

ENVISIONING THE NATIONAL POSTSECONDARY DATA INFRASTRUCTURE IN THE 21ST CENTURY

EXECUTIVE SUMMARY

Leveraging What We Already Know: Linking Federal Data Systems

DAVID A. BERGERON
CENTER FOR AMERICAN PROGRESS

MAY 2016

This publication is based on research funded in part by the Bill & Melinda Gates Foundation. The findings and conclusions contained within are those of the authors and do not necessarily reflect positions or policies of the Bill & Melinda Gates Foundation.

Executive Summary

Introduction

The federal government has access to a significant and growing body of information about individual citizens. While concerns about privacy and security are frequently expressed about the government holding such information, having access to this information is critical not only to conduct very basic government operations but also to serve citizens' needs, and the data are already subject to laws and regulations to ensure its safe and secure use.

The U.S. Department of Education (ED) holds a vast amount of data on individuals who applied for and received federal student aid. Each year, ED receives and retains income and asset information for the more than 20 million students who apply for federal aid on the Free Application for Federal Student Aid (FAFSA). ED augments the information received on the FAFSA with information from a number of other federal data sources. These data are retained and the body of information about aided students grows as the students progress through college, then graduate, and then enter student loan repayment.

But the federal government knows far more about individual citizens than does ED, especially if the citizen, whether a college graduate or not, did not receive federal student aid funds. Having access to information about these non-aided individuals is critical to the nation's postsecondary education data ecosystem. These data include information held by the U.S. Department of Health and Human Services' (HHS's) National Directory of New Hires (NDNH), the Social Security Administration's (SSA's) wage and earnings data, the Internal Revenue Service's (IRS's) tuition and required fees and financial aid data, the U.S. Department of Defense's (DoD's) military recruiting data, and data from the U.S. Department of Veterans Affairs (VA).

All personally identifiable data maintained by the federal government are governed by the provisions of the Privacy Act of 1974.¹ The Privacy Act prohibits the disclosure of information without the written consent of the person identified in the data, unless the disclosure is authorized under one of several statutory exceptions. The statutory exceptions include such uses as statistical purposes by the Census Bureau and the Bureau of Labor Statistics, routine uses within a U.S. government agency, archival purposes as a record that has sufficient historical value, law enforcement purposes, congressional investigations, and other administrative purposes.

Role in the National Postsecondary Data Ecosystem

At present, there is very little public or political support for making use of the broad range of information available to improve the national postsecondary education data ecosystem. When someone in the general public thinks about higher education, they likely think about how colleges and universities perform on college rankings and other publicly available data sources. But this information often leaves many unanswered questions:

- ▶ How likely is it that a student like me graduates from this or any institution of higher education on time?
- ▶ How much money will I likely make if I attend this institution and major in this field?
- ▶ How much will it really cost me to attend this college for a year? For the entire program? How does this compare with the same program at another college?
- ▶ What will likely happen if I transfer from this college majoring in this field to that college majoring in that field?
- ▶ How likely is it that I'll be able to repay my student loan if I attend this institution and major in this field?

Leveraging the data the federal government already has about its citizens would augment the information that ED already collects from students and families and institutions of higher education, and permit it to answer these vital questions that millions of students and their families ask every year.

At the national level, the federal agencies that are deeply invested in higher education—IRS, SSA, HHS, VA, and DoD—do not view themselves as data providers in the postsecondary education data ecosystem despite the roles they have played in recent years in describing the outcomes of that system.

A key challenge is to build public and elected-official support for an expanded view of what constitutes the national postsecondary education ecosystem to include these agency data resources by leveraging the data that already exists about postsecondary outcomes. Eliminating the data silos and allowing the federal government to share data more effectively would have significant benefits to students and families, state and institutional leaders, and federal policymakers.

Major Issues

Breaking down the barriers between siloed information that the federal government already has about citizens who opt into (or out of) postsecondary education would improve policymaking, improve the services provided to citizens, and help potential postsecondary students make better choices about their education. But these barriers are perhaps the greatest challenge. Each federal agency has a unique reason for collecting and holding the data it has, and it has little incentive to share this information with other federal agencies.

Sharing data presents real systemic problems. First, there are often specific legal constraints on the use of the data that federal agencies collect and hold. For example, in authorizing expansive data collections, like the NDNH, which collects information on a quarterly basis on everyone who works in the United States, Congress placed significant limitations on the data's use. Such limitations are extremely common.

Second, federal agencies often do not “speak the same language”—they lack common identifiers or define terms in inconsistent ways. The lack of common identifiers—like Social Security Numbers for individuals or common identifiers for institutions of higher education—make it very difficult for agencies to share information in a sufficiently accurate manner.

Finally, costs to the agencies involved in sharing information are frequently unknown at the outset and unbudgeted.

