

July 3, 2019

Karen Dunn Kelley, Deputy Secretary of Commerce
Department of Commerce
Docket Number USBC-2019-0001
Leveraging Data as a Strategic Asset Phase 1 Comments

Dear Deputy Secretary Kelley:

This letter is submitted on behalf of these 21 undersigned members and partners of the Postsecondary Data Collaborative (PostsecData), in response to the [Draft 2019-2020 Federal Data Strategy Action Plan](#). PostsecData is a nonpartisan coalition of organizations representing students, higher education institutions, states, employers, researchers, and privacy and security experts that seeks to advance the use of high-quality data to improve postsecondary student success and educational equity. We strongly believe that high-quality data are a prerequisite to improving student outcomes, addressing long-standing racial and socioeconomic gaps in educational opportunity, and giving all postsecondary stakeholders the information they need to make important decisions related to postsecondary education.

Understanding that this action plan intends to establish the foundation for activity alignment across agencies and define tools, processes and capacity for data stewardship and strategic data use, this letter includes the following recommendations that provide examples or additional actions needed to achieve Federal Data Strategy goals (questions 1 and 2):

- 1. Establish a data infrastructure that safeguards personally identifiable information, secures data, and promotes ethical data use.** (Actions 3 and 4)
- 2. Standardize and streamline access to data for researchers and agencies with clear and specific research questions.** (Actions 6 and 7)
- 3. Engage diverse cross-functional and cross-agency teams throughout the action plan.** (Actions 1 and 12)
- 4. Prioritize data governance, planning, and operational capacity to ensure success in each step of the action plan.** (Actions 2, 5, 13, 14, 15, and 16)

PostsecData strongly supports the goal of leveraging federal data as an asset to promote transparency and effectively steward the federal government's critical investments in higher education. Partners agree that the federal government plays a key role in collecting and reporting consistent, comprehensive postsecondary data in ways that promote student success and educational equity. We believe that thoughtful implementation of Federal Data Strategy plans can help us to achieve these goals.

1. Establish a data infrastructure that safeguards personally identifiable information, secures data, and promotes ethical data use.

As the Federal Data Strategy seeks to align and streamline data use across government agencies, researchers, and contractors, it is of the utmost importance that the federal government establish a data infrastructure that protects and secures personally identifiable information (PII) while allowing for appropriate data access and use.

PostsecData supports the Federal Data Strategy in its endeavor to streamline federal data collections and data systems in ways that centralize the importance of privacy, security, and confidentiality, especially when matching data across sources. Federal standards under the Federal Information Security Management Act (FISMA) and the Family Educational Rights and Privacy Act (FERPA), as well as those set by the National Institute of Standards and Technology (NIST), should be reviewed and incorporated, where appropriate, when implementing data matches, building data systems, or providing data access. It is imperative that all data collections and matches are limited to those metrics that have a specific and valuable purpose in meeting administrative, policymaking, consumer choice, or institutional improvement needs. These efforts also should prioritize robust reporting functions that return data back to colleges and states for benchmarking and improvement purposes, adding value for the data providers.

Action 4 requires that the federal government “develop a consistent approach for measuring and mitigating the risk of re-identification” through a data protection toolkit. We fully support this action step and recommend that the Federal Data Strategy consider the various points in the collection and reporting process where data are vulnerable and identify strategies for mitigating risk throughout the process. Recent research outlines a “Five Safes” framework that demonstrates how agencies and researchers can effectively diminish risk by creating “safe” spaces in the data infrastructure and flow.¹ The five safes aim to control data access and use through safe projects, safe people, safe settings, safe data, and safe outputs. The final safe—outputs—addresses *Action 4* directly by recommending additional protection of data outputs using techniques like rounding, aggregating, and suppressing results to obscure PII and safeguard against reidentification. For example, Federal Statistical Research Data Centers apply protections through disclosure review boards before publishing to avoid reidentification. The government should also be forward-looking in developing a data protection toolkit, taking advantage of new technologies and approaches that can minimize re-identification risk, such as differential privacy.

