

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.