# Strengthening Data Science Methods for Department of Defense Personnel and Readiness Missions

The Office of the Under Secretary of Defense (Personnel & Readiness), referred to throughout this document as P&R, is responsible for the total management of all Department of Defense (DoD) personnel including recruitment, readiness, and retention. This mission requires extensive data, a large number and variety of complex analyses, and access to skilled workers to extract meaningful information to guide DoD personnel and readiness policies. With the advent of newer sources of data, such as from social media, and modern data analytics, P&R has the opportunity to exploit new tools that may produce more powerful analyses and improve the effectiveness and efficiency with which it accomplishes its mission. At the request of DoD, the National Academies of Sciences, Engineering, and Medicine developed a framework for the integration of data analytics in support of decisions within the purview of P&R. This report presents strategies for improving data quality and sharing, identifying important analytic methods for supporting P&R in the long term, and strengthening data science on-the-job education to promote best practices within the P&R enterprise.

THE NEED FOR A DATA ANALYTICS FRAMEWORK

Currently, analyses developed to support P&R are often disjointed, one-time efforts that respond to immediate questions and may lack a continuity plan for future use of their data or methods. A comprehensive data and analytics framework, properly implemented, could add coherence to this work, expanding the types of questions that P&R can quickly examine, reducing the funds necessary to conduct analyses, improving the reliability of findings, and better informing policy decisions. This framework could also contribute to the goals of the Force of the Future initiative, which aims to make the DoD workforce more equitable, efficient, and flexible. Developing a data and analytics framework is complex, with many components that need to be addressed both individually and systematically. Data need to be easily accessible and shared across groups in a way that reduces the hurdles currently faced when researchers and analysts seek to find or share data, while ensuring proper privacy and security protections.
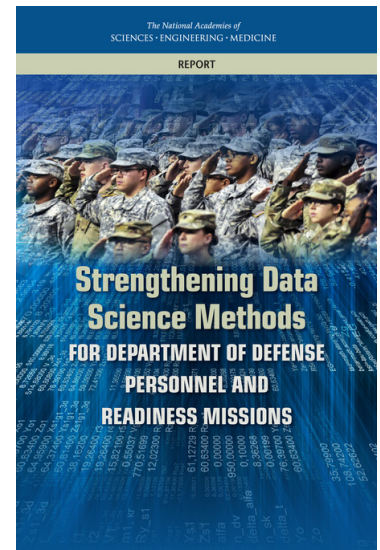
Read, purchase, or download a free PDF of this report at http://nap.edu

The National Academies of
SCIENCES · ENGINEERING · MEDICINE

IMPROVING DATA QUALITY AND SHARING

Collections of traditional administrative and transactional data used for P&R missions continue to grow due to improved technical abilities to track and store data and an increased interest in capturing data that may provide meaningful insights. One challenge is that despite the substantial data available relevant to DoD personnel, those data may not be appropriate for DoD's analytic tasks, or suitable only after considerable manipulation. While steps have been taken to simplify and unify data—such as the development of the Defense Manpower Data Center (DMDC), a unified personnel file, and the Civilian Personnel Data System—additional steps that enable more extensive data collection could lead to further benefits, including improved force readiness, better allocation of funds, and a more agile and adept workforce. Deeper data collection supports a richer picture of the individual, which allows for better matching of personnel based on skills and experience and improved identification of at-risk service members. Enhanced data sharing within DoD, across agencies, and with the research community could promote the creation of new statistical methods, tools, and products. However, the challenges of data sharing and repurposing are significant because Services and databases use different procedures, data definitions, and formatting.

RECOMMENDATION: The Office of the Under Secretary of Defense (Personnel & Readiness) should identify incentives to enhance data sharing and collection, such as the following:

- Tracking usage of data by source in repositories such as the Person-Event Data Environment and periodically reporting back to data providers on usage (e.g., number of uses, who the users are, the nature of the study, or analysis the data contributed to);

- Providing incremental funding on contracts that involve data collection and organization to cover the costs of archiving and documenting the data for other users; and

- Giving preference to projects for constructing or redesigning operational data systems that include explicit functionality to support data sharing.

RECOMMENDATION: The Office of the Under Secretary of Defense (Personnel & Readiness) should establish a working group with representation from the Services and other elements of the Department of Defense, as appropriate, to

- identify productive new fields and formats for personnel files, such as enabling the inclusion of unstructured data and free-form text in future records;

- identify opportunities for data sharing between Services and the Office of the Under Secretary of Defense (Personnel & Readiness) and within Ser-

vices and lower barriers to such sharing;

- work with organizations that provide operational data or collect it for analysis to improve data quality by providing standard ways for data users to report problems with data collections and channel those reports back to data providers when appropriate;

- clarify self-reporting rules and practices;

- identify legal and regulatory barriers to the appropriate and responsible sharing of data; and

- examine new hardware and software architectures that facilitate data access and data management.

Alternative Data Sources

The existence of alternative data sources such as social media presents a hope of deep insights relevant to the mission of P&R. Due to questions of privacy and appropriateness and to difficulties in ensuring statistical validity, the decision to follow this path requires careful consideration and additional research before being pursued further.

RECOMMENDATION: The Office of the Under Secretary of Defense (Personnel & Readiness) should investigate the feasibility of exploiting alternative data sources to augment traditional methods for measuring collective sentiment, evaluating recruitment practices, and classifying individuals (for credit-worthiness, perhaps, or for battle-readiness). Hand in hand with this effort there should be an investigation into privacy technology appropriate for these scenarios for data use.

