie Anne Egorin

Melanie Anne Egorin, PhD
Assistant Secretary for Legislation

Attachment

**Appendix II: Comments from the Department
of Health and Human Services**

GENERAL COMMENTS FROM THE DEPARTMENT OF HEALTH & HUMAN SERVICES ON THE GOVERNMENT ACCOUNTABILITY OFFICE'S DRAFT REPORT - DATA SCIENCE: NIH NEEDS TO IMPLEMENT KEY WORKFORCE PLANNING ACTIVITIES (GAO-23-105594)

The U.S. Department of Health and Human Services (HHS) appreciates the opportunity to review and comment on the draft report from the Government Accountability Office (GAO).

Strategic workforce planning is an essential component to ensuring NIH has trained and skilled personnel in the right positions at the right time. NIH uses a variety of tools to anticipate and respond to staffing needs. NIH plans to fully implement each of the eight key data science workforce planning activities that GAO describes in the report.

NIH has taken steps to develop the agency's data science workforce, including:

- Providing Institutes and Centers with a suite of [workforce planning tools](#) to address a variety of workforce challenges such as identifying competency or skill gaps, succession planning, knowledge management, engaging and retaining staff, and analytical models to identify trends and future staffing needs.
- Providing Institutes and Centers with direct expert support for workforce planning and analytics projects and challenges.
- Establishing a workforce planning community of practice to provide a forum for Institutes and Centers to share best practices and resources.
- Implementing a Data Science Workforce Group that is charged with improving recruitment, staffing, compensation, and retention of Data Scientists at NIH.
- Increasing opportunities for expert data scientists to contribute to the NIH mission through activities such as the Data and Technology Advancement National Service Scholars ([DATA Scholars](#)) Program. These scholars spend one to two years transforming NIH programs by applying cutting-edge methods to health-related challenges. NIH will also enhance a data science pathway program to bring to the NIH early-career technologists to spend a year working on data-related projects in NIH program offices.

GAO Recommendation 1:

The NIH Director should ensure that NIH establishes a comprehensive data science workforce planning process that addresses the shortfalls noted in this report.

HHS Response:

NIH concurs with GAO's recommendation. NIH is committed to effective data science workforce planning to ensure the NIH skilled and knowledgeable staff execute the range of functions critical to the agency's mission. NIH will provide an action plan to address the recommendation in our 180-day letter response to Congress.

**Appendix II: Comments from the Department
of Health and Human Services**

GAO Recommendation 2:

The NIH Director should ensure that NIH develops staffing requirements for the data science workforce.

HHS Response:

NIH concurs with GAO's recommendation. NIH will provide an action plan to address the recommendation in our 180-day letter response to Congress.

GAO Recommendation 3:

The NIH Director should ensure that NIH reassesses its data science competency and staffing needs periodically.

HHS Response:

NIH concurs with GAO's recommendation. NIH will provide an action plan to address the recommendation in our 180-day letter response to Congress.

GAO Recommendation 4:

The NIH Director should ensure that NIH analyzes its workforce to identify gaps in data science competencies and staffing.

HHS Response:

NIH concurs with GAO's recommendation. NIH will provide an action plan to address the recommendation in our 180-day letter response to Congress.

GAO Recommendation 5:

The NIH Director should ensure that NIH develops specific strategies and plans to address identified gaps in data science competencies and staffing.

HHS Response:

NIH concurs with GAO's recommendation. NIH will provide an action plan to address the recommendation in our 180-day letter response to Congress.

GAO Recommendation 6:

The NIH Director should ensure that NIH implements strategies and plans to address identified gaps in data science competencies and staffing.

HHS Response:

NIH concurs with GAO's recommendation. NIH will provide an action plan to address the recommendation in our 180-day letter response to Congress.

GAO Recommendation 7:

The NIH Director should ensure that NIH develops and tracks metrics to monitor the agency's progress in addressing data science competency and staffing gaps.

**Appendix II: Comments from the Department
of Health and Human Services**

HHS Response:

NIH concurs with GAO's recommendation. NIH will provide an action plan to address the recommendation in our 180-day letter response to Congress.

GAO Recommendation 8:

The NIH Director should ensure that NIH develops a process to track data science staff.