To ground the data infrastructure in an ethical framework as *Action 3* recommends, the Federal Data Strategy should ensure that individual data are protected, projects are thoughtfully implemented and follow the “Five Safes” protocols, that data are used in ways that do not compromise the integrity of the data or relationships developed as part of this initiative, and that data are used to help, never to harm, individuals.² As government develops technology and research, considerations about ethical and appropriate data use should be at the center of data governance, access, and control policies.

2. Standardize and streamline access to data for researchers and agencies with clear and specific research questions.

Research plays an essential role in informing postsecondary policy, consumer choice, and institutional improvement, meaning that researchers are key users of federal data. We are pleased to see that *Action 6* requires the federal government to pilot a one-stop standard application for accessing federal data for

¹ O’Hara, A. (2019) Postsecondary data infrastructure: What is possible today. Retrieved from Institute for Higher Education Policy website: <http://www.ihep.org/research/publications/postsecondary-data-infastructure-what-possible-today>

² O’Hara, A. (2019) Postsecondary data infrastructure: What is possible today. Retrieved from Institute for Higher Education Policy website: <http://www.ihep.org/research/publications/postsecondary-data-infastructure-what-possible-today>

research and evidence-building. Processes for accessing and protocols for use of data should be transparent for researchers and consistent across federal agencies.

The Department of Education's (ED) National Center for Educational Statistics' (NCES) longitudinal studies are an important example of quality data use. NCES provides researchers with differing levels of access to data, applying appropriate safeguards for each level of access while allowing researchers to study critical issues like college access, success, affordability, and equity. Researchers and other stakeholders are granted public access through the NCES data tool, PowerStats, to conduct aggregate analyses with no access to underlying student data. Qualified researchers conducting more in-depth analyses can apply for access to datasets with individual-level data if they adhere to strict privacy and security protocols, including being subject to random security checks.

The Federal Data Strategy should work with NCES, other federal statistical agencies, federal research centers, and data users to determine a process for the pilot that leverages best practices from each. The one-stop standard application should seek to ensure that only "safe people" have access to data and understand the parameters within which they can use it, by requiring clearance processes like background checks, training, or specific organizational affiliation.³ Examples of safe data access practice exist in a variety of agencies like NCES, the Centers for Medicare and Medicaid Services, and the Internal Revenue Service, among others described in recent research.⁴

In addition to standardizing the clearance required of data users, we are supportive of *Action 7*, which aims to streamline data and reduce duplication for federal agencies and data reporters through an automated inventory tool for Data.gov. By standardizing and streamlining data access and controls, NCES – which is leading this Action – can ensure that the government and researchers have access to the data they need while not overburdening federal resources or data providers. We strongly advise that while the federal government takes specific steps to identify duplicative reporting requirements, it should first put in place the proper legal agreements and build the technological capacity for data sharing across federal agencies before halting any collections.

3. Engage diverse cross-functional and cross-agency teams throughout the action plan.

As the Federal Data Strategy works toward creating systems and processes that allow agencies, policymakers, and researchers to build evidence and strategically leverage government data assets, it should engage diverse cross-functional and cross-agency teams through the process. We are heartened to see *Action 1* and *Action 12*, which create an Office of Management and Budget (OMB) Data Council and constitute a diverse data governance body, respectively. As outlined in the Foundations of Evidence-Based

³ O'Hara, A. (2019) Postsecondary data infrastructure: What is possible today. Retrieved from Institute for Higher Education Policy website: <http://www.ihep.org/research/publications/postsecondary-data-infastructure-what-possible-today>

⁴ O'Hara, A. (2019) Postsecondary data infrastructure: What is possible today. Retrieved from Institute for Higher Education Policy website: <http://www.ihep.org/research/publications/postsecondary-data-infastructure-what-possible-today>