Person-Event Data Environment

The development of DoD's Person-Event Data Environment (PDE) is a positive step toward making some data more easily accessible. However, some technical and cultural challenges pose a significant deterrent to utilizing this tool including a lengthy review process, data quality issues, and concerns over protecting sensitive personal information.

RECOMMENDATION: The Defense Manpower Data Center should assess how well the Person-Event Data Environment is working and whether it is serving its intended community. In doing so, the Center should consider taking the following steps to improve the usability of the Person-Event Data Environment and enhance its value:

- Assess if current privacy and security policies are adequate, taking into account modern methods of attack and sources of auxiliary information that can aid in these attacks, such as multiple releases of statistics and data sets, linkage attacks that make use of public sources, and chronological correlations with public sources;

- Analyze data usage information, both for privacy and determining value of assets;

- Do a better job of establishing and defining a user community for knowledge sharing. This includes improving relationships with the federally funded research and development centers doing work for the Department of Defense and determining which researchers would benefit from the capabilities of the Person-Event Data Environment;

- Remove unnecessary barriers for researchers to gain access to the system;

- Enhance computational power, memory, and tools;

- Respond to concerns about the quality and comprehensiveness of available data;

- Develop an explicit process for reporting data problems;

- Clarify data ownership rights to external data that are uploaded and merged;

- Assess protocols for accessing personally identifiable information;

- Review approval process for exporting analysis results; and

- Consider widening access to the data and/or re-balancing Institutional Review Board requirements by establishing a differentially private interface.

Improving Privacy Governance Practices

When conducting analyses relating to personnel data, it is essential that privacy, confidentiality, and fairness be built into the work plan from the start. The current privacy and confidentiality protections in place with government databases rest heavily on Institutional Review Board (IRB) supervision. However, the current use of IRBs (especially cumbersome re-reviews) can delay research and still do little to stop personal sensitive information from being compromised. Researchers often face multiple IRB reviews and re-reviews throughout a single study, which can significantly slow the analysis process and make it difficult to respond to policy needs in a timely manner.

RECOMMENDATION: In order to support timely and efficient research, the Office of the Under Secretary of Defense (Personnel & Readiness) should encourage streamlining of Institutional Review Board processes that involve multiple organizations—for example, federally funded research and development centers and the Department of Defense.

RECOMMENDATION: The Department of Defense should carry out research on the feasibility of differential privacy methods for its personnel analytics. These methods could reduce the need for Institutional Review Board oversight.

RECOMMENDATION: The Department of Defense should consider adopting or adapting the privacy and governance structure developed by the Office of Management and Budget for civilian statistical agencies. In particular, the Department should follow the guidance on use of administrative records and establishing of statistical units under the Confidential Information Protection and Statistical Efficiency Act for both military and civil service personnel. In doing so, the Department should examine the applicability of Fair Information Practice Principles in the treatment of Defense Manpower Data Center data.

ENHANCING ANALYTIC METHODS

While comprehensive and reliable data are essential in informed decision making, they would be of little use without advanced analytic capabilities. New methods of analysis are increasingly available, and many cutting-edge approaches are ready to be more thoroughly applied for P&R missions. While a wide range of problems is being addressed for P&R using data analytic techniques and rich data sources, these are often applied in response to specific questions rather than incorporated into a long-term plan.

RECOMMENDATION: The Office of the Under Secretary of Defense (Personnel & Readiness) should assess which predictive and prescriptive analyses would benefit its mission over the longer term, taking into account its understanding of which specific decisions could, if evaluated by applying more powerful data and/or methods, better enable the Department of Defense to prepare for future demands it may face. Some possible steps that might follow include these:
- Emphasizing the use of prescriptive analytics in conjunction with predictive what-if scenarios;

- Enhancing prescriptive analytics usage and disseminating best practices across the entire department; and

- Adapting the prescriptive analytics methods successfully used in the private sector for workforce and talent management.

DoD does not routinely employ controlled experiments to understand causes and effects of P&R policies—for example, revisions to enlistment standards or choices affecting family welfare—to judge whether they produce the intended effects and provide benefits that justify their costs. Controlled experiments could be used for a variety of areas important to P&R and could help support conclusions that contradict accepted propositions.

RECOMMENDATION: To the extent feasible and relevant, the Department of Defense should conduct carefully structured experiments to test the efficacy of policy.

STRENGTHENING DATA SCIENCE EDUCATION

By strengthening the data analytics expertise of a portion of its staff, both military and civilians, P&R could quickly answer questions and better integrate analyses into policy decisions. This expertise is also important to aid the transfer of best practices and skills across the P&R enterprise.

RECOMMENDATION: The Office of the Under Secretary of Defense (Personnel & Readiness) should enhance education in data science for its personnel, including civil service employees. This education could range from short courses in specific techniques for personnel who already have the requisite foundational knowledge, to overview seminars for managers who need to be acquainted with what their analytic staff can undertake, to formal degree programs, whether at Department of Defense or civilian universities.

Learn more about Math and Statistics at the National Academies by signing up for our mailing list:

eepurl.com/bPe5gb

Division on Engineering and Physical Sciences

The National Academies of
SCIENCES · ENGINEERING · MEDICINE

The nation turns to the National Academies of Sciences, Engineering, and Medicine for independent, objective advice on issues that affect people's lives worldwide.
www.national-academies.org