Technical Enhancements Needed to Improve the Ecosystem

Enhancements to Facilitate Data Exchanges

ED should begin to work with the Office of Management and Budget to fully implement provisions of existing law, and assign each institution participating in Title IV programs with a single ED identification number to uniquely identify each postsecondary education institution in all federal programs,² so that information about institutions can be readily shared and compared, to help identify effective best practices.

Enhancements to Improve Data Collection

The process of collecting information to determine veterans' eligibility for education benefits should be fully automated, and institutional reporting should be streamlined to make it easier for veterans to access their benefits while ensuring proper use of scarce resources. Similarly, the data collection for military recruiters needs to be automated and systematized as quickly as possible to reduce cost and burden, and improve coverage.

The multiple data collections from employers and institutions should be consolidated. The consolidation of the data collections into an integrated system would reduce cost to the government and burden on institutions and employers while increasing consistency and comprehensiveness.

The institutional collection could be accomplished by having institutions transmit the information needed by IRS, DoD, and VA to a centralized location. The core of the collection would be the information IRS already collects—essentially tuition and fees paid and scholarships received by those enrolled in postsecondary education—augmented to include the information needed by military recruiters—including the student's age and date of birth, place of birth, level of education, academic major, and degrees received.

ED should immediately begin to work with DoD, VA, and IRS on the development of such a system to exchange data with an eye toward obtaining information on completion for aided students.

Enhancements to Data Use

A mechanism to routinely generate summary statistical information from IRS, VA, and SSA records needs to be created. The exact characteristics of the information to be generated should be carefully explored but should include the calculation of net price using IRS data and the development of a standardized net price calculator. Estimates of labor market outcomes should also be developed at an institutional level, and this information should be combined with the net price to develop preliminary estimates of a return on investment for each institution of higher education.

Resources Needed to Improve the Ecosystem

The federal government already expends considerable resources to collect and analyze the data received from institutions and employers. Much of what is proposed in this paper could be accomplished without a significant increase in resources. To the extent that additional resources are needed, it is likely that they could be obtained from the Office of Management and Budget, which has resources available for cross-agency enhancements to federal data resources. The federal government could pilot a program to enhance the national postsecondary data infrastructure beginning with the automation of the military recruitment collection, including the capacity to identify veterans who are enrolled in postsecondary education.

Recommendations

Recommendations to Build Public Support

Advocates for enhancing our knowledge of the impacts of postsecondary education should use the critical moment when student and families are trying to make choices among postsecondary education institutions and programs to showcase what we do not—but need to—know about the outcomes of the investment. This should build public support for greater access to information about the outcomes of higher education. In this regard, business leaders and other employers must play a galvanizing role by demanding greater transparency from higher education institutions for information on outcomes for graduates from specific educational programs.

Recommendations for Federal Policymakers

ED must play a leadership role and begin the conversation across the government to leverage the existing data resources by demonstrating the benefits that would accrue to the other federal agencies.

The ban on a student unit record system should be repealed and linkages to other federal data resources like earnings information from SSA and completions information from DoD should be developed to reduce redundancy and cost and increase access and use.

Access to earnings data should be provided to the Secretaries of Education and Veterans Affairs, so they can evaluate the economic impact of the programs administered by their agencies. This would enable both agencies to comply with the requirements of the Government Performance and Results Act without imposing additional reporting burden on institutions of higher education. It would also ensure that students and families have access to the most accurate and timely data on the earnings of program completers as well as those who dropped out.

Consideration should be given to consolidating the NDNH with SSA's wage and earnings reporting. Requiring states to build an infrastructure that already largely exists at the national level is unnecessarily burdensome and expensive.

A statutory exception to the Privacy Act of 1974 should be provided to specifically permit the exchange of federal data on individuals for the purposes of improving service to students and families and to permit a better understanding of the effectiveness of the federal student aid system and our nation's higher education institutions.

Recommendations for State Policymakers

State policymakers should explore the possibility of using the NDNH to augment information from other postsecondary education data resources to enhance its primary purpose of improving collected information on newly hired individuals in order to aid in the enforcement of child support orders. Having access to information about the educational attainment of noncustodial parents who recently became employed might improve states' ability to craft and enforce child support orders.

States should explore whether their longitudinal data systems could be enhanced by including information from the NDNH. While the restrictions on NDNH use are significant, an opportunity might exist to exploit some of the permissive language on state use of the data to augment the state systems.

States should consider entering into an agreement with SSA to obtain aggregate earnings information in a manner that would allow the calculation of a return on investment for public colleges and universities. Having information on individual students is unnecessary to calculate a valid return on investment measure, so the limitation on release of SSA data likely would not inhibit the calculation of such a measure.

1 5 U.S.C. § 552(a).

2 20 U.S.C. § 1094(b).

Envisioning the National Postsecondary Data Infrastructure in the 21st Century is a project of the Institute for Higher Education Policy and is supported by the Bill & Melinda Gates Foundation.