Policymaking Act of 2018,⁵ the data governance body should include not only Chief Data Officers, but also information and evaluation officers from each participating agency, including statistical agencies, with each organization receiving equal representation, regardless of staff size. Further, these bodies should confer with other stakeholders, including information technology professionals, privacy experts, budget and policy analysts, and contractors who regularly use the data. The Federal Data Strategy should create additional governance subcommittees, comprised of representatives of agencies that regularly collaborate and match data, to ensure regular communication and continuous process improvement (e.g., representatives from the Departments of Education, Defense, Treasury, Labor, and Veterans Affairs would meet to discuss data sharing on postsecondary outcomes). Decision-making processes should include public comment periods, similar to governance processes outlined in the College Transparency Act.⁶ It is clear through the work of the Federal Data Strategy that there is cross-agency interest in more fully leveraging existing data and resources and it cannot be done by a single office. To ensure that data are collected, used, and reported in a way that protects individuals and builds a body of evidence for policymaking, a variety of perspectives must be incorporated into process-building and dissemination, including prioritizing racial, ethnic, and gender diversity on these bodies, in addition to job function.

4. **Prioritize data governance, planning, and operational capacity to ensure success in each step of the action plan.**

At the core of accomplishing all that the Federal Data Strategy action plan lays out is the ability to establish solid data governance protocols that are standardized across the federal data ecosystem and to build and maintain operational capacity of federal agencies to properly use and steward their data.⁷ *Actions 2, 5, 13, 14, 15, and 16* require different agencies to build a catalog of data science training and credentialing; to make certain that government workers can include data science in their professional development; to create a repository of tools and resources; and to assess readiness within each agency to undertake the changes necessary to meet the goals of the Federal Data Strategy, including identification of key datasets and research questions for agencies. By focusing on governance, planning and operational capacity, agencies will be well-positioned to leverage existing data assets to uncover and remedy systemic, long-standing racial and inequities. In the postsecondary context, for example, with better data, we can ensure that all students and outcomes are counted and understand how our education systems are serving all students, especially students of color, low-income students, student veterans, and adult students.

While the OMB Data Council and the diverse data governance body required by the Foundations of Evidence-Based Policymaking Act can and should guide agencies in standardization and implementation of data governance practices, each agency has a clear responsibility to safeguard and steward their data, which will require adherence to implementation plans and staff capacity to implement those plans well. It is important that the action plan includes assessment of and reflection on governance, maturity, and

⁵ Foundations of Evidence-Based Policymaking Act, Public Law 115-435, 115th Congress, <https://www.congress.gov/bill/115th-congress/house-bill/4174/text>

⁶ College Transparency Act, S. 800, 116th Congress, <http://www.congress.gov/bill/116th-congress/senate-bill/800>

⁷ Roberson, A.J., Rorison, J. & Voight, M. (2017) A blueprint for better information: Recommendations for a federal postsecondary student-level data network. Retrieved from Institute for Higher Education Policy website: <http://www.ihep.org/research/publications/blueprint-better-information-recommendations-federal-postsecondary-student>



capacity at each participating agency, and we encourage the Federal Data Strategy to embed this review at regular intervals after implementation and into the maintenance phase.

Thank you for your focus on improving the structure and utility of national data resources. We hope that this letter serves as a resource to improve the action plan of the Federal Data Strategy and look forward to seeing postsecondary data become more transparent, comprehensive, and efficient with your support. Ultimately, thoughtful use of high-quality postsecondary data can spur change by identifying postsecondary trends, shining a light on racial and socioeconomic inequities, and informing decision-making at all levels. If you have any questions about these comments then please contact IHEP's vice president of policy research, Mamie Voight (mvoight@ihep.org or 202-587-4967).

Sincerely,

Alloy Engineering Co., Inc.
Association for Career and Technical Education (ACTE)
Center for Law and Policy
Complete College America, Inc.
Credential Engine, Inc.
Data Quality Campaign
Future of Privacy Forum
Georgetown University Center on Education and the Workforce
Higher Learning Advocates
Institute for Higher Education Policy
Knowledge Alliance
Lehman College of The City University of New York
National Association of Student Financial Aid Administrators
National Center for Higher Education Management Systems (NCHEMS)
New America Higher Education Initiative
Student Veterans of America
The Century Foundation
The Education Trust
The Institute for College Access & Success
The Ohio State University
Workforce Data Quality Campaign