HHS Response:

NIH concurs with GAO's recommendation. NIH will provide an action plan to address the recommendation in our 180-day letter response to Congress.

GAO Recommendation 9:

The NIH Director should ensure that NIH requires reporting to agency leadership on progress made in addressing data science competency and staffing gaps.

HHS Response:

NIH concurs with GAO's recommendation. NIH will provide an action plan to address the recommendation in our 180-day letter response to Congress.

GAO Recommendation 10:

The NIH Director should ensure that NIH documents new time frames to complete the guidance its staff will need to assess data management and sharing plans, and ensure that the guidance is implemented.

HHS Response:

NIH considers this recommendation as closed – implemented.

In February 2023, NIH issued extramural staff (program and grants management) guidance for implementing the requirements and expectations provided by the Final NIH Policy for Data Management and Sharing and by supplemental notices that require researchers to prospectively plan for how scientific data will be preserved and shared through submission of a Data Management and Sharing (DMS) Plan. In March 2023, NIH provided GAO with this staff guidance.

GAO Recommendation 11:

The NIH Director should ensure that NIH documents new time frames to complete the guidance its staff will need to determine researchers' compliance with their data management and sharing plan, and ensure that the guidance is implemented.

HHS Response:

NIH considers this recommendation as closed – implemented.

In February 2023, NIH issued extramural staff (program and grants management) guidance for implementing the requirements and expectations provided by the Final NIH Policy for Data Management and Sharing and by supplemental notices that require researchers to prospectively plan for how scientific data will be preserved and shared through submission of a Data

**Appendix II: Comments from the Department
of Health and Human Services**

Management and Sharing (DMS) Plan. In March 2023, NIH provided GAO with this staff guidance.

Page 4 of 4

Appendix III: GAO Contact and Staff Acknowledgments

GAO Contact

David B. Hinchman, 214-777-5719 or hinchmand@gao.gov

Staff Acknowledgments

In addition to the individual named above, Sabine Paul (Assistant Director), Cheryl Dottermusch (Analyst-in-Charge), Christopher Businsky, Donna Epler, Angel Green, Franklin Jackson, Kimberly LaMore, Serena Lo, Thomas Murphy, and Ibrahim Suleman made contributions to this report.

GAO's Mission

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

Obtaining Copies of GAO Reports and Testimony

The fastest and easiest way to obtain copies of GAO documents at no cost is through our website. Each weekday afternoon, GAO posts on its [website](#) newly released reports, testimony, and correspondence. You can also [subscribe](#) to GAO's email updates to receive notification of newly posted products.

Order by Phone

The price of each GAO publication reflects GAO's actual cost of production and distribution and depends on the number of pages in the publication and whether the publication is printed in color or black and white. Pricing and ordering information is posted on GAO's website, <https://www.gao.gov/ordering.htm>.

Place orders by calling (202) 512-6000, toll free (866) 801-7077, or TDD (202) 512-2537.

Orders may be paid for using American Express, Discover Card, MasterCard, Visa, check, or money order. Call for additional information.

Connect with GAO

Connect with GAO on [Facebook](#), [Flickr](#), [Twitter](#), and [YouTube](#).
Subscribe to our [RSS Feeds](#) or [Email Updates](#). Listen to our [Podcasts](#).
Visit GAO on the web at <https://www.gao.gov>.

To Report Fraud, Waste, and Abuse in Federal Programs

Contact FraudNet:

Website: <https://www.gao.gov/about/what-gao-does/fraudnet>

Automated answering system: (800) 424-5454 or (202) 512-7700

Congressional Relations

A. Nicole Clowers, Managing Director, ClowersA@gao.gov, (202) 512-4400, U.S. Government Accountability Office, 441 G Street NW, Room 7125, Washington, DC 20548

Public Affairs

Chuck Young, Managing Director, youngc1@gao.gov, (202) 512-4800
U.S. Government Accountability Office, 441 G Street NW, Room 7149
Washington, DC 20548

Strategic Planning and External Liaison

Stephen J. Sanford, Managing Director, spel@gao.gov, (202) 512-4707
U.S. Government Accountability Office, 441 G Street NW, Room 7814,
Washington, DC 20548

